

***Proceedings of the 1st Interdisciplinary PhD
Workshop of Media Computer Science 2019***

Jun.-Prof. Dr. Danny Kowerko, Prof. Dr. Marc Ritter,
Prof. Dr. Thomas Köhler und Dr. Kristina Helle (Hrsg.)

TUD*press*

IBS Scientific Workshop Proceedings

Herausgegeben von Stiftung IBS, Wolfram Hardt

Band 8

***Proceedings of the 1st
Interdisciplinary PhD Workshop of
Media Computer Science 2019***

Jun.-Prof. Dr. Danny Kowerko, Prof. Dr. Marc Ritter,
Prof. Dr. Thomas Köhler und Dr. Kristina Helle (Hrsg.)

TUD*press*
2020

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im
Internet über <http://dnb.d-nb.de> abrufbar.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Na-
tionalbibliografie; detailed bibliographic data are available in the Internet
at <http://dnb.d-nb.de>.

ISBN 978-3-95908-196-2

© 2020 TUDpress
THELEM Universitätsverlag
GmbH und Co. KG
<http://www.tudpress.de>

Alle Rechte vorbehalten. All rights reserved.
Gesetzt von den Herausgebern.
Printed in Germany.

Vorwort

Im Zeitraum vom 5. bis 7. Januar 2019 fand im Internationalen Informatik- und Begegnungszentrum Sachsen IBS in Laubusch der „1st Interdisciplinary PhD Workshop of Media Computer Science 2019“ statt. Das übergeordnete Ziel des Workshops bestand darin, die Fallstricke einer institutionellen Forschung in der Startphase der Promotion zu überwinden und dabei die angehenden PromovendInnen zu befähigen, ihre wissenschaftlichen Beiträge innerhalb einer wohlwollenden, aber kritisch eingestellten Community einzuordnen.

Thematisch angesiedelt war der Doktorandenworkshop an der Schnittstelle verschiedener Disziplinen und Fachbereiche aus der Informatik und den Sozialwissenschaften. Er förderte durch kooperatives Engagement von neun Professoren und Betreuern aus drei sächsischen Hochschulen (HS Mittweida, TU Chemnitz und TU Dresden) in einem innovativen 3-tägigen Format das wissenschaftliche Arbeiten und die Entwicklung von Soft Skills der 18 PromovendInnen in besonderem Maße. Hierbei wurden klassische Elemente und Methoden eingesetzt, die von der gegenseitigen Vorstellung der individuellen Arbeitsthemen bis zu wissenschaftlichen Arbeitsanleitungen reichten. Darüber hinaus sollten die Promovenden mit Unterstützung erfahrener Wissenschaftler und Projektleiter innerhalb eines innovativen Formats wissenschaftliche Domänen in Kurzform gruppenbasiert ergründen und gewonnene Erkenntnisse präsentieren. Ein Open Review der eingereichten Promotionsbeiträge komplementierte die zuvor aufgeführten Formate.

Der Promovendenkreis entstammte den nachfolgenden Nachwuchsforschergruppen:

Agile Publika: Die vom European Social Fund (ESF) geförderte Nachwuchsforschergruppe, bestehend aus fünf Doktoranden und einem Postdoc, erforscht Themenstellungen, welche für die lokale Öffentlichkeit von besonderer Relevanz erscheinen. Dabei stehen Bewegungsmuster und Nahverkehr ebenso im Fokus wie beispielsweise Meinungsbildungsprozesse in sozialen Medien.

Methodisch liegt der Schwerpunkt bei der Untersuchung dieser Phänomene mittels computergestützter maschineller Analysen. Die Gruppe ist interdisziplinär und hochschulübergreifend zwischen den Fakultäten Medien und Angewandte Computer- und Biowissenschaften der Hochschule Mittweida und dem Medienzentrum der TU Dresden am Schnittpunkt von Informatik und Kommunikationswissenschaft aufgestellt.

localizeIT: Im Rahmen der Innovationsinitiative Neue Länder “Unternehmen Region” förderte das BMBF seit dem Jahr 2005 unterschiedliche Projekte, wozu das Projekt sachsmidia (2007 bis 2012) sowie die Stiftungsjuniorprofessur Media Computing mit ihren 5 Mitarbeitern im Projekt “localizeIT – Lokalisierung visueller Medien” an der Fakultät Informatik der TU Chemnitz (2014 bis 2019) zählen. Eine wesentliche Zielsetzung besteht darin, durch automatische Analyse von Ton-, Bild- und Videomaterialien zu einem semantischen Verständnis der Medien der lokalen Fernsehlandschaft zu gelangen, wobei insbesondere die Klassifikation und Verortung der darin enthaltenen Personen und Objekte eine große Rolle spielen.

Der Tagungsband präsentiert die im Ergebnis des Workshops überarbeiteten Beiträge, welche aus dem interdisziplinären Austausch von Wissenschaftlern und Nachwuchsforschern der TU Chemnitz, der Hochschule Mittweida und der TU Dresden hervorgegangen sind und sich in die Schwerpunkte Maschinelles Lernen und Life Sciences, Multimodale Datenanalyse und Big Data, Social Analytics und E-Learning einteilen lassen.

Viel Freude beim Lesen wünschen Ihnen
Ihre Herausgeber und Workshop-Organisatoren

Jun.-Prof. Dr. Danny Kowerko (TU Chemnitz, Jun.-Prof. Medieninformatik)
Prof. Dr. Marc Ritter (HSMW, Professur Medieninformatik)
Prof. Dr. Thomas Köhler (TU Dresden, Medienzentrum)
Dr. Kristina Helle (HSMW, Modellbildung und Simulation)

Inhaltsverzeichnis

Thema 1: Maschinelles Lernen und Life Sciences

Comparative Case-Study Between Different Deep Learning Frameworks on Varied OCT-Classification Problems	1
<i>R. Hasan, H. Langner, M. Ritter, M. Eibl</i>	
Size Doesn't Matter? Comparing Learning Curves for Data Subsets Using Different Classifiers	15
<i>R. Vogel, K. Schneider, K. B. Helle, M. Ritter</i>	
Hexagonale Bildverarbeitung im Kontext maschineller Lernverfahren: Konzeption eines biologisch inspirierten hexagonalen Deep Learning Frameworks	27
<i>T. Schlosser, M. Friedrich, D. Kowenko</i>	
Generierung hexagonaler Bilddaten und Vergleich mit konventioneller Rasterung in der Bildverarbeitung	39
<i>R. Manthey</i>	

Thema 2: Multimodale Datenanalyse und Big Data

Generierung eines Aktivitätsklassifikators für den Einsatz in mobilen kontextsensitiven Anwendungen	47
<i>T. Rolletschke</i>	
Anpassung multipler ML-Frameworks der Personen- und Ortserkennung zur Verwendung in der EMSML-Umgebung	59
<i>C. Roschke, R. Thomanek, B. Platte</i>	
Bildbasierte Aktivitätsanalyse mittels dezentraler Prozessinstanzen auf der Grundlage einer einheitlichen Datenbasis	73
<i>R. Thomanek, C. Roschke, B. Platte</i>	

Lexikonbasierte Sentimentanalyse von Online-Kundenbewertungen 87
C. Hösel

Entwurf einer interaktiven, dynamischen Logging-Schnittstelle in der Videospiel-Engine Unity 103
M. Heinzig, M. Ritter

Thema 3: Social Analytics und E-Learning

Optimization of a Hybrid Learning Approach for Power Electronics Course Using Virtual Laboratory 113
M. Ali, K.D. Irianto, D. Hariyanto, T. Koehler, D. Mardapi, and Sunarto

Interaktives Storytelling - Mit Geschichten spielend online lernen Erklärvideos im wissensvermittelnden Kontext an Hochschulen . . 123
R. Fleck

Eine netzwerkanalytische Aufarbeitung des Nahverkehrs in Mittelsachsen 137
C. Schulte

Wissenstransfer in der Bürgerbeteiligung auf kommunaler Ebene . 147
C. M. Brinkmann

Comparative Case-Study Between Different Deep Learning Frameworks on Varied OCT-Classification Problems

Rama Hasan¹, Holger Langner², Marc Ritter², Maximilian Eibl¹

¹ Chemnitz University of Technology, Media Informatics, 09111 Chemnitz, Germany

² Hochschule Mittweida, Media Computing, 09648, Mittweida, Germany
rama.hasan@s2017.tu-chemnitz.de

Abstract We introduce an experimental-based study to choose a Deep Learning (DL) framework between three of the mostly used: Tensorflow, Keras (Tensorflow backend) and Theano, to achieve a good classification accuracy of the Convolutional Neural Networks (CNNs). Two CNN architectures have been built and trained on three different image classification problems from macular OCT image data related to three training problems: discrimination between two forms of AMD, visual acuity quantification, and qualitative therapy outcome prediction. The challenge is the difficulty to classify this kind of problems, also by human ophthalmologists, for that, a CNN has been built, trained and evaluated using the three previous frameworks to choose the D-framework which achieves a better classification performance of the CNN for this kind of images. The experimental results illustrate the superiority of the CNN model built by Keras over Theano and Tensorflow.

Keywords CNN, Tensorflow, Keras, Theano, OCT

1 Introduction

Deep Learning (DL) is the branch of Machine Learning based on Deep Neural Networks (DNNs). Convolutional Neural Networks (CNNs) are one of the most popular DNN architectures. In recent years, CNNs have been used widely as a powerful tool to solve several machine-learning

tasks in computer vision [1] such as object detection [2], semantic segmentation [3], human pose estimation [4], etc. Various open-source DL-Frameworks are used to find easy solutions to complex problems, while they offer building blocks for designing, training and validating deep neural networks through a high-level programming interface. Widely used deep learning frameworks such as Tensorflow [5], Caffe [6], Torch [7], CNTK [8] and others rely on GPU-accelerated libraries such as cuDNN [9] and NCCL [10] to deliver high-performance multi-GPU accelerated training. Each of these frameworks has different features and performance characteristics, and each one provides different techniques to optimize its implementation of DL-algorithms. Choosing the proper framework to use it in research or an application takes a lot of time and effort and the effort invested by the research community in scientifically comparing these DL-frameworks is limited. Bahrampour et al. [11] did a comparative study between DL frameworks. The study was carried out on Caffe, Neon, Tensorflow, Theano, and Torch DL-frameworks, by training convolutional and stacked autoencoder networks on the MNIST dataset [12] and the ImageNet dataset [13]. For the experiments, a single machine for both CPU (multithreaded) and GPU (Nvidia Titan X) is used. The three considered aspects were extensibility (the ability to support various deep architectures), hardware utilization, and speed (gradient computation time and forward time). In terms of extensibility, the experiments clearly demonstrated the superiority of Theano and Torch. Torch took the first place, second was Theano in terms of suitability for any deep learning architecture of CPU. This study illustrated that the preferable performance was achieved by Theano on GPU for training and deployment of LSTM networks. On the other side, Caffe was the simplest for evaluating the performance for standard deep architectures, and Tensorflow appeared as a very flexible framework. However, it results in the worst performance on a single GPU based on the gradient computation time and forward step time compared to the remaining studied frameworks.

The Theano development team of Al-Rfou et al. [14] introduced a comparison between Theano, Torch, and Tensorflow on three types of machine learning models including CNN, RNN and sequence-to-sequence mapping RNN. Shatnawi [15] introduced a qualitative and quantitative study between Theano, CNTK, and Tensorflow. The aim of this study was to help different users to choose the most proper

DL-Framework related to their requirements. Several experiments are implemented using several benchmark datasets, the performance of the three framework implementations of different DL algorithms have been measured. Moreover, the results illustrated that for training networks, the processing time clearly demonstrates the advantage of GPU over CPU. Tensorflow tends to have the lowest GPU utilization metric, followed by Theano and CNTK respectively. On the other hand and by CPU utilization metric Theano has the lowest value followed by Tensorflow and CNTK, respectively. In terms of memory utilization on CPU and GPU, the results were close to each other. CNTK appears to be the best framework in terms of GPU processing time compared to Tensorflow and Theano. Kovalev [16] introduces a comparative study between the following DL framework: Theano, DeepLearning4J [17], Torch, Caffe, and Tensorflow in terms of convergence time, prediction (classification) time, classification accuracy, and the size of source code. The experiments were conducted on the MNIST hand written digits dataset using CPU. The goal of our study is to compare the three most popular DL-frameworks (Tensorflow, Keras, and Theano), in terms of classification accuracy using two different CNN architectures. Two CNN classifiers have been built and trained on three different OCT training problems, to help the ophthalmologists with their decision, and to help the researchers who used to train their own CNN models. The remainder of this study is organized as follows: in section two, the description of the OCT datasets is presented. A brief overview of our Convolutional Neural Network architectures is presented in section three. Section four demonstrates the Experimental Study and results. Finally, the conclusion of this study is summarized in section five.

2 OCT-datasets

The datasets under studying are real clinical datasets; they had been acquired during 10 years of practice at The Eye Center in Medical Center, University of Freiburg, during 2007 and 2018. They contain ophthalmological data for about 3600 patients. Each patient suffers from AMD or a related disease such as diabetic retinopathy or retinal vein occlusion. The data for each patient had been collected during a long-term application of Anti-VGEF-therapy, and it is unfiltered i.e. patients suffer from

other eye diseases too (e.g. glaucoma or cataract).

“Optical coherence tomography (OCT) is an optical analog of ultrasound imaging that uses low coherence interferometry to produce cross-sectional images of the retina. It captures optical scattering from the tissue to decode spatial details of tissue microstructures. It uses infrared light from a super-luminescent diode that is divided into two parts: one of which is reflected from a reference mirror and the other is scattered from the biological tissue. The two reflected beams of light are made to produce interference patterns to obtain the echo time delay and their amplitude information that makes up an A-Scan. A-Scans that are captured at adjacent retinal locations by transverse scanning mechanism are combined to produce a 2-dimensional image.” [18].

Age related Macular degeneration (AMD) is a general idiom that includes a group of eye diseases affecting the retina, causing progressive loss of the central vision. Despite this, it never causes overall blindness, it influences a person’s ability to read, drive, recognize faces, color, and contrast [18]. Figure 1.1a shows a healthy macula, the retinal pigment epithelium is a straight, smooth line. On the other hand, the presence of druses represents an optical marker for dry AMD, figure 1.1b. Typical signs refer to fluid AMD as the new but abnormal blood vessels grow (choroidal neovascularizations), figure 1.1c. The leakiness, which leads to aggregation of fluid i.e. intraretinal or subretinal edema, also leads to a scar (fibrosis) as in figure 1.1d and figure 1.1e respectively.

Visual Acuity (VA) is a general measure for the clearness of vision, especially form vision, which depends on the sharpness of the retinal focus within the eye, the sensitivity of the nervous elements, and the interpretative faculty of the brain. VA is the only measure of this kind that is available in our patient’s data set.

The study was conducted on three OCT-image classification problems.

- AMD-problem (discrimination problem): The system should be able to classify the two major classes of AMD, Dry Macular Degeneration (Dry AMD), Wet Macular Degeneration (Wet AMD).
- VA-problem (quantification problem): The model should be able to determine the visual acuity value in decimal scale [19] [20], from a given OCT image. The dataset consists of 7-class discretization of the Visual decimal scale (Visus): 0-zero, 1-ge-0-0125, 2-ge-0-

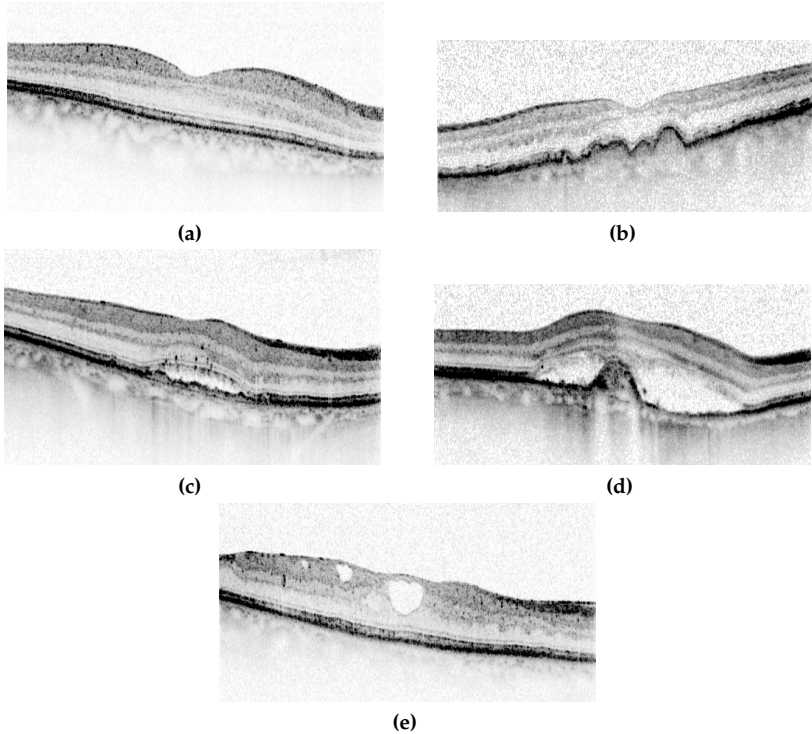


Figure 1.1: OCT-samples: (a) healthy macula, (b) dryAMD with druses, (c) sub-retinal fluid/oedema and choroidal neovascularizations, (d) subretinal fluid and a large scarf (fibrosis), (e) intraretinal fluid.

0600, 3-ge-0-1400, 4-ge-0-2800, 5-ge-0-4500, 6-ge-0-6600. Only the inner B-Scans of OCT-slices block are used (normally slice 8 to 18).

- VP-problem (outcome of the therapy quality): The model should be able to predict the development of VP through therapy in the following three categories, 1-decreasing: the visual acuity of the patient drops after a period of time, 2-stable: the visual acuity of the patient stabilizes after a long period. However, it needs consistent therapy. 3-increasing: the patient's visual acuity increased immediately from therapy.

Our dataset was split into three datasets, due to the previous training problems.

3 Experimental setup of the Convolutional Neural Network

A Convolutional Neural Network (CNN) consists of a series of layers organized in three dimensions, width, height and depth. An input layer, which receives the input images or characters, Convolutional layer, which represents the feature extraction stage by making a scalar product between local regions of the input and the convolutional filters (weights). Rectified Linear Unit (ReLU) layer is applied as an activation function to the output of the activation produced by the previous layer, to make the output non-linear. Pooling layer performs down sampling along the spatial dimensionality of the given input, further reducing the number of parameters within that activation. The last layer is the Fully-Connected layer will compute the class scores from the activations to be used in classification [21] [22].

In this work, we applied our two CNN architectures. The first one follows the traditional strategy, which chooses an increased number of filters as we go deeper in the CNN layers. It consists of three convolution layers with filter-weights of sizes $(3 \times 3 \times 96)$, $(4 \times 4 \times 128)$, $(5 \times 5 \times 256)$, respectively. Max-Pooling of size (3×3) is used as a down sampling process. Adam Optimizer [23] has been used with a learning rate between 0.00015 and 0.0002. Three dense layers are used with 1024 neurons in the first and second dense layers, and 512 neurons in the third dense layer.

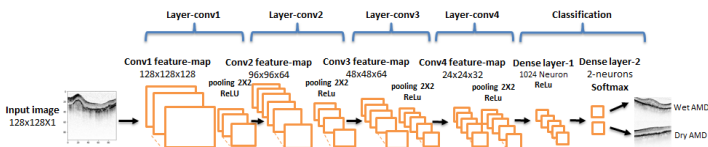


Figure 1.2: Our second CNN architecture used to classify OCT images.

Our second CNN structure consists of four convolutional layers, with filter-weights of sizes $(7 \times 7 \times 128)$, $(7 \times 7 \times 64)$, $(5 \times 5 \times 64)$, $(5 \times 5 \times 32)$ respec-

tively. We use a decreasing number of filters as long as going deeper through our layers, in contrast to the traditional way, which tends to choose an increased number of filters as long as going deeper through the layers. Each of convolutional layer followed by (2x2) Max-Pooling and ReLU activation. We use Adama Optimizer with learning rate 0.0001. In the first dense layer, 1024 neurons are used, followed by ReLU activation. The last dense layer (output layer) in both CNN architectures has a number of neurons, which matches the number of classes and is followed by softmax activation. We set the train batch size to 100. Figure 1.2 shows our CNN architectures to classify two classes (dry AMD, wet AMD) related to AMD problem. The input image shape is image-height x image-wide x number-of-channels. RGB images have three channels whereas grayscale images have one channel. In our study, we use grayscale images. The real OCT image size is 256x256 but we use 128x128 due to the memory problem. The architectures and the hyper parameters configurations were chosen by experiments, with the aim of designing an architecture, which results in better classification results.

Three Deep Learning Frameworks are used in this study to build the Convolution Neural Network. Theano, an open source Python library, developed by Montreal Institute for Learning Algorithms (MILA), University of Montreal in 2007 [14]. Tensorflow, an open source DL-framework developed by Google Brain Team in 2015 [24]. Keras, an open source Deep Learning library developed in Python on top of Theano, Tensorflow and CNTK frameworks in 2018. Keras wraps the numerical computing complexity of Theano and Tensorflow providing a compact API, in this work Keras (Tensorflow back-end) has been used [25].

Three Convolution Neural Networks are built and trained using three DL-Frameworks (Tensorflow, Theano, and Keras). Our first CNN architecture is built by Theano. The second CNN architecture is built using Tensorflow and Keras.

4 Experimental results

Our two CNN architectures are built, trained and evaluated using Theano, Tensorflow and Keras on three different OCT-training problems, to obtain the highest classification accuracy values for OCT im-

ages. Then we choose the best framework performance and CNN architecture based on the resulting classification accuracy values.

Implementing the experiments (training and evaluating our CNNs) has done for multiple runs using Nvidia GPU GeForce GTX 1080 Ti, in order to find the classification accuracy. The average value of accuracies after multiple runs on test data is shown in the table 1.1.

It is clear that CNN second architecture (Tensorflow-Model and Keras-Model) using less number of epochs achieved better than the tradition architecture (Theano-Model). We can notice the superiority of CNN-Keras model over Tensorflow-Models with the same architecture configurations. The classification accuracies resulted by CNN-Keras model after 60 epochs for three training problems (AMD , VP, VA) are 88.15%, 80.46%, 39.62%, respectively, figure 1.3.

The classification accuracy values achieved by Tensorflow-Model are 87.0%, 76.46%, and 35.0% for the classification problems (AMD , VP, VA) respectively, figure 1.3.

dataset	train	val	test	classes	Tensorflow-Model		Keras-Model		Theano-Model	
					epochs	Acc	epochs	Acc	epochs	Acc
AMD	22248	2472	6182	2	50 - 60	87.0%	50 -60	88.15%	70-80	87.30 %
VP	6271	696	777	3	50 - 60	76.46%	50 -60	80.46%	70 - 80	77.00 %
VA	50400	5600	14000	7	50 - 60	35.0 %	50 - 60	39.62%	70- 80	35.16 %

Table 1.1: The classification accuracy of the three CNN-models over the three training OCT problems sets.

CNN-Theano model required more number of epochs and it resulted the lowest classification accuracy values, table 1.1, 87.30%, 77.00%, 35.16% on three training problems (AMD, VP, VA), respectivel, figure 1.5.

5 Conclusion and Future Work

In this work, we introduce two CNN architectures implemented by three DL-frameworks (Tensorflow, Keras, and Theano). The goal of this study is to obtain the highest classification accuracy for three OCT training problems. The three CNN models were implemented by the three frameworks mentioned above, then the best framework was chosen.

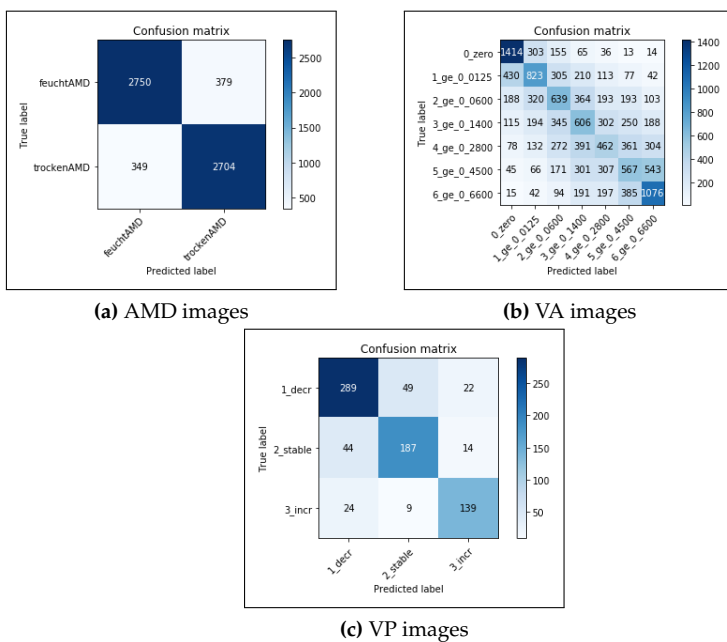
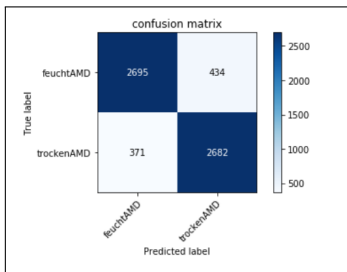
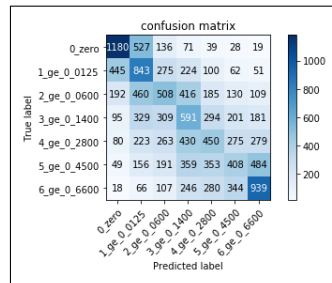


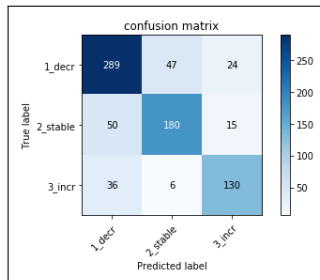
Figure 1.3: Confusion matrices of CNN-Keras model.



(a) AMD images



(b) VA images



(c) AP images

Figure 1.4: Confusion matrices of CNN-Tensorflow model.

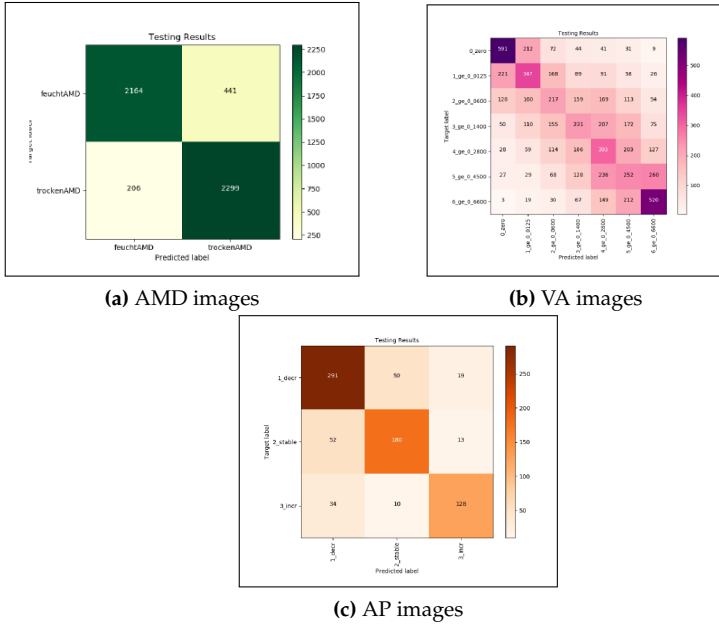


Figure 1.5: Confusion matrices of CNN-Theano model.

Performance was based on the resulting values according to classification accuracy. The results illustrated that the second CNN architecture built by Keras and Tensorflow achieved better results than the first architecture built by Theano. With the same CNN configuration and architecture, the results show that Keras-model achieved better than Tensorflow. In the future work we will try to improve our classification results and reduce training time by using transfer learning. Furthermore, we will try to use the real image size using appropriate hardware.

Acknowledgements

We like to acknowledge that the OCT image data that was used in this study, as well as the ophthalmological background, was provided by Prof. Andreas Stahl and the collaborators of the TOPOs project. TOPOs

(“Therapievorhersage durch Analyse von Patientendaten in der Ophthalmologie”) is a collaborative project that is funded by BMBF (“Bundesministerium für Bildung und Forschung”) (FKZ: 13GW0170B) from 03.2017-01.2021.

This work was funded by the European Social Fund (ESF).



References

1. A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep learning for computer vision: a brief review,” *Computational intelligence and neuroscience*, vol. 2018, 2018.
2. A. Diba, V. Sharma, A. Pazandeh, H. Pirsiavash, and L. Van Gool, “Weakly supervised cascaded convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 914–922.
3. J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
4. A. Toshev and C. Szegedy, “DeepPose: Human pose estimation via deep neural networks,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.
5. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, software available from tensorflow.org. [Online]. Available: <http://tensorflow.org/>
6. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” *arXiv preprint arXiv:1408.5093*, 2014.

7. R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A matlab-like environment for machine learning," in *BigLearn, NIPS Workshop*, 2011.
8. F. Seide and A. Agarwal, "Cntk: Microsoft's open-source deep-learning toolkit," in *KDD*, 2016.
9. S. Chetlur, C. Woolley, P. Vandermersch, J. Cohen, J. Tran, B. Catanzaro, and E. Shelhamer, "cudnn: Efficient primitives for deep learning," *arXiv preprint arXiv:1410.0759*, 2014.
10. S. NCCL, "Reference method for broth dilution antifungal susceptibility testing of filamentous fungi," *Approved standard, M38-A. National Committee on Clinical Laboratory Standards*, vol. 22, p. 16, 2002.
11. S. Bahrapour, N. Ramakrishnan, L. Schott, and M. Shah, "Comparative study of deep learning software frameworks," *arXiv preprint arXiv:1511.06435*, 2015.
12. G. Cohen, S. Afshar, J. Tapson, and A. van Schaik, "Emnist: an extension of mnist to handwritten letters," *arXiv preprint arXiv:1702.05373*, 2017.
13. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.
14. R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Balas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky *et al.*, "Theano: A python framework for fast computation of mathematical expressions," *arXiv preprint*, 2016.
15. A. Shatnawi, G. Al-Bdour, R. Al-Qurran, and M. Al-Ayyoub, "A comparative study of open source deep learning frameworks," in *Information and Communication Systems (ICICS), 2018 9th International Conference on*. IEEE, 2018, pp. 72–77.
16. V. Kovalev, A. Kalinovskiy, and S. Kovalev, "Deep learning with theano, torch, caffe, tensorflow, and deeplearning4j: Which one is the best in speed and accuracy?" 2016.
17. D. Team *et al.*, "Deeplearning4j: Open-source distributed deep learning for the jvm," *Apache Software Foundation License*, vol. 2, 2016.
18. M. Bhende, S. Shetty, M. K. Parthasarathy, and S. Ramya, "Optical coherence tomography: A guide to interpretation of common macular diseases," *Indian journal of ophthalmology*, vol. 66, no. 1, p. 20, 2018.
19. A. Colenbrander, "The historical evolution of visual acuity measurement," *Visual impairment research*, vol. 10, no. 2-3, pp. 57–66, 2008.
20. Wikipedia contributors, "Visual acuity — Wikipedia, the free encyclopedia," 2019, [Online; accessed 23-September-2019]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Visual_acuity&oldid=916573407

21. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
22. A. Mahendran and A. Vedaldi, "Visualizing deep convolutional neural networks using natural pre-images," *International Journal of Computer Vision*, vol. 120, no. 3, pp. 233–255, 2016.
23. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
24. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *Osd*, vol. 16, 2016, pp. 265–283.
25. F. Chollet *et al.*, "Keras," <https://keras.io>, 2015.

Size Doesn't Matter? Comparing Learning Curves for Data Subsets Using Different Classifiers

Richard Vogel, Marc Ritter, Kristina B. Helle, Kristan Schneider

Mittweida University of Applied Sciences, D-09648 Mittweida, Germany,
vogel2@hs-mittweida.de

Abstract Collecting and annotating training and test data, as needed for supervised learning, often is time consuming and costly as it may require expert knowledge. For study design, reliable sample size determination to achieve the desired performance goal is crucial to avoid unnecessary efforts and costs. Here, we advocate a graphical heuristic based on inspecting learning curves, obtained from training data sub-sampling, to determine their predictive power for sample size estimation. Namely, if learning curves, obtained from increasingly large sub-samples, appear to converge, they can be employed to reliably estimate the required amount of training data, imposed by the performance goal. Additionally, this approach is applicable to evaluate the appropriateness of the underlying machine learning pipeline. As examples, we apply the method to cancer classification based on RNA expression levels and handwritten digit recognition.

Keywords Sample Size Estimation, Classification, Classifier Significance, Learning Curves, Sub-sampling, Gene Expression, Cancer, Image Recognition

1 Introduction

Classification problems, where data X belongs to a set of finitely many classes (or targets, labels, etc.) \mathcal{C} , are the most common in machine learning. In multi-class classification, to a data point x we typically assign the

class $\hat{y} \in \mathcal{C}$ that maximizes its conditional probability

$$\hat{y} = \operatorname{argmax}\{P(Y = y|X = x)|y \in \mathcal{C}\},$$

where X denotes the random vector and Y the corresponding labels. Here, $P(Y = y|X = x)$ is modeled as a function $f(x, y)$ (classification algorithm). It is based on a general model (e.g., Decision Trees) and the parameters are determined by training it first on a set of pairs of data and their correct class labels $\{(x_1, y_1), \dots, (x_n, y_n)\} =: \mathcal{L}_n$ (training data).

Unfortunately, data acquisition and labelling is often costly, limiting the availability of large training data sets. Additional constraints on data collection might apply to particular applications (e.g., ethical restrictions in life- and medical sciences [1,2]).

The aim is to train the algorithm such that its error is limited by a predefined level e^* . Here we present an approach to estimate the required amount of training data: Given training data \mathcal{L}_n of size n archiving performance e_n , the required sample size n' yielding maximal error $e_{n'} < e^*$ can be estimated.

Learning curves A learning curve [3] plots the performance of a model as a function e of the training data sample size n . For different sizes n the model f is trained on a dataset of n samples \mathcal{L}_n respectively, to determine a classification function f_n . This function is tested (on a new test dataset \mathcal{L}') to quantify its performance.

Here performance is measured by the prediction error rate, the proportion of misclassifications of f_n :

$$e_n = \frac{|\{(x_i, y_i) \in \mathcal{L}' : f_n(x_i) \neq y_i\}|}{|\mathcal{L}'|}.$$

LCs based on the prediction error commonly follow an inverse power law [3]

$$e(z) = az^{-\alpha} + b \text{ subject to } a, \alpha, b \in \mathbb{R}_{\geq 0}, \quad (2.1)$$

where, z is the sample size, a the learning rate, α the exponential decay, and b the bias or Bayes error. The functional choice (2.1) is well motivated by the learning behavior of humans [2] and animals [4].

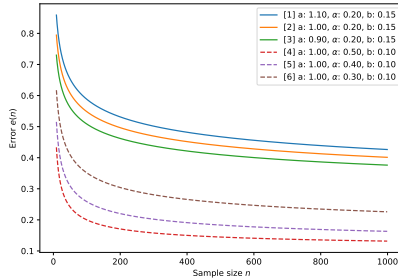


Figure 2.1: Demonstration of the parameter’s influence on learning curves. Curves 1-3: different learning rates (α) — the lower, the better the performance for all sample sizes; curves 3-6: different decay rates (a) — greater values result in a steeper curve that converges faster to b .

To determine the parameters of the LC (2.1) for a model f , the model’s performance is first estimated for different sample sizes n_1, \dots, n_k based on T training data sets of each size n_i , respectively, yielding performance estimates $e_{n_i,t}$ ($i = 1, \dots, k; t = 1, \dots, T$). Then the curve of the inverse power law is fitted to these points, for example by gradient descent.

The fitted LC’s parameters $\hat{\alpha}$, \hat{a} , and \hat{b} depend on the combination of classifier and training data (cf. Fig. 2.1):

- The bias b estimates the asymptotic error rate, the remaining error even if full information ($n \rightarrow \infty$) is available: performance cannot be better than this limit, no matter the size of the training data. Large b indicates that the model does not explain much of the functional relationship between input X and output Y . A reason could be independence of X and Y , that does not allow for a meaningful model. However, it is also possible that an existing dependence is not reflected well by the model type, the training process has flaws, or the data preparation (e.g., normalization, dimension reduction) is unsuitable.
- Large learning rate α results in a steeper LC, the information gain per training sample is higher than for small α .
- Larger exponential decay a reflects a higher demand for samples

needed to reach the performance goal, indicating high complexity of data in the underlying problem (e.g., existence of many data clusters).

Fitting the LC by a power law according to (2.1) allows to extrapolate the model's performance beyond the training data size n [5–8]. Based on the parameter estimates the required sample size n' to achieve a desired performance $e^* > b$ can be estimated as follows:

$$n' \approx e^{\frac{\ln \hat{a} - \ln(p^* - \hat{b})}{\hat{\alpha}}} = \left(\frac{\hat{a}}{p^* - \hat{b}} \right)^{\frac{1}{\hat{\alpha}}}. \quad (2.2)$$

However, this estimate is only justified if the estimated LC is accurate, which is typically unclear as the maximal size of available training data n may be much smaller than n' . Here, we inspect the behavior of LCs for different sample sizes n . In particular, we argue when they can be used to reliably estimate n' according to (2.2).

2 Approach

In practice, usually one training dataset \mathcal{L}_n of a size n is available. The learning curve $\hat{e}_n(z)$ is estimated by fitting it to error rates $e_{n_i,t}$ based on training data $\mathcal{L}_{n_i,t}$ which are sub-samples of the original data (i.e., $\mathcal{L}_{n_i,t} \subseteq \mathcal{L}_n$).

Our focus is to study the impact of the size n on the accuracy of the LC $\hat{e}_n(z)$. For this attempt, training datasets $\mathcal{L}_{n_{\text{sub}}}$ of different sizes n_{sub} are required; then for each, a LC $\hat{e}_{n_{\text{sub}}}$ is fitted based on error rates related to training on its subsets. The sets $\mathcal{L}_{n_{\text{sub}}}$ are sampled from the original training data $\mathcal{L}_{n_{\text{sub}}} \subset \mathcal{L}_n$ for different sizes $n_{\text{sub}} \leq n$.

The rationale now is: if the LCs $\hat{e}_{n_{\text{sub}}}(z)$ 'converge' for $n_{\text{sub}} \rightarrow n$, this indicates that for sufficiently large $n_{\text{sub}} \leq n$ the LCs $\hat{e}_{n_{\text{sub}}}(z)$ are accurate approximations for the true LC $e(z)$.

3 Methods and data and implementation

Calculating performances After training the model f to a sub-sample $\mathcal{L}_{n_i,t} \subset \mathcal{L}_{n_{\text{sub}}}$ (for $n_i < n_{\text{sub}}$) its error rate $e_{n_i,t}^{n_{\text{sub}}}$ is calculated on randomly drawn test data of fixed size (cf. figures) from $\mathcal{L}_{n_{\text{sub}}} \setminus \mathcal{L}_{n_i,t}$.

Curve fitting algorithm The generation of a single LC follows the basic algorithm described in Mukherjee [5] to fit $e_{n_{\text{sub}}}(z)$ according to (2.1) to the performances $e_{n_i, t}^{n_{\text{sub}}}$. Deviating from Mukherjee we did not perform the significance test described there. The additional computing time and complexity to fit a hyperparameter were not justified as it does not provide additional information relevant to the presented approach.

Machine learning algorithms As models two commonly used machine learning classifiers f were employed, Decision Trees (DT) [9] and Ridge Regression Classifier (RRC) [10], a form of Logistic Regression using ℓ_2 regularization.

Computational setup The experiments were carried out using Python 3.6. As toolbox scikit-learn [11] was adapted for the used machine learning models and matplotlib for visualizations [12]. Calculations were performed on a Microsoft[®] Surface Book 2[™] equipped with an Intel[®] Core[™] i7-8650U CPU. The CPU features 4 physical cores having a maximum frequency of 4.2 GHz.

Data The approach was tested on two datasets of contrasting properties. **RNA expression levels** [13] (5 classes) are a subset of the PAN-CAN data set [14], consisting of 20531 RNA-Seq gene expression levels (measured by the illumina[®] HiSeq platform) of 801 randomly selected patients (samples) having five types of cancer. **Handwritten digits (MNIST)** (10 classes) are a collection of 80,000 images, each as a 28×28 pixels (amounting to 784 features) picture on a gray-scale from 0 to 256 [15].

4 Results

RNA expression levels Learning curves based on subsets of the full dataset (Figure 2.2, top row) show that both models learn the data reasonably well; however, RRC clearly outperforms DT on average. This is not surprising as in this high dimensional parameter space DTs might suffer from the curse of dimensionality [16], whereas RRC is extremely

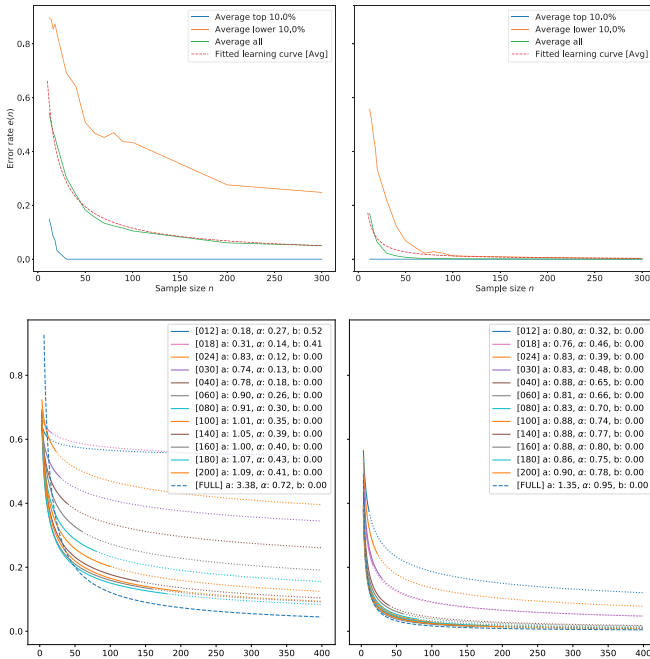


Figure 2.2: Learning curves for RNA expression level data for two classification algorithms (left: DT; right: RRC with $\alpha=1.0$); error rate e based on test data of size 5. Top: LC for full dataset, based on $T = 1000$ sub-samples of 18 sizes $1 \leq n_i \leq 600$, respectively (best/worst 10% and average per sample size shown). Computation time (using 4 cores): 0.4h for RRC, 1.4h for DT. Bottom: LCs fitted to sub-samples of increasing sizes $[n_{\text{sub}}]$, each based on $T = 1000$ samples of size $n_i = 2, 4, \dots, n_{\text{sub}} - 5$, respectively (continuous: interpolation, dotted: extrapolation) and to full dataset (dashed line).

flexible as the dimension of its parameters space (neglecting the intercept) coincides with that of the feature space. Additionally, it is a shrinkage method that controls for the dimensionality of the feature space. Although under best circumstances 6 – 7 samples may yield errors close to 0 (notably this could be a sampling artifact), a much larger training data set should be chosen to obtain reasonably small errors on average.

The LC estimates $\hat{e}_{n_{\text{sub}}}(z)$ (cf. Section 2) of both models indicate a Bayes error of $\hat{b} = 0$ for big n_{sub} , but else have a very different character-

istic (Figure 2.2, bottom row). For increasing n_{sub} the $\hat{\alpha}$ values increase generally. Beginning at $n_{\text{sub}} \geq 80$ the curves exhibit converging behavior for both classifiers. Particularly, $\hat{\alpha}$ stagnates with higher n_{sub} . The convergence is more pronounced for the better-performing RRC.

Taking into account this convergence, estimates of n' based on the estimated curves $\hat{e}_{n_{\text{sub}}}(z)$ according to (2.2) are reliable for $n_{\text{sub}} \geq 80$. To achieve an error rate of $e^* = 0.05$ the learner RCC requires training data of size $n' = 58$ (Table 2.1, left). Basing the estimate on smaller sub-samples $n_{\text{sub}} < 80$ it becomes too conservative. The worse performance of DT results in a larger required sample size.

Handwritten digits (MNIST) For this dataset the learning curves based on the whole dataset are similar for DT and RCC (Figure 2.3, top row). The LC estimates $\hat{e}_{n_{\text{sub}}}(z)$ tend to show lower error rates for RRCs than for DTs (Figure 2.3, bottom row). However, the LCs seem to converge better for DT than for RRC, indicating that inference of required sample size is more reliable for DT. For this model to achieve an error rate of $e^* = 0.15$ a sample of about $n' = 50,000$ images is estimated from $n_{\text{sub}} \geq 2000$ (Table 2.1, right). In general, the lack of converging behavior of LCs for both classifiers suggest that both methods are not optimal for this kind of application if the size of training data is less than 10,000.

5 Discussion

The two examples of training models to RNA expression levels and handwritten digits (MNIST) illustrate the potential power of engaging in sample size determination by fitted LCs based on sub-samples of the training data. Although we rely on visual inspection of LCs, the approach can be quantified by defining distance metrics, (e.g., Kolmogorov or uniform distance), which goes beyond the scope of this work. Furthermore, the procedure can be adapted to facilitate optimization of data collection and annotation efforts by employing adaptive sampling designs. This will be subject to future work.

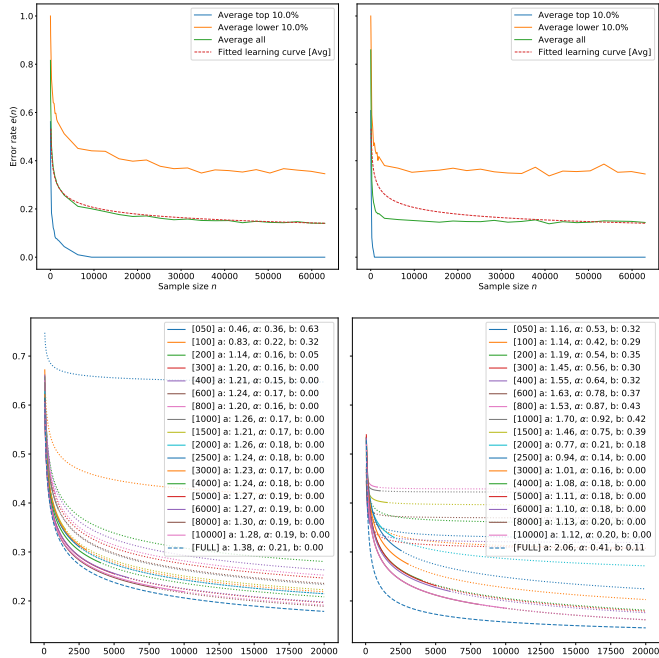


Figure 2.3: Learning curves for MNIST handwritten digits data for two classification algorithms (left: DT; right: RRC with $\alpha=1000$); error rate e based on test data of size 10. Top: LC for full dataset, based on $T = 1000$ sub-samples of 31 sizes $10 \leq n_i \leq 62,990$, respectively (best/worst 10% and average per sample size shown). Computation time (on 4 threads at 100% load): 7.5h for RRC, 26h for DT. Bottom: LCs fitted to sub-samples of increasing sizes $[n_{\text{sub}}]$, each based on $T = 600$ samples in 20 uniformly spaced steps $10 \leq n_i \leq n_{\text{sub}} - 10$, respectively (continuous: interpolation, dotted: extrapolation) and to full dataset (dashed line).

Required sample size n' to achieve e^* — according to (2.2)					
RNA Expression Levels			MNIST Handwritten Digits		
$e^* = 0.05$			$e^* = 0.15$		
n_{sub}	DT	RRC	n_{sub}	DT	RRC
12	-	10,566	200	1,615,651	-
18	-	498	300	169,729	-
24	$29,7 * 10^8$	1,644	400	280,126	-
30	$12,9 * 10^8$	434	600	102,313	-
40	$10,2 * 10^5$	87	800	202,813	-
60	96,162	76	1000	88,907	-
80	22,775	58	1500	78,970	-
100	5,352	50	2000	50,382	-
140	2,496	43	2500	54,857	477,268
160	1,837	37	3000	65,394	123,198
180	1,219	46	4000	46,754	41,066
200	1,665	40	5000	29,916	33,416
			6000	31,785	31,626
			8000	22,395	17,470
			10000	25,222	18,349
FULL	213	31		13,534	5,097

Table 2.1: Estimated sample size n' to reach target error rate e^* for two datasets (left: RNA expression levels; right: handwritten digits (MNIST)) and two learning algorithms (DT, RRC). Based on learning curves fitted to samples of different maximal size n_{sub} .

Acknowledgements

This study was carried out in the group of junior scientists "Agile Publika" which is funded by the European Social Fund (ESF) and the Free State of Saxony.



Diese Maßnahme wird mitfinanziert durch Steuermittel auf Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

References

1. S. Festing and R. Wilkinson, "The ethics of animal research. Talking point on the use of animals in scientific research," *EMBO reports*, vol. 8, no. 6, pp. 526–530, 2007.
2. M. B. Kapp, "Ethical and legal issues in research involving human subjects: do you want a piece of me?" *Journal of clinical pathology*, vol. 59, no. 4, pp. 335–339, 2006.
3. C. Cortes, L. D. Jackel, S. A. Solla, V. Vapnik, and J. S. Denker, "Learning curves: Asymptotic values and rate of convergence," in *Advances in Neural Information Processing Systems 6*, J. D. Cowan, G. Tesauro, and J. Alsppector, Eds. Morgan-Kaufmann, 1994, pp. 327–334. [Online]. Available: <http://papers.nips.cc/paper/803-learning-curves-asymptotic-values-and-rate-of-convergence.pdf>
4. J. R. Anderson, *The architecture of cognition*, 3rd ed., ser. Cognitive science series. Cambridge, Mass.: Harvard Univ.Pr, 1983, vol. 5.
5. S. Mukherjee, P. Tamayo, S. Rogers, R. Rifkin, A. Engle, C. Campbell, T. R. Golub, and J. P. Mesirov, "Estimating dataset size requirements for classifying DNA microarray data," *Journal of computational biology : a journal of computational molecular cell biology*, vol. 10, no. 2, pp. 119–142, 2003.
6. C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft, and J. Popp, "Sample size planning for classification models," *Analytica Chimica Acta*, vol. 760, pp. 25–33, 2013.
7. R. L. Figueroa, Q. Zeng-Treitler, S. Kandula, and L. H. Ngo, "Predicting sample size required for classification performance," *BMC medical informatics and decision making*, vol. 12, p. 8, 2012.

8. N. Ali, S. Girmus, P. Rösch, J. Popp, and T. Bocklitz, "Sample-size planning for multivariate data: A raman-spectroscopy-based example," *Analytical chemistry*, vol. 90, no. 21, pp. 12 485–12 492, 2018.
9. J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
10. R. de Vlaming and P. J. F. Groenen, "The current and future use of ridge regression for prediction in quantitative genetics," *BioMed research international*, vol. 2015, p. 143712, 2015.
11. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
12. J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing In Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
13. S. Fiorini. Gene expression cancer RNA-seq data set. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/gene+expression+cancer+RNA-Seq>
14. J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. M. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, and J. M. Stuart, "The cancer genome atlas pan-cancer analysis project," *Nature genetics*, vol. 45, no. 10, pp. 1113–1120, 2013.
15. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
16. R. Bellman, *Adaptive Control Processes: A Guided Tour*, ser. Princeton legacy library. Princeton University Press, 1961.

Hexagonale Bildverarbeitung im Kontext maschineller Lernverfahren

Konzeption eines biologisch inspirierten hexagonalen Deep Learning Frameworks

Tobias Schlosser, Michael Friedrich und Danny Kowerko

Juniorprofessur Media Computing, Technische Universität Chemnitz,
D-09107 Chemnitz, tobias.schlosser@informatik.tu-chemnitz.de

Zusammenfassung Inspiriert durch das visuelle Wahrnehmungssystem des Menschen befasst sich die hexagonale Bildverarbeitung im Kontext maschineller Lernverfahren mit der Entwicklung bildverarbeitender Systeme, die nach biologischem Vorbild die Vorteile evolutionär begründeter Strukturen in sich vereinen. Während klassische bildverarbeitende Systeme nach aktuellem Stand der Technik von Aufnahme- und Ausgabegeräten fast ausschließlich quadratisch gerasterte Verfahren zur Anwendung bringen, bieten ihre hexagonalen Pendanten eine Reihe entscheidender Vorteile, von denen Forscher sowie Anwender profitieren können. Diese Arbeit beschäftigt sich als ein erster anwendungsorientierter Ansatz mit der Synthese des hierfür konzipierten Frameworks, genannt Hexnet, den Verarbeitungsschritten der hexagonalen Bildtransformation und abhängigen Verfahren. Wie die Ergebnisse der erstellten Testumgebung zeigen, übertrifft das realisierte Framework aktuelle Ansätze hexagonaler bildverarbeitender Systeme, während künstliche neuronale Netze von der implementierten hexagonalen Architektur profitieren können.

Schlagwörter Pattern and Image Recognition, Deep Learning, Convolutional Neural Networks, Hexagonal Image Processing, Hexagonal Lattice, Hexagonal Sampling

1 Motivation

Mit den Entwicklungen der letzten Jahre kommen maschinellen Lernverfahren in Form künstlicher neuronaler Netze ein immer größerer Stellenwert zu. In diesem Kontext stellen Convolutional Neural Networks (CNN) zur Erkennung und Klassifikation von Objekten einen der Ansätze dar, die sich im Fokus aktueller Forschung des Deep Learnings wiederfinden. Um den stetig zunehmenden Anforderungen komplexerer Problemstellungen entsprechen zu können, werden indessen nach *Krizhevsky et al.* [1] größere Datensätze erstellt und neuartige Modelle und Verfahren trainiert.

Während die Struktur und Funktionsweise künstlicher neuronaler Netze durch biologische Prozesse inspiriert ist, werden diese aufgrund des aktuellen Standes der Technik von Aufnahme- und Ausgabegeräten insbesondere durch ihre zugrundeliegende Struktur limitiert. So kommen größtenteils quadratische Strukturen zum Einsatz, wodurch anschließende bildverarbeitende Systeme maßgeblich eingeschränkt werden.

Im Vergleich hierzu weist das visuelle Wahrnehmungssystem des Menschen insbesondere durch das menschliche Auge eine alternative, evolutionär begründete Struktur auf. So besitzt die Netzhaut des menschlichen Auges, auch Retina genannt, nach *Curcio et al.* [2] eine stark ausgeprägte hexagonale Anordnung der Sinneszellen (Abbildung 3.1), wodurch nach *Middleton und Sivaswamy* [3] die Verarbeitung von auftretenden Signalen deutlich effizienter erfolgen kann. Über die Schichten der Retina erfolgt die Reduktion der aufgenommenen Informationen auf die Nervenfasern, die über den Sehnerv Afferenzen in das

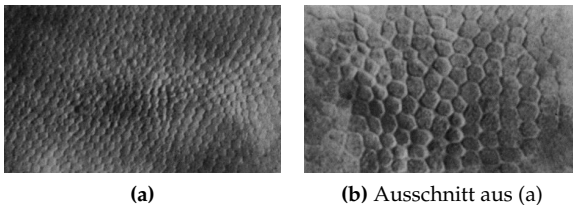


Abbildung 3.1: Ausschnitt der Netzhaut des menschlichen Auges (reprinted from *Curcio et al.* [2], Copyright 1987 AAAS)

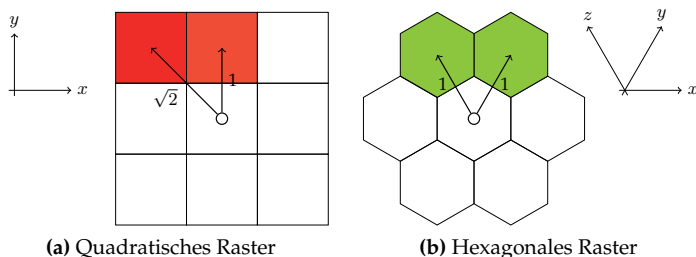


Abbildung 3.2: Quadratisches und hexagonales Rasterformat im Vergleich [7]

Gehirn besitzen. Eine Abbildung in den visuellen Kortex erfolgt nach *Hubel und Wiesel* [4] retinotop, sodass benachbarte Strukturen der Retina bildgetreu erhalten und durch die folgenden Bereiche des Gehirns verarbeitet werden.

Somit ergibt sich die Verwendung hexagonaler Strukturen als ein durch den Aufbau des visuellen Wahrnehmungssystems des Menschen evolutionär inspirierter Ansatz. Im Vergleich zu seinem quadratischen Pendant besitzt das hexagonale Rasterformat (Abbildung 3.2) eine Reihe entscheidender Vorteile, die unter anderem den Gegenstand aktueller Forschung aus den Bereichen der sensorbasierten Bildverarbeitung nach *Jung et al.* [5] und der Bildsynthese nach *Theussl et al.* [6], aber auch der Videokompression nach *Schlosser* [7] darstellen. Hierzu zählt die Homogenität des hexagonalen Rasterformates, wodurch die Eindeutigkeit der Nachbarschaft und eine erhöhte radiale Symmetrie gegeben sind. Zusammen mit der nach *Mersereau* [8] und *Nel* [9] um ca. 13,4 % erhöhten Abbildungseffizienz, die eine Speicherung größerer Mengen an Informationen auf der gleichen Anzahl Bildpunkte ermöglicht, resultieren nach *Golay* [10] weiterhin geringere Rechenzeiten und eine erhöhte Effizienz beim Programmieren.

Um die Vorteile der Forschungsfelder der biologisch inspirierten hexagonalen Bildverarbeitung und der künstlicher neuronaler Netze vereinen zu können, befasst sich diese Arbeit mit der Konzeption eines ersten prototypischen hexagonalen Deep Learning Frameworks, das inspiriert durch das visuelle Wahrnehmungssystem des Menschen insbesondere der Erkennung und Klassifikation von Objekten dienen kann. Mit Fokus auf die Umsetzung des zugrundeliegenden hexagonalen

Rasterformates umfasst das konzipierte Framework, genannt Hexnet, zunächst sämtliche hierfür notwendigen Vorverarbeitungsschritte. Die im Weiteren implementierten Architekturen und Verfahren sowie die dazugehörigen Tests werden indessen der Forschungscommunity via Open Science über das Projekt von Hexnet¹ zur Verfügung gestellt.

2 Grundlagen

Während sich die im Kontext der hexagonalen Bildverarbeitung von *Burt* [11] und *Gibson und Lucas* [12] vorgestellte eindimensionale Spiral-Architektur-Adressierung (engl. Spiral Architecture Addressing, SAA) als bevorzugtes Adressierungsschema hexagonaler Rasterformate etablierte (Abbildung 3.2b), birgt diese nach *Schlosser* [7] trotz ihrer erhöhten Lokalität den entscheidenden Nachteil einer erhöhten Komplexität bei der Hin- und Rücktransformation mittels auf quadratischem Rasterformat basierender Eingaben.

Diese Arbeit verfolgt hingegen einen hybriden Ansatz, der die auf klassischem quadratischem Rasterformat basierende eindimensionale zeilenweise Adressierung mit SAA vereint. Durch eine Vereinigung der respektiven Schemata und der damit verbundenen erhöhten Lokalität sowie der nach *Schlosser et al.* [7, 13, 14] resultierenden Verringerung von Laufzeiten ergibt sich weiterhin die Möglichkeit der vollständigen Abbildung des auf quadratischem Rasterformat basierenden Originalbildes.

Die Konstruktion dieser erfolgt entlang der dargestellten Nummerierung, wobei ein hexagonaler Bildpunkt als Hexint und eine Menge von 7^o Hexints als Hexarray der Ordnung o bezeichnet werden ($o \in \mathbb{N}_{\geq 0}$). Entlang der dargestellten Nummerierung erfolgt auf Basis der eindimensionalen Spiralarchitektur die Konstruktion eines jeden Hexarrays der Ordnung o aus sieben Sub-Hexarrays der Ordnung $o - 1$ (c), deren Hexints sich einer eindeutigen Position im Hexarray zuordnen lassen.

Die im Kontext des maschinellen Lernens zur Verfügung stehenden klassischen, fast ausschließlich auf Basis des quadratischen Rasterformates zum Einsatz kommenden CNNs unterteilen sich nach *Lecun et al.* [15] in eine Menge alternierender Schichten, die bestehend

¹ <https://github.com/TSchlosser13/Hexnet>

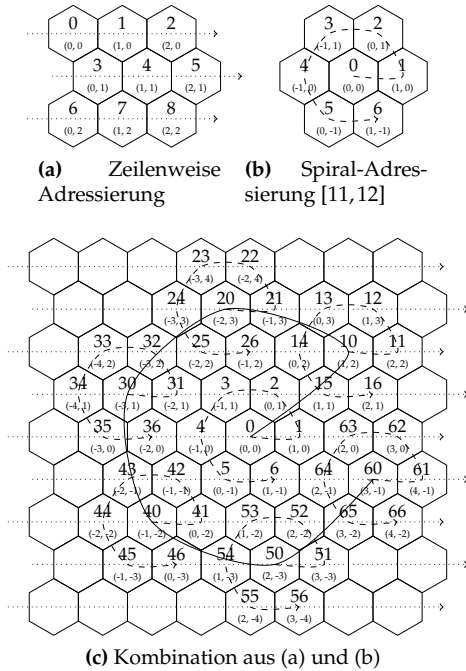


Abbildung 3.3: Konstruktion hexagonaler Adressierungsschemata

aus Convolutional (CL) und Pooling Layer (PL) entsprechend ihrer Tiefe der Filter-basierten Konvolution von Merkmalen dienen, durch ihre räumliche Dimensionalität abstrahieren und die extrahierten Merkmale in Form von Feature Maps (FM) darstellen. Die sich anschließenden Fully Connected Layer (FCL) dienen der Klassifikation der repräsentierten Merkmale. Um schließlich eine Zielfunktion eines gegebenen Optimierungsproblems, wie beispielsweise in Form der Erkennung und Klassifikation von Objekten, finden zu können, erfolgt eine Fehlerrückführung anhand der dem CNN zur Verfügung gestellten Trainingsdaten, gefolgt von einem Update des Netzes.

Das in Abbildung 3.4 dargestellte hexagonale Convolutional Neural Network, auch HCNN genannt, basiert auf dem in Abbildung 3.3c eingeführten Adressierungsschema für hexagonale Architekturen und ver-

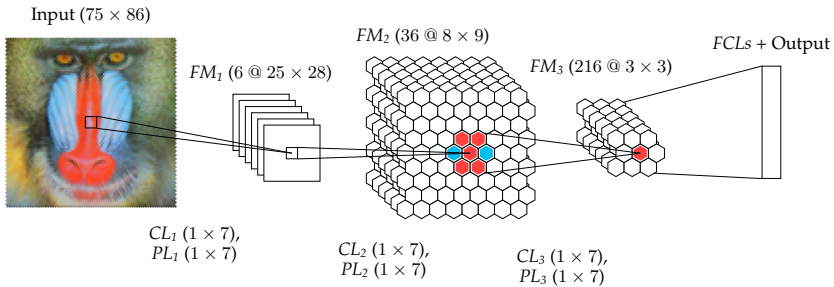


Abbildung 3.4: Beispielhafter Aufbau eines hexagonalen Convolutional Neural Networks

deutlicht das Vorgehen am Beispiel des Testbildes »Baboon« [16]. Die in einer jeden Schicht abgebildeten Ergebnisse enthalten für FM_1 bis FM_3 die Feature Maps der jeweiligen Konvolution mit anschließendem Pooling, wobei die zum Einsatz kommenden Filter durch die Unterstrukturen des hexagonalen Adressierungsschemas gegeben sind.

Nach *Cohen und Welling* [17] stellen sich für CNNs auf quadratischem Rasterformat eine Translationsäquivalanz sowie eine Translations- und Skalierungsinvarianz als entscheidende Merkmale heraus, während im Hexagonalen aufgrund der Unterstrukturen verschiedener Ordnungen weiterhin der Vorteil einer Rotationsinvarianz besteht, wodurch nach *Dieleman et al.* [18] insbesondere die Notwendigkeit von Data Augmentation maßgeblich reduziert werden kann. Zusammen mit der erhöhten axialen Symmetrie hexagonaler Darstellungen ergibt sich nach [17, 18] die Möglichkeit der Verwendung von Gruppen-Konvolutionen erhöhter Diversität, die fernerhin eine verringerte angulare Dependenz aufweisen und in erhöhten Konvergenzen und Konfidenzen resultieren können.

3 Implementierung

Die in Hexnet als initialer Datenreduktionsschritt implementierte hexagonale Bildtransformation basiert auf dem in Abbildung 3.3c dargestellten Adressierungsschema, wobei die zu erzeugenden Hexints entsprechend des gegebenen quadratischen Rasterformates der Eingabe zeilen-

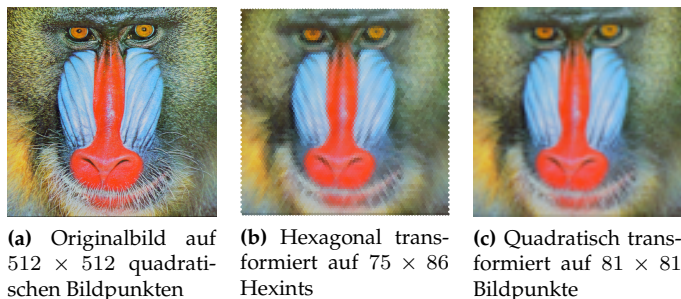


Abbildung 3.5: Testdurchlauf der hexagonalen Bildtransformation in Hexnet

weise interpoliert werden. Abbildung 3.5 zeigt einen Testdurchlauf der hexagonalen Bildtransformation (b) im Vergleich mit dem durch das Softwarepaket ImageMagick² bereitgestellten Pendant für quadratische Rasterformate (c).

Um eine adäquate Visualisierung des hexagonalen Rasterformates durch Hexnet gewährleisten zu können, erfolgt eine durch General-Purpose Graphics Processing Units (GPGPU) beschleunigte Darstellung, wodurch die Funktionalität der Translation und der Skalierung gegeben sind. So erfolgt entsprechend des aktuellen Detailgrades eine Shader-basierte Approximation eines jeden Hexints.

4 Evaluation

Um die Transformationseffizienz der hexagonalen Bildtransformation quantifizieren und evaluieren zu können, erfolgt zunächst eine Gegenüberstellung beider Rasterformate am Beispiel der »USC-SIPI Image Database« [16]. Ein Vergleich beider Rasterformate wird ermöglicht, indem die durch eine gegenseitige Abbildung beider Rasterformate resultierenden Teilflächen der Transformationsergebnisse in ein Verhältnis zum gegebenen Flächeninhalt mit der zum Einsatz kommenden Vergleichsmetrik gesetzt werden.

Im Vergleich (Abbildung 3.6) dargestellt sind für die im Weiteren als Transformationseffizienz T bezeichneten Differenzen der

² <https://www.imagemagick.org/Usage/filter/#triangle>

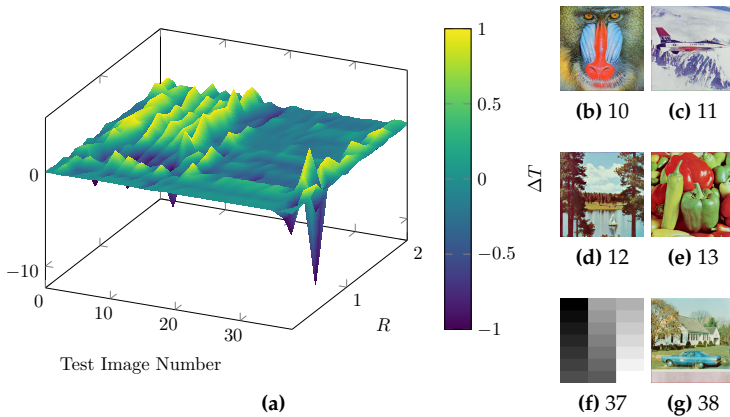


Abbildung 3.6: Quadratische und hexagonale Bildtransformation im Vergleich mit Ausschnitt des Testsets [16]

quadratischen (T_q) und hexagonalen Bildtransformation (T_h) die Transformationseffizienz-Differenzen ($\Delta T = T_h - T_q$) am Beispiel der Peak Signal-to-Noise Ratio, wobei im Hexagonalen mittels des Umkreisradius R eines jeden Hexints und im Quadratischen auf die hieraus resultierende Anzahl Hexints interpoliert wurde.³ So weisen insbesondere Testbilder, die sich nach *Field* [19] durch natürliche Formen charakterisieren (b–e), im Test bildinhaltsbedingt eine evident erhöhte Transformationseffizienz auf, wodurch hexagonale Strukturen effizienter erhalten werden konnten. Ein Gegenbeispiel hierfür stellen die Testbilder Nummer 37 (f) und 38 (g) dar, die sich aufgrund der Vielzahl auftretender senkrechter und waagerechter Strukturen a priori nur unzuverlässig hexagonal approximieren ließen.

Für einen Performanzvergleich bieten sich das nach *Middleton und Sivaswamy* [3] entwickelte Hexagonal Image Processing Framework HIP und das nach *Schlosser et al.* [7, 13, 14] auf Basis dessen implementierte Hexagonal Image Processing Framework CHIP als zwei der verfügbaren Frameworks der hexagonalen Bildverarbeitung an. Diese führen im Vergleich zu Hexnet ihre hexagonale Bildtransformation

³ <https://github.com/TSchlosser13/Hexnet/releases/tag/v0.1-rc0>

Ordnung	Hintransformation			Rücktransformation		
	5	6	7	5	6	7
HIP	6	30	180	3.600	18.000	30 [†]
CHIP	0,002	0,013	0,109	0,005	0,062	0,545
Hexnet	0,001	0,009	0,067	0,004	0,023	0,169

Tabelle 3.1: Performanzvergleich aktueller Frameworks der hexagonalen Bildverarbeitung (Angaben in Sekunden bzw. Stunden[†])

durch eine iterative Traversierung des aus Abbildung 3.3b resultierenden, durch SAA aufgespannten Septrees durch, wobei Nachbarschaftsbeziehungen über die Unterstrukturen hinweg in einem jeden Transformationsschritt erneut berechnet werden. Tabelle 3.1 zeigt eine Gegenüberstellung der hexagonalen Hin- und Rücktransformation für die Hexarrays der Ordnungen 5 bis 7 für HIP, CHIP und Hexnet.

Während das HIP-Framework einen ersten anwendungsorientierten Ansatz hexagonaler bildverarbeitender Frameworks darstellt, weist das CHIP-Framework als eine Weiterentwicklung dessen eine stark erhöhte Performanz auf. Die in Hexnet als initialer Datenreduktionsschritt basierend auf Abbildung 3.3c implementierte hexagonale Bildtransformation vereint die Vorteile der auf quadratischem Rasterformat zum Einsatz kommenden zeilenweisen Architektur mit der eindimensionalen Spiralarchitektur im Hexagonalen. So indizieren sich die Vorteile dieser durch die verringerten Laufzeiten, wobei Hexnet im Vergleich zu CHIP für die Ordnungen 5 bis 7 im Test eine um ca. Faktor 2, 1,4 und 1,63 (Hintransformation) sowie 1,25, 2,7 und 3,23 (Rücktransformation) verringerte Laufzeit aufweist.⁴

5 Zusammenfassung und Ausblick

Die implementierten und konzipierten Verfahren stellen die Grundlage eines ersten hexagonalen Frameworks dar, dass der Synthese der biologisch inspirierten hexagonalen Bildverarbeitung mit Ansätzen des maschinellen Lernens in Form künstlicher neuronaler Netze dient. Die in

⁴ CPU: »Intel Core i7-6500U @ 2.50GHz«, BogoMips: 5184, CPU-Load: ca. 99 % (ein Kern), VmPeak: ca. 814 MiB (Max. für Ordnung 7)

Hexnet als Basis der hexagonalen Bildtransformation realisierte Architektur vereint hierzu die Vorteile klassischer, auf quadratischem Rasterformat basierender Systeme und hexagonaler Adressierungsschemata in sich.

Wie die Ergebnisse der erstellten Testumgebung zeigen, erzielt die hexagonale Bildtransformation im Vergleich mit ihrem quadratischen Pendant ein überzeugendes Transformationsergebnis, wobei die in Hexnet umgesetzte Funktionalität nach aktuellem Stand der Technik von Aufnahme- und Ausgabegeräten die Realisierbarkeit der hexagonalen Bildverarbeitung im Kontext maschineller Lernverfahren aufzeigt. So erzielt die hexagonale Bildtransformation als notwendiger initialer Verarbeitungsschritt im Test eine bildinhaltsbedingt erhöhte Transformationseffizienz, während ein Vergleich mit aktuellen Frameworks der hexagonalen Bildverarbeitung insbesondere eine Verringerung der Laufzeit aufweist.

Anschließende Verfahren des maschinellen Lernens können von den vorverarbeiteten Bilddaten in ihrer hexagonalen Darstellung sowie von den Vorteilen einer erhöhten Transformationseffizienz profitieren. Somit ergeben sich für die in dieser Arbeit eingeführten hexagonalen Convolutional Neural Networks, auch HCNN genannt, die Vorteile der Translations- und Skalierungsinvarianz sowie die durch die eindimensionale Spiralarchitektur im Framework gegebene Rotationsinvarianz. Weiterhin bestehen die Vorteile der hexagonalen Darstellung, wodurch unter anderem eine erhöhte Symmetrie, wie beispielsweise bei dem Einsatz von Gruppen-Konvolutionen, gegeben ist. Zukünftige Arbeiten können auf diesem Framework aufbauen und das hexagonale Pendant zu klassischen künstlichen neuronalen Netzen darstellen.

Danksagung

Diese Arbeit wurde teilweise durch die Europäische Union und den Europäischen Sozialfonds für Deutschland gefördert.



Europäische Union

Europa fördert Sachsen.



Europäischer Sozialfonds



Diese Maßnahme wird mitfinanziert durch Steuermittel auf Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Literatur

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2012, pp. 1097–1105. [Online]. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks>
2. C. A. Curcio, K. R. Sloan, O. Packer, A. E. Hendrickson, and R. E. Kalina, "Distribution of Cones in Human and Monkey Retina: Individual Variability and Radial Asymmetry," *Science (New York, N.Y.)*, vol. 236, no. 4801, pp. 579–582, 5 1987. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/3576186>
3. L. Middleton and J. Sivaswamy, *Hexagonal Image Processing: A Practical Approach*. Springer, 2005. [Online]. Available: <https://dl.acm.org/citation.cfm?id=1076360>
4. D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *The Journal of Physiology*, vol. 195, no. 1, pp. 215–243, 3 1968. [Online]. Available: <http://doi.wiley.com/10.1113/jphysiol.1968.sp008455>
5. S. Jung, R. Thewes, T. Scheiter, K. Goser, and W. Weber, "A Low-Power and High-Performance CMOS Fingerprint Sensing and Encoding Architecture," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 7, pp. 978–984, 7 1999. [Online]. Available: <http://ieeexplore.ieee.org/document/772413/>
6. T. Theussl, T. Moller, and M. Groller, "Optimal Regular Volume Sampling," in *Proceedings Visualization, 2001. VIS '01*. IEEE, pp. 91–98. [Online]. Available: <http://ieeexplore.ieee.org/document/964498/>
7. T. Schlosser, *Entwurf und Implementierung eines CHIP-basierten Systems und Demonstrators zur Aufnahme, Speicherung und Visualisierung hexagonal gerasterter Bild- und Videodaten (MSc thesis)*, 2018.
8. R. Mersereau, "The Processing of Hexagonally Sampled Two-Dimensional Signals," *Proceedings of the IEEE*, vol. 67, no. 6, pp. 930–949, 1979. [Online]. Available: <http://ieeexplore.ieee.org/document/1455625/>
9. A. Nel, "Hexagonal image processing," in *COMSIG 1989 Proceedings: Southern African Conference on Communications and Signal Processing*. IEEE, 1989, pp. 109–113. [Online]. Available: <http://ieeexplore.ieee.org/document/129027/>
10. M. Golay, "Hexagonal Parallel Pattern Transformations," *IEEE Transactions on Computers*, vol. C-18, no. 8, pp. 733–740, 8 1969. [Online]. Available: <http://ieeexplore.ieee.org/document/1671349/>

11. P. J. Burt, "Tree and Pyramid Structures for Coding Hexagonally Sampled Binary Images," *Computer Graphics and Image Processing*, vol. 14, no. 3, pp. 271–280, 11 1980. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0146664X80900568>
12. L. Gibson and D. Lucas, "Vectorization of Raster images Using Hierarchical Methods," *Computer Graphics and Image Processing*, vol. 20, no. 1, pp. 82–89, 9 1982. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0146664X82900752>
13. T. Schlosser, *Entwurf und Implementierung von Optimierungs- und Funktionserweiterungen der hexagonalen Bildraasterung in der Videokompressionssoftware x264HMod (BSc thesis)*, 2016.
14. T. Schlosser, R. Manthey, and M. Ritter, "Entwurf und Implementierung von Optimierungs- und Funktionserweiterungen der hexagonalen Bildraasterung in der Videokompressionssoftware x264HMod," in *Studierendensymposium Informatik 2016 der TU Chemnitz*, 2016, pp. 63–74.
15. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. [Online]. Available: <http://ieeexplore.ieee.org/document/726791/>
16. A. G. Weber, "The USC-SIPI Image Database: Version 4," Tech. Rep., 1993. [Online]. Available: <http://sipi.usc.edu/reports/pdfs/Scanned/USC-SIPI-244.pdf>
17. T. S. Cohen and M. Welling, "Group Equivariant Convolutional Networks," in *International Conference on Machine Learning*, 2 2016, pp. 2990–2999. [Online]. Available: <http://proceedings.mlr.press/v48/>
18. S. Dieleman, J. De Fauw, and K. Kavukcuoglu, "Exploiting Cyclic Symmetry in Convolutional Neural Networks," in *International Conference on Machine Learning*, 2 2016, pp. 1889–1898. [Online]. Available: <http://proceedings.mlr.press/v48/>
19. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A*, vol. 4, no. 12, pp. 2379–2394, 12 1987. [Online]. Available: <https://www.osapublishing.org/abstract.cfm?URI=josaa-4-12-2379>

Generierung hexagonaler Bilddaten und Vergleich mit konventioneller Rasterung in der Bildverarbeitung

Generation of Hexagonal Image Data and Comparison With Conventional Grids in Image Processing

Robert Manthey

Juniorprofessur Media Computing, Technische Universität Chemnitz,
D-09107 Chemnitz, robert.manthey@informatik.tu-chemnitz.de

Zusammenfassung Augen und Kameras stellen als optische Wahrnehmungssysteme eine wichtige Informationsquelle für biologische und technische Systeme dar um z.B. die Essbarkeit von Früchten, die Entfernung zu Beutetieren, die Qualität von Bauteilen oder geografische Gegebenheiten zu beurteilen. Hierzu werden über größere Entfernungen Form, Struktur und bestimmte Eigenschaften der Oberfläche der Objekte durch Analyse der Farbe und Intensität des von der Oberfläche reflektierten sichtbaren Lichts, der abgegebenen Infrarotstrahlung oder auch deren räumliche und zeitliche Veränderung erfasst, Abbildung 4.2, und von nachfolgenden Verarbeitungs- und Analyseschritten entsprechend ausgewertet. Dabei kommt der zur technischen Weiterverarbeitung notwendigen Diskretisierung der Szenenprojektion als erstem Glied der Verarbeitungskette eine besondere Bedeutung zu, da die unterschiedliche Abbildungseffizienzen verschiedener Diskretisierungsraster mit einer Reduzierung der Menge an Informationen einhergehen [1]. Vorarbeiten zu möglichen Diskretisierungsrastern von [2], [3], [4] u.a. zeigten dass eine Verwendung von dreieckigen, rechteckigen und sechseckigen Grundstrukturen als Raster für digitalen Aufnahmesystemen geeignet sind und deutliche Unterschiede im Bereich der Abbildungseffizienz [4,5] aufweisen, allerdings entsprechende native Aufnahmesysteme für Drei- und Sechsecke fehlen. Im Gegensatz dazu weisen biologische, evolutionär über Jahrmlionen entwickelte und optimierte Aufnahmesysteme wie Facetten- oder Linsenaugen eine

hexagonale Rasterung auf, wie in Abbildung 4.1 zu sehen. Im Rahmen dieser Arbeit werden die verschiedenen Möglichkeiten zur Aufnahme hexagonal gerasterter Bilder untersucht und ihre Leistungsmerkmale bestimmt. Die Verfahren werden konsolidiert und Optimiert in einem einheitlichen Framework zusammengefasst.

Schlagwörter Rasterung, Hexagon, Bildgewinnung, Bilderzeugung, Bildverarbeitung

1 Einleitung

Für die Wahrnehmung der Umgebung, speziell über größere Entfernungen, werden in technischen und biologischen Systemen Kameras bzw. Augen verwendet. Mit diesen können Entfernungen und Oberflächenbeschaffenheiten abgeschätzt werden, wie beispielsweise bei rauem Sandstein oder glatter Keramik. Allerdings unterscheiden sich diese Systeme bezüglich der von ihnen verwendeten Sensoranordnung sowie deren Form. In Kameras werden nahezu ausschließlich rechteckige bzw. quadratische Sensoren verwendet, während evolutionär optimierte Augen meist sechseckige Formen aufweisen, wie in Abbildung 4.1 dargestellt. Untersuchungen dazu zeigen die Vorteile der Hexagonale, stellen aber immer wieder das Fehlen geeigneter Aufnahmesysteme als großes Hindernis für weitere und umfassendere Analysen dar [2–5]. In dieser Arbeit wird dieses Problem adressiert und ein möglicher Lösungsansatz vorgestellt.

2 Grundlagen

Für die Erkennung und Analyse von Objekten, Personen und Oberflächen der kontinuierlichen dreidimensionalen Welt wird die von diesen ausgehende bzw. reflektierte informationstragende Strahlung auf eine zweidimensionale Ebene projiziert und wie in Abbildung 4.2 dargestellt mit Hilfe eines Rasters auf dessen Elemente abgebildet. Durch diese Diskretisierung in Bezug auf räumliche und zeitliche Verteilung der Informationen ist eine nachfolgende Verarbeitung und Analyse realisierbar.

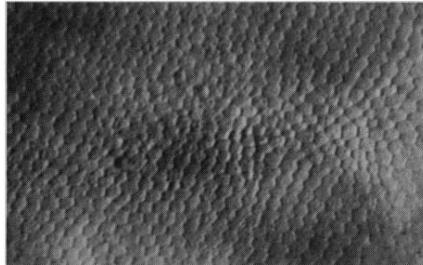


Abbildung 4.1: Detailaufnahme der Netzhautstruktur eines menschlichen Auges mit ihren meist sechseckigen Fotorezeptoren [6].

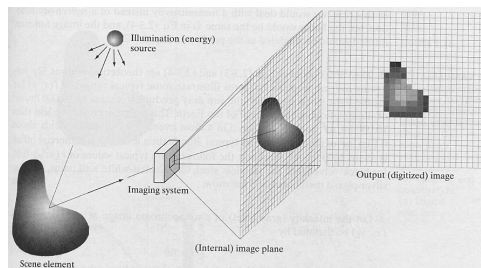


Abbildung 4.2: Schematischer Bildaufnahmeprozess [7, S.51] mit Beleuchtung, Objekt, Aufnahmesystem und diskretisiertem Bild mit Rechteckraster.

Jeder Teil einer Verarbeitungskette wie in Abbildung 4.3 reduziert allerdings auch die Informationsmenge bzw. führt charakteristische Artefakte und Fehler ein. Eine beträchtliche Minderung stellen dabei die durch die Rasterung verursachten Verluste dar, wodurch diesem Schritt eine beträchtliche Bedeutung zukommt und bereits von [2, 3] mögliche Kandidaten untersucht wurden. So reduziert ein auf gleichseitigen Dreiecken basierendes Raster die Informationsmenge auf ca. 60%, ein Quadratbasiertes auf ca. 79% und ein Hexagonbasiertes knapp 91%.

3 Implementierung

Um diese unterschiedlichen Raster untersuchen zu können und das Problem des Aufnahmesystems zu beheben wurde die Raytracing-

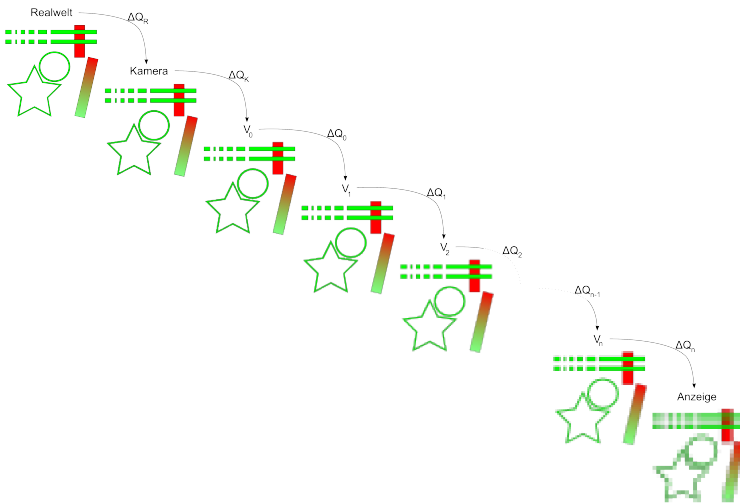


Abbildung 4.3: Beispiel einer Verarbeitungskette mit den Realweltobjekten, der Transformation auf die Bildebene der Kamera und den durch nachfolgende Verarbeitungsschritte V_0, V_1, \dots, V_n und die Darstellung erlittenen Informationsverlusten $\Delta Q_R, \Delta Q_K, \Delta Q_0, \dots, \Delta Q_n$.

Software POV-Ray¹ benutzt mit welcher der Strahlenverlauf simulierbar und realistische Bilddaten erstellbar sind. Aufgrund der Fähigkeit dieser Software das bei der Kamera zum Einsatz kommende Raster mittels eines Dreiecksnetzes wahlfrei definieren zu können sind somit auch hexagonbasierte Rasterungen erzeugbar, wie in Abbildung 4.4. Durch die abstrakte mathematische Definition der Szene sind exakte Größen und Formen erstellbar.

4 Ergebnisse

Für die Untersuchung der Eigenschaften der verschiedenen Rasterungen wurden verschiedene Szenen mit bekannten Test- und Demonstrationsobjekten aus dem Bereich der Bildverarbeitung erstellt sowie

¹ <http://www.povray.org>

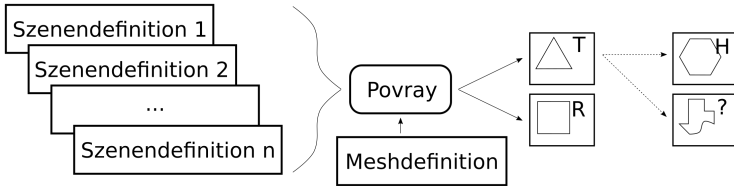


Abbildung 4.4: Schematischer Workflow zur Generierung von Bilddaten mit dreieckbasiertem, rechteckbasiertem, hexagonbasiertem oder beliebigem Raster aus den gegebenen Szenendefinitionen. [8]

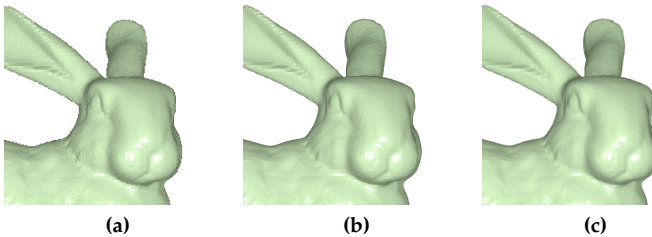


Abbildung 4.5: Kopfbereich des Stanford Bunny² gerastert mit (a) 256×256 rechteckigen, (b) 476×824 dreieckigen und (c) 238×275 sechseckigen Pixeln. Deutlich treten die Treppenstufenartefakte im Rechteckraster an den Kanten und den Flächen der Schulter hervor. Letztere sind auch im Dreiecksraster noch auffälliger als im Sechseckraster. [8]

weitere vorhandenen Testszenen [9] genutzt, wie beispielhaft in Abbildung 4.5 und 4.6 dargestellt.

5 Zusammenfassung

Die Abbildung der Welt in eine verarbeitbare Repräsentation stellt eine beträchtliche Reduktion von Informationen dar, deren Umfang durch die Wahl geeigneter Raster beeinflusst werden kann. Hierzu

² <http://sipi.usc.edu/database/download.php?vol=misc&img=4.2.04>

³ <http://graphics.stanford.edu/data/3Dscanrep/#bunny>

stellte das Fehlen eines Aufnahmesystems für nichtrechteckige Raster ein erhebliches Problem für die bisherigen Forschungen und Analysen dar, welches durch die vorgestellte Erstellung eines Aufnahmesystems für virtuelle Szenen behoben wurde. Dies ermöglichte erste direkte Vergleiche, zeigt die Vorteile hexagonaler Rasterung und stellt die Grundlage für komplexere Untersuchungen dar. Somit können Qualitätsbetrachtungen und auch das Auftreten optischer Artefakte, aber auch Algorithmenvergleiche durchgeführt werden.



Abbildung 4.6: Testbild Lena³ gerastert mit (a) 64×64 rechteckigen und (b) 59×68 sechseckigen Pixeln bei gleicher Sensorfläche. Runde Kanten weisen im Rechteckraster deutlich sichtbare Treppenstufenartefakte und mehr Rauschen auf. [8]

Danksagung

Diese Arbeit wurde vom BMBF im Rahmen des Unternehmens Region geförderten InnoProfile-Transfer-Initiative localizeIT (Förderkennzeichen: 03IPT608) durchgeführt.

Literatur

1. X. He and W. Jia, "Hexagonal structure for intelligent vision," in *International Conference on Information and Communication Technologies*. IEEE, August 2005, pp. 52–64.
2. M. J. Golay, "Hexagonal parallel pattern transformations," *Transactions on Computers*, vol. C-18, no. 8, pp. 733–740, August 1969.

3. E. Deutsch, "Thinning algorithms on rectangular, hexagonal, and triangular arrays," *Communications of the ACM*, vol. 15, no. 9, pp. 827–837, September 1972. [Online]. Available: <http://doi.acm.org/10.1145/361573.361583>
4. C. Wüthrich and P. Stucki, "An algorithmic comparison between square- and hexagonal-based grids," *Graphical Models and Image Processing*, vol. 53, no. 4, pp. 324–339, July 1991.
5. A. L. Nel, "Hexagonal image processing," in *Southern African Conference on Communications and Signal Processing*. IEEE, June 1989, pp. 109–113.
6. C. A. Curcio, J. Kenneth R. Sloan, O. Packer, A. E. Hendrickson, and R. E. Kalina, "Distribution of cones in human and monkey retina: individual variability and radial asymmetry." *Science*, vol. 236, no. 4801, pp. 579–582, May 1987. [Online]. Available: <http://www.jstor.org/stable/1699166>
7. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Pearson, 2008.
8. R. Manthey and D. Kowerko, "Visuelle szenendiskretisierung mittels wahl-freien bildrastern am beispiel von drei-, vier- und sechseckpixeln," in *Forum Bildverarbeitung 2018*, T. Längle, F. Puente León, and M. Heizmann, Eds. KIT Scientific Publishing, Karlsruhe, 2018, p. 9.
9. R. Manthey, M. Ritter, M. Heinzig, and D. Kowerko, "An exploratory comparison of the visual quality of virtual reality systems based on device-independent testsets," 05 2017, pp. 130–140.

Generierung eines Aktivitätsklassifikators für den Einsatz in mobilen kontextsensitiven Anwendungen

Ein Workflow zur Aktivitätserkennung

Tony Rolletschke

Hochschule Mittweida, D-09648 Mittweida, rolletsc@hs-mittweida.de

Zusammenfassung In den letzten Jahren wurden verschiedene Methoden zur automatisierten Erkennung von menschlichen Aktivitäten durch maschinelle Lernverfahren auf Grundlage großer Datenmengen untersucht. Durch die stetig steigende Hardwareleistung mobiler Endgeräte wie Smartphones und Wearables und der kontinuierlichen Weiterentwicklung der verwendeten Verfahren waren dabei fortlaufend höhere Erkennungsraten möglich. Diese jüngsten Fortschritte führten gleichzeitig zu einer zunehmenden Nachfrage nach intelligenten und kontextsensitiven mobilen Anwendungen. Eine größere Herausforderung bleibt jedoch weiterhin die Generierung valider Trainingsdatensätze (engl. ground truth information) mit einer angemessenen Güte. Zudem gibt es aktuell kein standardisiertes Verfahren, um einen Aktivitätsklassifizierer für den Einsatz in individuellen Anwendungsbereichen zu generieren. Der hier vorgestellte Workflow beschreibt die Erkennung von Aktivitäten mittels Sensoren mobiler Endgeräte. Hierzu wird durch die Aufzeichnung und Annotation von Sensor- und Videodaten ein *Ground Truth* -Datensatz erzeugt. Anschließend folgt die Weiterverarbeitung des Datensatzes in einem Framework für maschinelles Lernen, um einen Lerntransfer durchzuführen. Das daraus resultierende Modell ermöglicht die Echtzeitklassifikation auf mobilen Endgeräten. Als Datenquelle dient eine aktuelle Studie, deren gegenwärtiger Fortschritt abschließend vorgestellt wird.

Schlagwörter Mobile Sensorik, Aktivitätserkennung, Maschinelles Lernen, Video Annotation

1 Einleitung

Durch die fortschreitende Entwicklung zu einer ubiquitär von tragbaren Computern durchzogenen Welt, wuchs die Domäne des mobilen Aktivitäts- und Mobilitäts-Monitorings mit Hilfe von weit verbreiteten Technologien wie Sensoren aus Smartphones, Wearables und Kameras zu einem der größeren Trends der letzten Jahre heran. Derzeit gibt es verschiedene wissenschaftliche Ansätze zur Aktivitätserkennung, die neben der Auswertung von Bewegungssensoren [1] auch kamerabasierte Lösungen bevorzugen [2]. Auf Grundlage dieser Technologien werden Anwendungen für die unterschiedlichsten Bereiche, wie beispielsweise Gesundheitsüberwachung [3], Fitness-Spiele [4] und Auswertung von Fahrzeugtelemetriedaten [5] entwickelt. Außerdem untersuchen große Unternehmen Verfahrensweisen zur Unterstützung von Arbeitern bei komplexen Arbeitsaufgaben, um dadurch Fehler zu minimieren und die Effizienz in der Fertigung zu steigern [6]. In diesen Anwendungsfällen liegt der Fokus darauf, verschiedene Arten von Bewegungen zu erkennen, zu unterscheiden und qualitativ zu beurteilen. Entscheidend ist dabei eine hohe Erkennungsgenauigkeit in annähernder Echtzeit. Um diese Präzision zu erreichen, sind annotierte Datensätze in qualitativ hochwertiger Güte nötig. Einige vergangene Forschungsstudien greifen hierzu auf bestehende Datensätze zurück. In diesen befinden sich jedoch meistens Messwerte von kontextspezifischen Aktivitäten, die sich nur in den seltensten Fällen auf andere Domänen übertragen lassen. Andere Forschungsarbeiten wie Ward et al. [7] und Yang et al. [8] setzen wiederum auf aufwendige Messinstallationen, in denen separate Sensoren an Körperteilen von Studienteilnehmer befestigt werden, um die verschiedenen Aktivitätsmerkmale zu bestimmen. Das lange Tragen dieser Sensoren ist jedoch für die Probanden unangenehm und von daher nicht alltagstauglich. Außerdem ist die Anschaffung und Initialisierung solcher komplexen Messsysteme mit einem hohen Gesamtkostenaufwand verbunden und aus diesem Grund für den Einsatz im privaten Hausgebrauch ungeeignet. Durch immer leistungsfähigere Hardware lassen sich die mittlerweile von mobilen Betriebssystemen bereitgestellten Schnittstellen für die Verwendung von *Deep Learning*-Technologien nutzen [9]. Dadurch lassen sich - mit dieser weit verbreiteten und überall verfügbaren Technik - kontextsensitive Anwendungen, welche ihr Verhalten auf die aktuelle Umgebung eines

Nutzers abstimmen können, implementieren. Der vorliegende Text beschreibt einen Workflow, welcher die Erkennung von Aktivitäten mittels Sensoren aus handelsüblichen mobilen Endgeräten ermöglicht. Zur Entwicklung des Aktivitätsklassifikators wird in einer laufenden Forschungsstudie ein Trainingsdatensatz generiert und daraus resultierende Ergebnisse aufgezeigt.

2 Beschreibung des Workflows

Im Folgenden wird der in Abbildung 11.2 dargestellte Workflow zur Klassifikation von Aktivitäten beschrieben. Hierzu wird eine Methode zur Generierung eines *Ground Truth* -Datensatzes mittels einer mobilen Anwendung für Smartphone und Weareables vorgestellt. Anschließend wird der erstellte Datensatz zum Training eines maschinellen Lernalgorithmus eingesetzt. Abschließend wird das daraus resultierende Modell zur Echtzeitklassifikation von Aktivitäten auf mobilen Endgeräten angewandt.

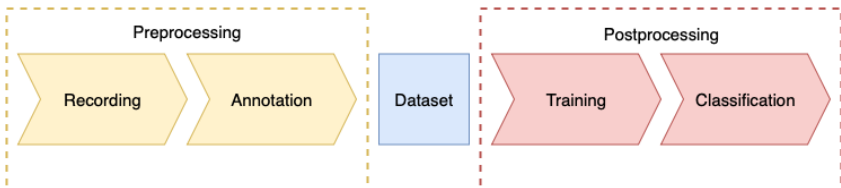


Abbildung 5.1: Workflow zur Aktivitätserkennung bestehend aus vier Schritten

2.1 Aufzeichnung der Sensordaten

Im Rahmen einer Untersuchung wird eine Gruppe von sechs Studenten mit aktueller Hardware ausgestattet. Dazu zählt eine Smartwatch (Apple Watch Series 3), ein Smartphone (Apple iPhone 7) und eine Action-Cam (GoPro Hero 4). Die Trageposition der Geräte ist fest vorgeschrieben und wird mit Hilfe einer Checkliste zu Beginn jeder Messung überprüft. Dadurch ist eine Vergleichbarkeit der einzelnen Messreihen gewährleistet. Die Smartwatch ist beispielsweise am Handgelenk der dominanten Hand des Probanden anzulegen. Dabei ist ein weiterer

wichtiger Aspekt die Ausrichtung der AppleWatch. Apple betont das eine genaue Messung der Herzfrequenz nur durch optimalen Sitz der Apple Watch gesichert ist [10]. Dazu muss die Rückseite der Uhr direkt auf der Haut aufliegen. Außerdem ist das Display zur Oberseite des Handgelenkes und die “Digital Crown“ in Richtung des Ellenbogens auszurichten. Zusätzlich sollte darauf geachtet werden das die Apple Watch weder zu locker noch fest sitzt, da dies nicht nur die Messergebnisse verfälschen würde, sondern auch zu Hautirritationen führen kann [10]. Ähnliche Vorgaben sind auch bei der Positionierung des Smartphones zu beachten. Dieses ist in der vorderen Hosentasche mit dem Display zum Oberschenkel und den Lightning-Anschluss nach oben zu platzieren. Die Actioncam ist sichtbar mit Hilfe eines Brustgurtes zentral am Körper der Testperson zu befestigen. Auf diese Weise wird das Sichtfeld der Versuchsperson großflächig in Form einer *First Person View* aufgezeichnet. Erst nach der ordnungsgemäßen Abarbeitung der Checkliste ist dem Probanden die Durchführung einer Messung erlaubt. Für diesen Zweck ist eine eigens für die Untersuchung implementierte App auf der Hardware installiert. Innerhalb der Benutzeroberfläche ist

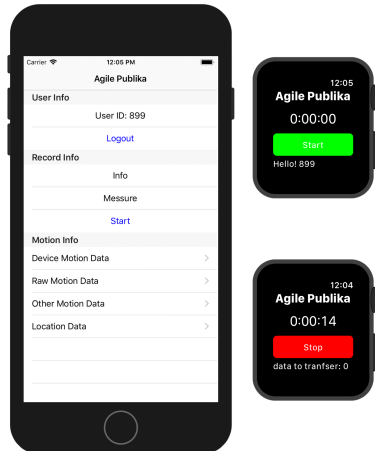


Abbildung 5.2: Screenshot der iOS- und watchOS-Benutzeroberfläche

eine Anmeldemaske vorgesehen, über die die Eintragung einer anonymen Identifikationsnummer durch den Anwender zu Beginn jeder Mes-

sung erfolgt (vgl. Abbildung 5.2). Dies verknüpft die einzelnen Messreihen mit den dazugehörigen Testpersonen und ermöglicht die Evaluation der benutzerabhängigen und benutzerunabhängigen Genauigkeit. Die restliche Bedienung der App erfolgt über die Benutzeroberfläche der AppleWatch. Diese bietet dem Anwender die Möglichkeit, die Messung zu starten bzw. zu stoppen und weitere Statusinformation - wie beispielsweise die bisherige Messdauer - abzulesen (vgl. Abbildung 5.2). Dies hat den Vorteil, dass der Nutzer während der Messung nur mit einem Gerät interagiert und sich somit auf die Ausführung der Aktivitäten konzentrieren kann. Darüber hinaus wird eine Verfälschung der Sensorwerte, wie beispielsweise durch das Herausnehmen und Hineinstecken des iPhones in die Hosentasche des Nutzers, vorgebeugt.

Die Schnittstellen in mobilen Betriebssystemen (z.B. Core Motion Framework) ermöglichen den Zugriff auf verschiedene Arten von Sensoren (vgl. Tabelle 5.1). Unterschieden wird hierbei zwischen Hardware Sensoren, die das Auslesen von Rohdaten erlauben und Software Sensoren, die eine bereits verarbeitete Version der Rohwerte bereitstellen. Aus den Daten der Software Sensoren sind bereits Verzerrungen durch Rauschen oder andere externe Einwirkungen wie der Einfluss durch die Erdanziehungskraft herausgerechnet. Dadurch ist es direkt möglich, lediglich die vom Benutzer aufgewandten Kräfte in der weiteren Auswertung zu analysieren. Die Aufzeichnung der Sensordaten erfolgt zusätzlich in Relation zu einem vordefinierten Weltkoordinatensystem, wodurch die Verdrehung des Gerätes um eine beliebige Achse im dreidimensionalen Raum exakt verfolgt werden kann [12]. Somit lässt sich die Abhängigkeit von der Ausrichtung des Gerätes minimieren und es ist bspw. irrelevant ob das Telefon mit den Anschlüssen nach unten oder oben in der Hosentasche platziert wird.

2.2 Annotation der Sensordaten

In dem hier beschriebenen Workflow kommt eine Form des überwachten Lernens zum Einsatz. Das Ziel von überwachtem Lernen ist, ein Model mit Hilfe von gekennzeichneten Trainingsdaten und bekannten Features zu trainieren und damit die Fähigkeit zu erlernen, Vorhersagen über unbekannte oder zukünftige Daten zu treffen [13]. Zur Erstellung eines Trainingsdatensatzes ist eine Zuordnung der aufgezeichneten Sensordaten zu den entsprechenden Aktivitätsklassen

Tabelle 5.1: Vereinfachte Übersicht der erfassten Sensordaten (nach [11])

Sensor	Typ	Messgröße
Beschleunigungssensor	Hardware	Beschleunigungsrohdaten für alle drei Achsen des Gerätes.
Gyroskop	Hardware	Rohdaten des Gyroskops für alle drei Achsen des Gerätes.
Magnetometer	Hardware	Das vom Magnetometer gemessene Erdmagnetfeld.
Barometer	Hardware	Der in Kilopascal gemessene Luftdruck.
Höhenmesser	Software	Die relative Höhenänderung in Bezug auf einen Startpunkt.
Schrittzähler	Software	Informationen über die zu Fuß zurückgelegte Wegstrecke.
Ausrichtungssensor	Software	Die Ausrichtung des Gerätes in einem vordefinierten Bezugssystem.
Drehratenmesser	Software	Die Drehrate des Geräts in Bezug zu einem vordefinierten Bezugssystem.
Gravitationssensor	Software	Die Erdanziehung die auf das Gerät einwirkt.
Benutzerbeschleunigung	Software	Die Beschleunigung die ein Benutzer auf Gerät ausübt.
Magnetfeldsensor	Software	Kalibrierte Magnetfelddaten mit einer gewissen Genauigkeit.
Kompass	Software	Der Kurswinkel des Gerätes bezogen auf ein vordefiniertes Bezugssystem.
Pulssensor	Hardware	Die Herzfrequenz des Benutzers.

nötig. Zu diesem Zweck folgt die Annotation des parallel zur Sensordatenerfassung aufgezeichneten Videostreams in einer webbasierenden Anwendung. Die Motivation für die Verwendung eines videobasierenden Annotations-Verfahrens ist die Entlastung der Probanden während der Durchführung von Aktivitäten, da eine zusätzliche Belastung durch

die parallele schriftliche Dokumentation entfällt. Außerdem wird eine Verfälschung der Sensordaten-Aufzeichnung durch die Bewegung der Hand, wie sie bei einer schriftlichen Protokollierung auftreten würde, verhindert. Darüber hinaus enthält das aufgezeichnete Videomaterial einen Großteil der ursprünglichen Informationen, wodurch Unschärfen bei der Erfassung von Aktivitäten durch Erinnerungslücken der Probanden weitestgehend auszuschließen sind. Dies trägt zur Verminderung der Nachteile, die sich durch eine retrospektive Arbeitsweise ergeben, bei. Die Benutzeroberfläche der Anwendung enthält eine Vorschau des Videomaterials, eine funktionale Zeitleiste zur Übersicht des Annotationsvorschlages und eine Auflistung der, zur Verwendung des Tools benötigten, Tastenbelegung. Es sind Hotkeys für die Videonavigation und die Zuordnung von Aktivitäten implementiert. Über diese erfolgt mit Hilfe manueller Tastatureingaben eine Annotation der Zeitbereiche, in denen die relevanten Aktivitäten auftreten. Ziel bei der Gestaltung der Benutzeroberfläche ist, den Zeitaufwand, der zur Annotation eines Videos benötigt wird, möglichst zu minimieren.

2.3 Training des Aktivitätsklassifikators

Das Ziel eines Aktivitätsklassifizierers ist, anhand von Sensordaten zwischen verschiedenen von Menschen ausgeführten Aktivitäten zu differenzieren. Hierbei existieren unterschiedliche Lernverfahren zur Analyse von Aktivitätsdaten. Akhavian und Behzadan [14] untersuchten fünf verschiedene Ansätze zur Klassifikation von Sensordaten während der Ausführung menschlicher Aktivitäten bei Bauvorgängen und erzielten mit künstlichen neuronalen Netzen die besten Resultate.

Ordóñez et al. zeigten darüber hinaus mit ihrer Veröffentlichung, dass die Verwendung von CNNs und LSTMs eine hohe Genauigkeit bei der Erkennung von menschlichen Aktivitäten erzielt und das sie damit andere Verfahren, welche auf eine Verwendung von rekurrenten neuronalen Netzen verzichten, um durchschnittlich 4% übertreffen konnten. Des Weiteren beschreiben sie in ihrer Forschungsarbeit einen Ansatz für ein generisches *Deep Learning* -Framework, das eine Reihe vorteilhafter Funktionen bietet: [15]

- Verarbeitung von Daten von multimodalen tragbaren Sensoren
- Durchführung von Sensordatenfusion

- Extraktion von Merkmalen ohne Fachkenntnisse
- Modellierung von zeitlichen Merkmalen

Basierend auf dieser Veröffentlichung hat Apple eine Funktionalität zur Erstellung von Aktivitätsklassifikatoren in ihr eigenes Framework für maschinelles Lernen namens *Turi Create* integriert. *Turi Create* wurde 2017 von Apple unter einer Open Source Lizenz auf *GitHub* veröffentlicht und ist für Entwickler gedacht, die wenig Erfahrung im Bereich des maschinellen Lernens haben. Deshalb sind zur Erstellung eines benutzerdefinierten maschinellen Lernmodells nur wenig Codezeilen nötig [16]. Zur Weiterverarbeitung der einzelnen Sensordaten in *Turi Create* werden diese zusammengeführt und in ein passendes Inputformat (*SFrame*) gebracht. Dabei sind alle Informationen zeitlich aufsteigend sortiert und die einzelnen Sensorwerte in einer Zeile aneinandergereiht. Jede Aufnahme entspricht einem Satz von Messwerten und wird als Session bezeichnet. Der Aktivitätsklassifizierer erwartet dass jeder Messwert einer Session zugeordnet ist und über ein Aktivitätslabel verfügt. Zur quantitativen Evaluierung des Klassifikators wird der generierte Datensatz über eine Zufallsfunktion in Trainings- und Testdaten unterteilt. Danach wird das Modell trainiert und dessen Generalisierungsfähigkeit über eine Evaluierungsfunktion bestimmt. Ein Test auf lediglich Trainingsdaten würde ansonsten nur einem Abrufen von gespeicherten Daten entsprechen [17]. Sobald das Modell zufriedenstellende Ergebnisse erreicht, wird der Trainingsvorgang abgeschlossen, das entstandene Modell gespeichert und in ein für iOS geeignetes Format (*.mlmodel*) exportiert. Anschließend erfolgt der Einsatz im Rahmen der Echtzeit-Klassifizierung auf dem iPhone.

2.4 Klassifikation von Aktivitäten

Mitte 2017 im Rahmen der Vorstellung von iOS 11 führte Apple ein neues Framework namens *Core ML* ein. Dieses Framework bietet leistungsstarke Funktionen, um maschinelles Lernen in Anwendungen für iOS, macOS, watchOS und tvOS zu integrieren. Es ermöglicht beispielsweise rechenintensive Klassifikationsverfahren mittels künstlicher neuronaler Netze lokal auf einem mobilen Endgerät durchzuführen. Dadurch entfällt die Notwendigkeit, Daten an

einen Server zu übertragen und die Verarbeitungsgeschwindigkeit wird durch Wegfall von Übertragungszeiten deutlich verkürzt. [18]

Ersten Implementierungen mit *Core ML* haben bereits vielversprechende Ergebnisse bei Klassifizierung von Bildern, Texten und Erkennung von menschlichen Posen geliefert [19]. In dieser Arbeit liegt der Fokus in der Verwendung eines mit *Turi Create* erstellten Modells zur Klassifizierung von Aktivitäten (vgl. Abschnitt 2.3). Hierzu ist es nötig die in Tabelle 5.1 aufgeführten Sensoren auf dem mobilen Endgerät zu aktivieren und mit der gewünschten Abtastfrequenz abzufragen.

Anschließend folgt die temporäre Speicherung der resultierenden Messwerte der einzelnen Sensoren in einem Array mit der Kapazität der gewünschten Fenstergröße. Die Fenstergröße entspricht hierbei der Länge des Vorhersageintervalls, mit dem die Klassifizierung der Aktivitäten erfolgt. Ist das Array mit Sensordaten gefüllt, wird dessen Inhalt an die *Prediction*-Funktion des Modells übergeben. Abschließend liefert das Modell eine Vorhersage zurück. Diese beinhaltet die erkannte Aktivität mit der höchsten Erkennungswahrscheinlichkeit. [20]

3 Ergebnisse

In einem ersten Pilotversuch wurde die mobile iOS-Anwendung von sechs Studenten im Rahmen einer Projektarbeit verwendet. Die Gruppe nahm hierzu verschiedene Aktivitäten auf und protokollierte ihre Erfahrungen und Erkenntnisse. Dabei hat sich ergeben, dass nach einer gewissen Einarbeitungsphase die Bedienung der App auch ohne technische Fachkenntnisse gelingt. Anfängliche Schwierigkeiten hingen mit der falschen Positionierung der Hardware und fehlerhaften Bedienung der Anwendung zusammen. Diese Hürden wurden im Laufe des Versuchszeitraums durch die steigende Vertrautheit im Umgang mit der Messtechnik und die Einführung der in Abschnitt 2.1 erwähnten Checkliste gelöst. Zusätzlich ließ sich ein Fehler durch ein abgelaufenes Entwicklerzertifikat, der zu einem Löschen der App führte, rasch beheben. Ein langfristiges Problem, das Verbindungsprobleme zwischen iPhone und Apple Watch verursachte, verschwand nach einem Bugfix innerhalb eines Updates des watchOS-Betriebssystems.

Es wurden innerhalb eines Semesters ca. 24 Stunden an Sensor- und Videodaten erfasst. Dabei ist es gelungen, die einzelnen Informationen in

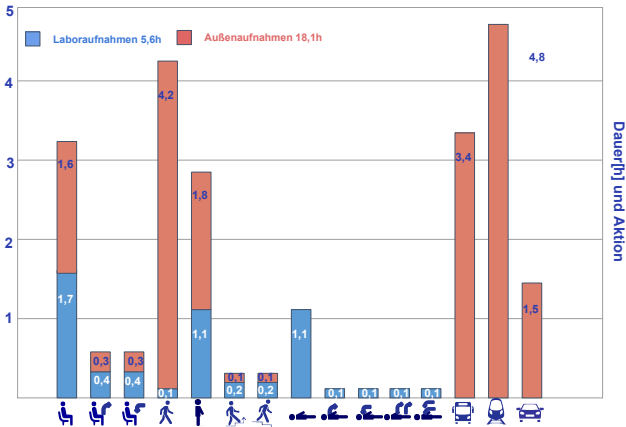


Abbildung 5.3: Übersicht der Aufnahmen der ersten Anwendungsstudie [21]

einer gemeinsamen Datenbank abzuspeichern und über die Zeitstempel miteinander zu verknüpfen. Aufgenommen wurden Aktivitäten wie Liegen, Sitzen, Stehen, Laufen, Treppensteigen und die Übergänge zwischen den Bewegungsformen. Außerdem wurden Messungen in unterschiedlichen Verkehrsmitteln wie Auto, Bus und Bahn durchgeführt. Eine Übersicht der getätigten Aufnahmen ist in Abbildung 5.3 dargestellt. Verursacht durch die o.g. Herausforderung sind ca. 10% der getätigten Aufnahmen unbrauchbar oder nur teilweise brauchbar. Der restliche Teil des Datensatzes wird aktuell - mit dem im Abschnitt 2.2 beschriebenen Annotations-Tool - ausgewertet und dabei die Sensordaten mit dem dazugehörigen Aktivitäts-Label versehen.

4 Schlussfolgerung

In diesem Paper wurde ein Workflow vorgestellt, der bereits in einer laufend Untersuchung Anwendung findet. Erste Aufzeichnungen der hier vorgestellten Smartphone-App befinden sich bereits in einer zentralen Datenbank und werden aktuell, nachdem in dieser Arbeit beschriebenen Verfahren, annotiert. Das Ergebnis zeigt, dass die Generierung eines Aktivitätsdatensatzes mit handelsüblichen Smartpho-

nes und Weareables möglich ist. Zukünftig folgt das in 2.3 beschriebene Training des Aktivitätsklassifikators und die Evaluierung des resultierenden Modells. Abschließend ist angedacht, die Güte des Modells während der Klassifikation von Aktivitäten in einer Echtzeitanwendung zu analysieren. Außerdem ist geplant die Handhabung, Skalierbarkeit und Robustheit des Workflows in weiteren Feld- und Langzeittests zu optimieren. Weiterhin ist beabsichtigt, vergleichbare Testdatensätze für andere Aktivitäten zu generieren und die Performance des Aktivitätsklassifikators anderen Verfahren gegenüberzustellen.

Danksagung

Diese Arbeit wurde gefördert durch den Europäischen Sozialfonds (ESF) und den Freistaat Sachsen.

Literatur

1. Y. Chen and C. Shen, "Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition." *IEEE Access*, vol. 5, pp. 3095–3110, 2017.
2. A. Jalal, Y. Kim, Y.-J. Kim, S. Kamal, and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused features." *Pattern Recognition*, vol. 61, pp. 295–308, 2017.
3. L. E. Moser and P. M. Melliar-Smith, "Personal Health Monitoring Using a Smartphone," in *2015 IEEE International Conference on Mobile Services (MS)*. IEEE, 2015, pp. 344–351.
4. A. Direito, Y. Jiang, R. Whittaker, and R. Maddison, "Smartphone apps to improve fitness and increase physical activity among young people: protocol of the Apps for IMproving FITness (AIMFIT) randomized controlled trial," *BMC Public Health*, vol. 15, no. 1, p. 635, Dec. 2015.
5. N. Cardoso, J. Madureira, and N. P. 0001, "Smartphone-based transport mode detection for elderly care." *HealthCom*, pp. 1–6, 2016.
6. DFKI and Hitachi jointly develop AI technology for human activity recognition of workers using wearable devices.
7. J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, "Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers,"

IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 10, pp. 1553–1567, 2006.

8. A. Y. Yang, S. Iyengar, P. Kuryloski, and R. Jafari, “Distributed segmentation and classification of human actions using a wearable motion sensor network,” in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*. IEEE, 2008, pp. 1–8.
9. J. Newnham, *Machine Learning with Core ML: An iOS developer’s guide to implementing machine learning in mobile apps*. Packt Publishing, Jul. 2018.
10. Apple. (2018) Tragen der Apple Watch - Apple Support. <https://support.apple.com/de-de/HT204665#bestfit> besucht am: 28.01.2019.
11. A. Dev. (2018) Core Motion Framework - Apple Developer. <https://developer.apple.com/documentation/coremotion> besucht am: 21.08.2018.
12. A. Henpraserttae, S. Thiemjarus, and S. Marukatat, “Accurate Activity Recognition Using a Mobile Phone Regardless of Device Orientation and Location.” *BSN*, pp. 41–46, 2011.
13. S. Raschka, *Python Machine Learning*. Packt Publishing Ltd, Sep. 2015.
14. R. Akhavian and A. H. Behzadan, “Wearable sensor-based activity recognition for data-driven simulation of construction workers’ activities.” *Winter Simulation Conference*, pp. 3333–3344, 2015.
15. F. Ordóñez and D. Roggen, “Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition,” *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016.
16. Apple. (2018) Turi Create - User Guide. <https://apple.github.io/turicreate/docs/userguide/> besucht am: 21.01.2019.
17. W. Ertel, *Grundkurs Künstliche Intelligenz*. Wiesbaden: Springer Fachmedien Wiesbaden, 2016.
18. Apple. (2018) Machine Learning - Apple Developer. <https://developer.apple.com/machine-learning> besucht am: 15.01.2019.
19. K. Li. (2018) Awesome CoreML Models - GitHub. <https://github.com/likedan/Awesome-CoreML-Models> besucht am: 18.01.2019.
20. Apple. (2018) Deployment to Core ML GitBook. <https://apple.github.io/turicreate/docs/userguide> besucht am: 15.01.2019.
21. F. M. Hammer, Hironymus, “Mobile Analyse - Abschlusspräsentation - Hochschule Mittweida,” 2018.

Anpassung multipler ML-Frameworks der Personen- und Ortserkennung zur Verwendung in der EMSML-Umgebung

Analyse und Entwicklung eines generischen Workflows

Christian Roschke, Rico Thomanek und Benny Platte

Hochschule Mittweida, D-09648 Mittweida, roschke@hs-mittweida.de

Zusammenfassung Die Kombination der Personen- und Ortserkennung erschließt multiple neue Anwendungsfelder, wie beispielsweise das Entwickeln von Ansätzen, um mittels Echtzeitüberwachung vermisste Personen im öffentlichen Raum wiederzufinden. Dabei ist es notwendig Frameworks beider Domänen auf gegebene Datensätze anzuwenden und erhaltene Ergebnisse zu fusionieren. Für diesen Einsatzzweck wurde von Thomanek et al. 2019 ein Evaluations- und Managementsystem für Maschinelles Lernen (EMSML) entwickelt, welches die Verschaltung verschiedener Frameworks und die Fusionierung von Ergebnisvektoren ermöglicht [1]. In diesem Beitrag wird das EMSML bezüglich der Schnittstellen und Komponenten untersucht, um einen generischen Workflow zu entwickeln, der die Integration verschiedener Frameworks der Personen- und Ortserkennung ermöglicht. Diesbezüglich wird auf notwendige Anpassungen bestehender Frameworks an die implementierte Infrastruktur fokussiert. Aus den Analyseergebnissen lässt sich ein Konzept in Form eines generischen Workflows ableiten. Dieses Konzept ist sich zur Evaluation an zwei exemplarischen Frameworks anwend- und jeweils prototypisch implementierbar. Anschließend werden entwickelte Testfälle genutzt, um die funktionale Validität der Prototypen und die Anwendbarkeit des Konzeptes aufzuzeigen.

Schlagwörter Generischer Workflow, Machine Learning, Personenerkennung, Ortserkennung

1 Einleitung

Die automatische Detektion und Wiedererkennung von Personen in Videos hat viele Anwendungen in multiplen Bereichen. Dies umfasst beispielsweise Vorteile für die Sicherung von Zugangsseiten, für passwortfreie Computeranmeldung und auch für die Analyse von Verhaltensweisen von Einzelpersonen. Die Ortserkennung in audiovisuellen Medien kann darüber hinaus eingesetzt werden, um eine videobasierte Navigation in Gebäuden zu realisieren oder nicht annotierte Datensätze im Nachhinein adäquat zu annotieren. Durch die Kombination der beiden beschriebenen Anwendungsfelder erschließen sich neue mögliche Anwendungsbeispiele. Beispielsweise ergibt sich dadurch das Problemfeld der Personen- und Ortserkennung in Echtzeit-Überwachungsdaten. Dabei werden Ansätze gesucht, um beispielsweise vermisste Personen in Überwachungsmaterial wiederzufinden.

Zur Bewältigung dieses Problemfeldes entstand im Rahmen der internationalen Evaluationskampagne TRECVID der Aufgabenbereich Instance Search, welcher seit acht Jahren durchgeführt wird [2]. Diese Kampagne wird vom National Institute of Standards and Technology (NIST, USA) durchgeführt und hat zum Ziel, die Forschung im Bereich des robusten Multimedia Information Retrieval (MIR) zu fördern. In diesem Umfeld konnten auf Basis eines definierten Datensatz von 464 Stunden Videomaterial der langjährigen BBC-Serie EastEnders, mannigfaltige Systeme und Ansätze entwickelt und getestet werden. [3]

Ansätze aus den Domänen der Personen- und Ortserkennung wie Peng et al. 2018 [4], Chen et al. 2018 [5], Makhayeri et al. 2019 [6] und Zeng et al. 2018 [7] sind meist problemspezifisch und lassen sich schwer auf andere Domänen übertragen. Weiterhin werden sowohl bestehende Frameworks genutzt sowie eigene CNNs trainiert, wobei sich die ermittelten Merkmale ausschließlich für die direkte Auswertung im Kontext des zu lösenden Problems verwenden lassen. Eine standardisierte Speicherung, Verwaltung und Anwendung auf andere Problemfelder ist nicht vorgesehen.

In Thomanek et al. 2019 wird ein Ansatz beschrieben, bei dem ein Evaluations- und Management System für das Maschinelle Lernen (EMSML) zur Evaluierung und einfachen Kombination bestehender Frameworks entwickelt wurde [8]. Das dabei entstandene System hat zum Ziel den kompletten Workflow zu standardisieren, Ressour-

cen zielführend zur Verfügung zu stellen, Forschungsdaten persistent zu speichern sowie für die Weiterverarbeitung aufzubereiten und alle erhaltenen Ergebnisse miteinander zu fusionieren. Diesbezüglich entstand eine komplexe Infrastruktur, welche auch in anderen Anwendungsdomänen verwendet werden kann. Ein Defizit des entstandenen Systems besteht aktuell in der Notwendigkeit bestehende Frameworks zur Integration in die Infrastruktur anzupassen. [1]

Dieser Beitrag zeigt einen generischen Workflow aufzeigen, welcher notwendig ist, um Frameworks der Personen- und Ortserkennung in das EMSML zu integrieren. Dabei werden vorhandene Schnittstellen untersucht und insbesondere auf die Anpassungen bestehender Frameworks sowie Technologien an die Anforderungen der implementierten Infrastruktur eingegangen.

2 Methode

Zur Integration bestehender Frameworks sowie Merkmalsextraktoren aus den Domänen der Personen- und der Ortserkennung in das EMSML ist es notwendig, dieses bezüglich der Schnittstellen und Komponenten zu analysieren und einen generischen Arbeitsablauf konzeptionell abzuleiten. Dieses Konzept ist auf ausgewählte Frameworks anzuwenden und jeweils prototypisch zu implementieren. Die Prototypen dienen der Evaluierung, ob das generische Konzept funktioniert. Dabei wird das holistische System inklusive der Prototypen auf einen gegebenen annotierten Datensatz angewandt. Dies ermöglicht Speicherstrukturen des EMSML im zeitlichen Verlauf zu betrachten und zu ermitteln, ob sich das System erwartungsgemäß verhält. Diesbezüglich ist zu prüfen, ob Teilergebnisse gespeichert werden können und ob am Ende der Prozesskette die Annotationen mit den ermittelten Endergebnissen vergleichbar erscheinen.

2.1 EMSML-Architektur

Das EMSML ist ein Werkzeug zum Aufbau holistischer Prozessketten im Kontext des Maschinellen Lernens. Dabei werden multiple Komponenten dezentralisiert gespeichert, miteinander vernetzt und der Datenfluss über standardisierte Schnittstellen ermöglicht. Dies kann genutzt

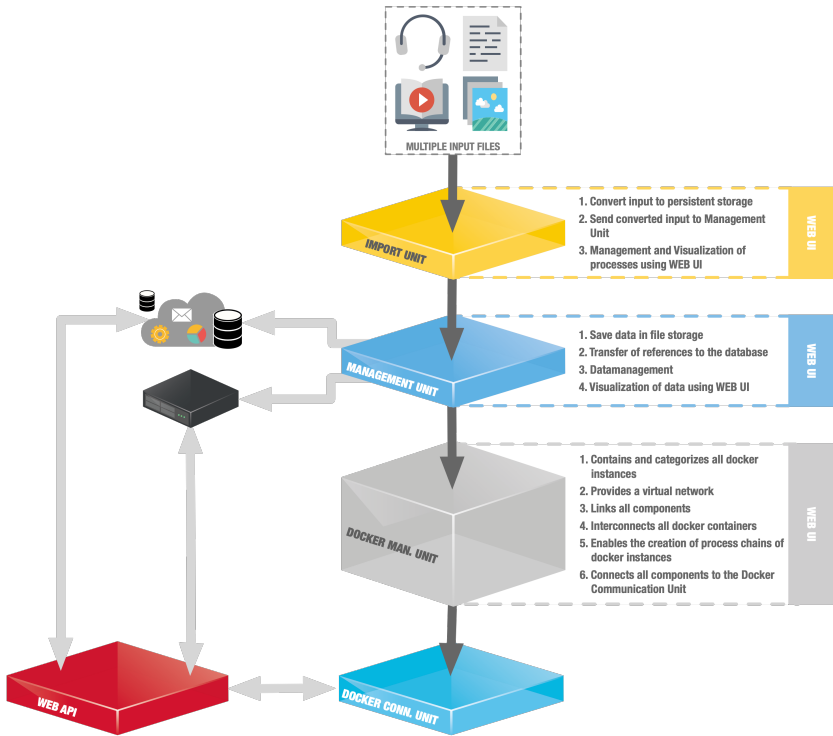


Abbildung 6.1: Schematische Darstellung der Systemarchitektur des EMSML

werden, um bestehende Prozesse abzubilden oder neue zu entwickeln, wobei durch den grundlegenden Aufbau eine Parallelisierbarkeit, Skalierbarkeit und Verteilbarkeit gegeben ist. Somit können verschiedenste Frameworks miteinander kombiniert werden und die Verarbeitung bei ressourcenärmeren Computern auf weitere Rechner verteilt werden. Abbildung 6.1 zeigt den grundlegenden Aufbau des EMSML. Dieses System besteht im Wesentlichen aus mehreren nachfolgend beschriebenen Kernkomponenten zur Durchführung von Prozessen und einem Datenbankserver zur dauerhaften Speicherung aller Ergebnisse und Rohdaten.

Import und Datenverwaltung

Die *Import Unit* dient zum Importieren multipler Dateien wie beispielsweise Bildern, Texten und audiovisuellen Medien. Diesbezüglich lassen sich diese einzeln oder gesamtheitlich in das System überführen. Dabei werden alle importierten Daten gepackt, aufbereitet und an die *Management Unit* weitergeleitet. Besonders dabei ist, dass Import-Prozesse beliebig anpassbar sind. Bei der *Management Unit* handelt es sich um die zentrale Einheit zur Verwaltung der Daten. Dabei werden die von der *Import Unit* übergebenen Dateien in einen zentralen Speicher transferiert und alle Referenzen in entsprechenden Datenbankstrukturen vorgehalten. Sowohl die Daten als auch die Datenbank können dabei mittels WEB-Schnittstelle eingesehen werden. Alle Dateien werden auf einem zentralen Server gespeichert und mittels diverser Protokolle, wie beispielsweise SMB, NFS, HTTP und FTP, zugreifbar gemacht. Somit ist der Zugriff auf den Speicher auch über andere Werkzeuge möglich. Anschließend werden ausschließlich die absoluten Referenzen und nicht die Daten selbst in das Datenbanksystem transferiert. Bei der Datenbank handelt es sich um PostgreSQL.

Docker Management Unit

Erweiternd zu der *Import Unit* und der *Management Unit*, welche im Kern das Datenmanagement inklusive Importmechanismen inkludieren, stellt die *Docker Management Unit* Logik bereit, um weitere Funktionalitäten in Form von Docker-Containern hinzuzufügen. Diese Unit basiert auf Portainer IO, einem einfachen Management UI für Docker, wobei Docker-Container verwaltet und kategorisiert werden können. Überdies ist es möglich ein virtuelles Netzwerk aufzubauen und Container miteinander zu verknüpfen. Dies ermöglicht den geregelten Informationsaustausch zwischen einzelnen Komponenten. Darüber hinaus lassen sich Prozessketten erstellen und verwalten. Eine weitere zentrale Funktion dieser Unit ist das Management und die Verknüpfung, der in Abbildung 6.2 dargestellten Teilbereiche *Machine Learning Frameworks*, *Processing*, *Scoring* und *Export*. Besonders dabei ist, dass für jeden Teilbereich beliebig viele neue Teilkomponenten in Form von Docker-Containern hinzugefügt werden können.

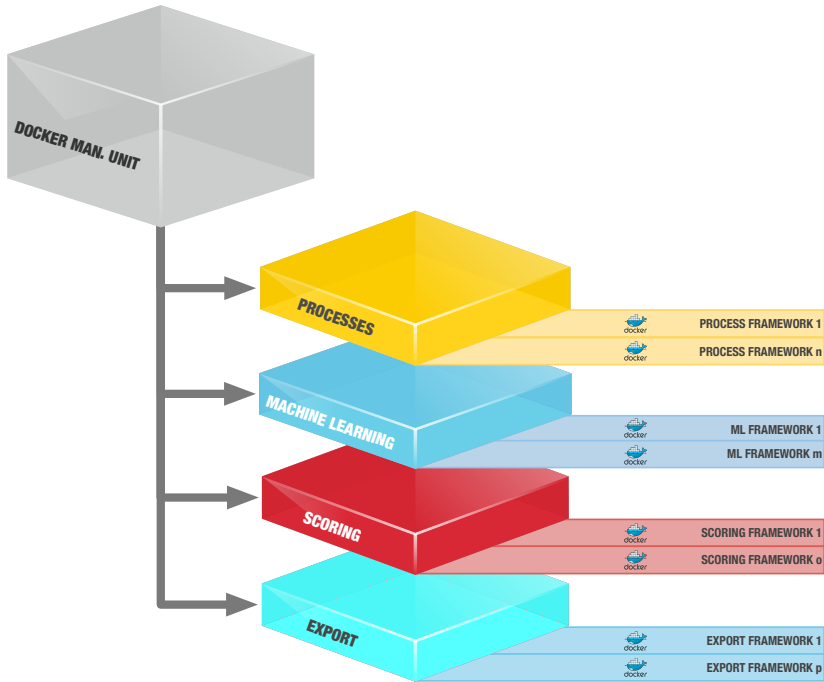


Abbildung 6.2: Schematische Darstellung der *Docker Management Unit*

Verknüpfung Datenverwaltung und Dockermanagement

Um die Daten aus der Datenbank filterbar und individualisiert austauschen zu können, verfügt das EMSML über eine API. Diese ist in der Lage HTTP-Anfragen entgegenzunehmen und weiterzuverarbeiten sowie gezielte Anfragen an die Datenbank zu stellen. Darüber hinaus ermöglicht diese Komponente Datensätze beliebig zu kombinieren und in einem beliebigen Austauschformat zu übertragen. Überdies lassen sich darüber Dateien direkt aus dem Dateispeicher beziehen. Zur Kommunikation aller Docker-Container mit anderen Werkzeugen außerhalb der virtualisierten Umgebungen und der API existiert die *Docker Connection Unit*. Diese ruft mittels HTTP Datensätze zur Verarbeitung ab, leitet Ergebnisse weiter und bildet die Schnittstelle zur *Docker*

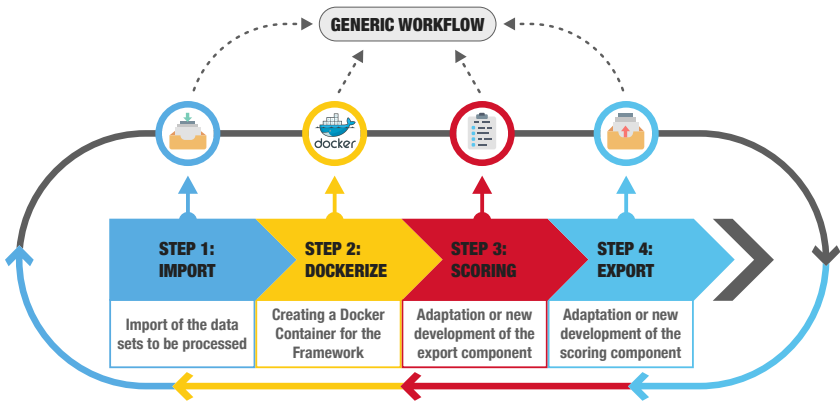


Abbildung 6.3: Generischer Workflow zur Anpassung eines Frameworks zur Integration in das EMSML

Management Unit.

2.2 Anpassung multipler Frameworks

Basierend auf der Architektur des EMSML lassen sich durch die Schnittstellen bedingte in Abbildung 11.2 dargestellte und nachfolgend beschriebene Schritte identifizieren, welche durchzuführen sind, um ein weiteres Framework dem EMSML hinzuzufügen.

Schritt 1: Import der zu verarbeitenden Datensätze

Der zu verarbeitende Datensatz muss mittels der *Import Unit* in das System übertragen werden. Dabei lässt sich auf einen Satz an Hilfswerkzeugen zurückgreifen. Für alle importierten und vorverarbeiteten Daten werden Datenbanktabellen erzeugt und ein Datensatz-Label vergeben, was eine Zuordnung jedes importierten Elements zu einem Datensatz ermöglicht. Alternativ zum Import können bereits angelegte Datensätze verwendet werden.

Schritt 2: Erstellung eines Docker-Containers für das Framework

Insofern ein Framework nicht bereits als Docker-Container vorliegt ist ein neues Docker-Image zu erstellen und in der *Docker Management Unit* hinzuzufügen. Je nach verwendeten HOST Betriebssystem und benötigter Bibliotheken ist dabei auf die Verwendung eines geeigneten Basis-Images zu achten. Basierend darauf lassen sich alle weiteren Werkzeuge und das Framework selbst im Docker-Container installieren.

Nach der allgemeinen Einrichtung des Containers und dem Test, ob das Framework in der virtualisierten Umgebung funktioniert, müssen Anpassungen an die EMSML-Infrastruktur vorgenommen werden. Dabei ist eine Komponente hinzuzufügen, welche mittels HTTP mit der Web-API kommunizieren kann. Diesbezüglich stellt der Container eine Anfrage und übermittelt mittels POST-Request den Datensatz und das benötigte Austauschformat. Die Web-API wird daraufhin die Referenzen auf die Dateien wie angefordert aufbereiten und an den Container weiterleiten. Das Framework ist so anzupassen, dass das Austauschformat ausgelesen wird und die Dateien mittels GET-Requests an die aufbereiteten Referenzen zur weiteren Verarbeitung bezogen werden können. Weiterhin ist die Ausgabe des Frameworks dahingehend zu modifizieren, dass die erzeugten Daten mittels HTTP im JSON-Format an die Web-API gesendet werden. In die Verarbeitung der Daten innerhalb der Frameworks kann verzichtet werden, da ausschließlich die Schnittstellen anzupassen sind.

Nach der Anpassung der Schnittstellen innerhalb des Containers, muss dieser noch in der *Docker Connection Unit* mit der Web-API verbunden werden um die Kommunikation zu ermöglichen. Überdies ist in der Datenbank für das entsprechende Framework noch eine Datenbanktabelle anzulegen, welche die zu übertragenden Daten abbilden kann. Die Attribute der Tabelle werden automatisch aus der übermittelten JSON erstellt. Dabei werden die Ergebnisse mit den Datensatz in Relation gesetzt, um eine spätere Verarbeitung mittels Scorer zu ermöglichen.

Schritt 3: Die Scoring-Komponente

Nach der Datenverarbeitung des Frameworks sind die Ergebnisse in der zugehörigen Tabelle vorhanden. Um diese auswerten zu können, ist eine Leistungsbewertung (Scorer) notwendig. Im System ist bereits ein Scorer in Form eines Docker-Containers integriert, wobei dieser beliebig anpassbar ist. Mittels des Containers ist es möglich, mehrere Ergebnistabellen zu verwenden, wobei alle Ergebnisse einer Tabelle auf einer Punkte-Skala abgebildet werden. Dabei bekommt ein Ergebnis mindestens 0 Punkte und maximal 100 Punkte. Überdies können zur Berechnung eines Scores auch weitere Ansätze des Maschinellen Lernens, welche die *Machine Learning-Komponente* zur Verfügung stellt, zur Datenfusionierung verwendet werden. Die durch den Scorer erzeugten Ergebnisse werden in einer zugehörigen Tabelle gespeichert und mit den Datensätzen in Relation gesetzt.

Schritt 4: Die Export-Komponente

Basierend auf der durch den Scorer entstandenen Ergebnistabelle kann die Export-Komponente die Ergebnisse in einem Austauschformat exportieren. Bereits im System integriert ist ein Docker-Container, welcher die Scorer-Ergebnisse anhand einer XML Schema Definition (XSD) aufbereitet und als XML exportiert.

2.3 Anwendung des entwickelten Workflows

Sowohl für die Ortserkennung als auch für die Personenerkennung werden zwei Datensätze ausgewählt und in Schritt 1 in das System importiert. Für die Ortserkennung wird der Datensatz „Places 2nd generation“ und das Framework „ResNet152-places365“ verwendet. „Places 2nd generation“ ist eine szenenzentrierte Datenbank mit 205 Szenenkategorien und 2,5 Millionen annotierten Bildern [9]. Aus ressourcentechnischen Gründen werden davon 10.000 intellektuell selektierte Bilder zum Test verwendet. Basis für die Personenerkennung ist der Datensatz „SCface - Surveillance Cameras Face Database“ mit 4160 statischen Bildern von 130 Personen [10]. Zur Überführung der Datensätze in das System wird die *Import Unit* genutzt. Nach der Durchführung wird erwartet, dass das EMSML 14.160 Bilder aus beiden Datensätzen in er-

stellte Verzeichnisse kopiert hat und in der Datenbank zwei mit Referenzen gefüllte Tabellen angelegt wurden. Dabei ist insbesondere darauf zu achten, dass alle Bilder mit dem korrekten Datensatzlabel versehen werden.

Für den Test ausgewählte Frameworks sind für die Domäne der Ortserkennung „ResNet152-places365“ und für die Gesichtserkennung aus der Domäne der Personenerkennung „FaceNet“. „ResNet152-places365“ ermöglicht es, für jedes Bild zehn Klassifizierungen mit abnehmender Genauigkeit zu bestimmen [11]. Darüber hinaus werden die ersten zehn gefundenen Attribute des *Sun Attribute*-Datensatzes [12, 13] für jedes verarbeitete Bild bereitgestellt. „FaceNet“ ist ein One-Shot-Modell, das direkt eine Zuordnung von Bildern zu einem kompakten euklidischen Raum lernt, in dem Entfernungen direkt dem Maß der Gesichtähnlichkeit entsprechen [14]. Derzeit stehen mehrere Implementierungen von „FaceNet“ zur Verfügung. Genutzt wird die Tensorflow Implementierung von David Sandberg. Für beiden Frameworks existiert bereits ein Docker-Container, welcher anzupassen ist und in das EMSML integriert werden kann. „ResNet152-places365“ und „FaceNet“ verfügen über PythonAPIs zur Anwendung der Frameworks. Diese APIs sind in beiden Implementationen einfach erweiter- und anpassbar, sodass eine Routine integriert werden kann, um Daten direkt von der Web-API zu laden und Ergebnisse als JSON an diese zu übergeben. Für beide Frameworks sind diesbezüglich noch zwei Datenbanktabellen zu erstellen. Nach der Verknüpfung der Docker-Container mit der *Docker Connection Unit* wird erwartet, dass beide Datenbanktabellen mit frameworkspezifischen Attributen und mit detektierten Ergebnissen gefüllt sind. Ebenso muss eine relationale Verknüpfung zwischen den Datensatzbildern und den erzeugten Daten existieren.

Für die Leistungsbewertung wird in beiden Fällen die im System vorhandene Scoring-Komponente verwendet. Dabei ist syntaktisch zu testen, dass die Ergebnisse in ein Punktesystem überführt werden und eine Endergebnistabelle erzeugt wird. Mittels einer XSD wird dem System ein Schema vorgegeben um die Daten zu exportieren. Basierend auf den Scoring-Ergebnissen und dem übergebenen Schema muss eine valide XML erzeugt werden.

3 Ergebnisse und Diskussion

Alle Erwartungen des Tests wurden in normierte Testfälle überführt. Bei dem zugrundeliegenden Benennungsschema setzt sich die ID aus T für Testfall, S für Schritt, $\{1-4\}$ für die Nummer des jeweiligen Schritts sowie $F1-F9$ für die Fallnummer zusammen. Die Durchführung aller Schritte des Workflows konnte durchgeführt werden. Ein Auszug aus dem entstandenen Testdokument ist in Tabelle 6.1 dargestellt. Alle Testfälle ließen sich erfolgreich durchführen. Der generische Workflow war für die exemplarisch ausgewählten Frameworks erfolgreich einsetzbar. Bei der Anwendung des Workflows sind mehrere nachfolgend erläuterte Vor- und Nachteile aufgefallen. Der modulare Aufbau des Workflows erlaubte eine strukturierte Abarbeitung und individuelle Anpassungen des Transferprozesses der Frameworks in das EMSML. Durch die modulare Sicht auf den kompletten Prozess ist es nun möglich Optimierungen schrittweise durchzuführen und den Workflow im Kontext der Automatisierbarkeit zu untersuchen. Dies wäre insofern notwendig, da die intellektuelle Abarbeitung des Prozesses fehleranfällig und zeitintensiv ist. Insbesondere die Überführung der Frameworks in einen Docker-Container oder die Anbindung an die standardisierten Schnittstellen des Systems ist nicht trivial und verbleibt frameworkspezifisch. In diesem Kontext muss der Workflow anhand weiterer Frameworks untersucht werden.

4 Zusammenfassung und Ausblick

In diesem Beitrag konnte ein generischer Workflow entwickelt werden um Frameworks aus der Domäne der Gesichtserkennung- und Ortserkennung in das EMSML zu integrieren. Mittels mehrerer exemplarisch ausgewählter Frameworks war es möglich diesen Workflow prototypisch umzusetzen und zu testen. Die Untersuchung zeigt, dass der entwickelte Ablauf funktioniert und als systematische Anleitung dienen kann. Der modulare Aufbau des Ablaufs ermöglicht weitere Untersuchungen im Kontext der Automatisierbarkeit, um eine generische sowie automatisierte Integrationskomponente für das EMSML zu entwickeln. Dies würde die bisher intellektuell durchzuführende Anpassung einzelner Frameworks im Sinne der Effizienz optimieren. Weiter-

Tabelle 6.1: Testergebnisse EMSML mit angepassten Frameworks

ID	Erwartung	Status
TS1F1	Zwei Verzeichnisse werden erstellt.	✓
TS1F2	14.160 Bilder werden auf beide Verzeichnisse aufgeteilt.	✓
TS1F3	In der Datenbank werden zwei Tabellen angelegt.	✓
TS1F4	Beide Tabellen sind mit Referenzen gefüllt.	✓
TS1F5	Alle Bilder sind mit dem korrekten Datensatzlabel versehen.	✓
TS2F1	Frameworkspezifische Attribute wurden angelegt.	✓
TS2F2	Alle Relationen sind vorhanden.	✓
TS3F1	Zwei Datenbanktabellen wurden erzeugt.	✓
TS3F2	Alle attributbasierten Daten wurden eingefügt und verknüpft.	✓
TS4F1	Eine Scoring-Tabelle wurde erstellt.	✓
TS4F2	Alle Ergebnisse wurden in Punkte überführt.	✓
TS4F3	Eine XSD kann übergeben werden und eine XML wird erstellt.	✓
TS4F4	Erzeugte XML ist valide und enthält Scoring-Daten.	✓

hin können die Erkenntnisse genutzt werden, um weitere Frameworks anzupassen und iterativ die bisher rudimentär implementierte Scoring-Komponente zu verbessern oder mehrere Scoring-Verfahren miteinander zu vergleichen.

Danksagung

Diese Forschung wurde aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01PL17062 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor_innen.



Literatur

1. R. Thomanek, C. Roschke, B. Platte, R. Manthey, T. Rolletschke, M. Heinzig, M. Vodel, F. Zimmer, M. Eibl, and M. Ritter, "A Scalable System Architecture for Activity Detection with Simple Heuristics," *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019.
2. G. Awad, J. Fiscus, M. Michel, D. Joy, W. Kraaij, A. F. Smeaton, G. Quénot, M. Eskevich, R. Aly, and R. Ordelman, "TRECVID 2016: Evaluating Video Search, Video Event Detection, Localization, and Hyperlinking," in *Proceedings of TRECVID 2016*. NIST, USA, 2016.
3. G. Awad, A. Butt, J. Fiscus, D. Joy, A. Delgado, M. Michel, A. Smeaton, Y. Graham, W. Kraaij, G. Quénot, M. Eskevich, R. Ordelman, G. J. Jones, and B. Huet, "Trecvid 2017: Evaluating ad-hoc and instance video search, events detection, video captioning and hyperlinking — The Insight Centre for Data Analytics," in *Proceedings of TRECVID 2017*, 2017.
4. Y. Peng, X. Huang, Y. Zhuo, Z. Ye, Z. Wen, J. Chi, and J. Zhang, "PKU ICST at TRECVID 2018: Instance Search Task," *Proceedings of TRECVID Workshop*, vol. Proceeding, 2018.
5. C. Chen, J. Huang, C. Pan, and X. Yuan, "Military Image Scene Recognition Based on CNN and Semantic Information," in *2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*. IEEE, 9 2018, pp. 573–577. [Online]. Available: <https://ieeexplore.ieee.org/document/8537624/>
6. F. Mokhayeri, E. Granger, and G.-A. Bilodeau, "Domain-Specific Face Synthesis for Video Face Recognition From a Single Sample Per Person," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 3, pp. 757–772, 3 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8443382/>
7. J. Zeng, X. Zhao, J. Gan, C. Mai, Y. Zhai, and F. Wang, "Deep Convolutional Neural Network Used in Single Sample per Person Face Recognition," *Computational Intelligence and Neuroscience*, vol. 2018, pp.

- 1–11, 8 2018. [Online]. Available: <https://www.hindawi.com/journals/cin/2018/3803627/>
8. R. Thomanek, C. Roschke, R. Manthey, B. Platte, T. Rolletschke, M. Heinzig, M. Vodel, D. Kowerko, S. Kahl, F. Zimmer, M. Eibl, and M. Ritter, "University of Applied Sciences Mittweida and Chemnitz University of Technology at TRECVID 2018," in *Proceedings of TRECVID Workshop*, 2018.
 9. B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning Deep Features for Scene Recognition using Places Database," *Advances in Neural Information Processing Systems* 27, 2014.
 10. P. Tome, J. Fierrez, R. Vera-Rodriguez, and D. Ramos, "Identification using face regions: Application and assessment in forensic scenarios," in *Forensic Science International*, 2013.
 11. X. Chen, Z. Ji, Y. Fan, and Y. Zhan, "Restful API Architecture Based on Laravel Framework," in *Journal of Physics: Conference Series*, 2017.
 12. M. Grgic, K. Delac, and S. Grgic, "SCface - Surveillance cameras face database," *Multimedia Tools and Applications*, 2011.
 13. G. Patterson and J. Hays, "SUN attribute database: Discovering, annotating, and recognizing scene attributes," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012.
 14. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015.

Bildbasierte Aktivitätsanalyse mittels dezentraler Prozessinstanzen auf der Grundlage einer einheitlichen Datenbasis

Rico Thomanek, Christian Roschke und Benny Platte

Hochschule Mittweida, D-09648 Mittweida, rthomane@hs-mittweida.de

Zusammenfassung Die Videoüberwachung von sicherheitsrelevanten Bereichen wird immer häufiger eingesetzt. Auswertung und das Erkennen von problematischen Szenen sind jedoch durch personelle Ressourcen begrenzt. Einen Lösungsansatz bieten automatisierte Systeme, welche die Bewegungsabläufe von Objekten verfolgen und detektieren sollen. Zur automatisierten Aktivitätserkennung sind mehrere Teilschritte notwendig, die optimal miteinander interagieren müssen. Für die Erfüllung der benötigten Teilaufgaben stehen aktuell verschiedene State-of-the-Art-Frameworks zur Verfügung. Ziel der vorliegenden Arbeit ist die Bestimmung einfacher Aktivitäten in Überwachungsvideos, basierend auf dem Einsatz heterogener Frameworks und deren Verschaltung, unter Verwendung einer Management-Umgebung.

Schlagwörter Aktivitätserkennung, Objektverfolgung, Schnittstellen, Datenaustauschformate

1 Einleitung

In den vergangenen Jahren stieg die Zahl der Überwachungskameras weltweit an. Durch CCTV-Kameras (Closed Circuit Television) werden ständig immer größere Mengen an Bilddaten erfasst, welche aufgrund begrenzter personeller Ressourcen üblicherweise erst nach Auftreten eines bestimmten Ereignisses auf Anzeichen einer problematischen Aktivität überprüft werden. Auch im Bereich der Verkehrssicherheit (z.B. Überwachung von Verkehrsknotenpunkten) oder anderer sensibler Bereiche wächst der Wunsch, Videos aussagekräftig und ressourcenschonend auszuwerten, denn die manuelle Aktivitätsüberwachung

erfordert ein hohes Maß an Konzentration, was wiederum schnell zu Ermüdungserscheinungen führen kann. Dieser Effekt wird als „unaufmerksame Blindheit“ bezeichnet, was bewirkt, dass bei hoher Konzentration auf ein Objekt oder eine Tätigkeit, unerwartete Objekte selbst im Betrachtungsmittelpunkt nicht wahrgenommen werden [1]. Einen Ansatz zur Lösung bieten automatisierte Systeme, die im Sinne der „prädiktiven Polizeiarbeit“ in der Lage sind, Bewegungsabläufe von Objekten einschließlich Personen zu erfassen und auszuwerten.

Im Bereich der Videoüberwachung leidet das zugrunde liegende Material unter Problemen im Zusammenhang mit der Komplexität der realen Umgebung. So existiert beispielsweise eine starke Variation in den Perspektiven der Aufzeichnungskameras, den Abständen zu relevanten Objekten und der Qualität der einzelnen Aufnahmen. Zudem wird die Lokalisierung und deren Verfolgung in Mehrkamerasystemen durch fehlende Information über das Weltkoordinatensystem und die eigene Kameraposition erschwert. Obwohl für eine nur auf Bilddaten basierte Analyse, bezüglich der im Video enthaltenen Objekte und deren Merkmale, verschiedene Implementationen von Algorithmen (z.B. Mask R-CNN, Faster R-CNN, RetinaNet, ...) zur Verfügung stehen, ist eine direkte Verkettung, aufgrund mangelnder Schnittstellen, zwischen verschiedenen Frameworks nicht möglich.

Die automatische Aktivitätserkennung setzt sich aus verschiedenen Teilaufgaben zusammen, welche ausgehend von der Extraktion der Einzelframes, über die Objekterkennung, deren Verfolgung bis zur Detektion der Aktivität reicht. Hierbei wird deutlich, dass diese verschiedenen Systemblöcke über definierte Schnittstellen ihre Dienste zur Verfügung stellen müssen. Der Beitrag dieser Arbeit beschreibt den Aufbau der in Abbildung 7.1 dargestellten Prozesskette zur Erkennung einfacher Aktivitäten in Videos. Als zentrale Verwaltungsinstanz dient ein selbst entwickeltes Management-System (EMSML-Evaluation- und Management-System für Maschinelles Lernen) für Machine-Learning-Frameworks, welches im Rahmen der internationalen Evaluationskampagne TRECVID (Maryland USA) entwickelt wurde [2, 3]. Um eine Interoperabilität des Systems zu gewährleisten, steht der Einsatz etablierter und möglichst standardisierter Technologien für die Interprozesskommunikation im Vordergrund. Mittelpunkt des Systems bildet eine zentrale Datenbank, die von den Frameworks als Datenbasis und Datenspeicher verwendet werden kann. Ziel ist es, die Prozesskette

so zu gestalten, dass ein Einsatz in zukünftigen Echtzeitumgebungen gewährleistet ist. Zu diesem Zweck sollen alle Komponenten autonom handeln und ihre Daten über eine zentrale Stelle bereitstellen. Die Kommunikation zwischen den Instanzen erfolgt über standardisierte Datenaustauschformate und Protokolle.



Abbildung 7.1: Realisierte Prozesskette zur Aktivitätserkennung

2 Systemarchitektur

Im Kontext der Szenenerkennung existieren bereits Lösungen, die ausgewählte Aktivitäten detektieren können [4–7]. Nachteilig ist jedoch die Proprietät der Systeme, was dazu führt, dass neue Aktivitäten bzw. bessere Algorithmen nur schwer integrierbar sind. Bezüglich dieses Nachteils soll die zu entwickelnde Prozesskette den einfachen und aufwendungsarmen Austausch von einzelnen Blöcken ermöglichen. Basierend auf diesem Ansatz ist das Management-System für verschiedenste Anwendungsfälle im Kontext des maschinellen Lernens einsetzbar.

2.1 Das verwendete Management-System

Ziel der EMSML-Umgebung ist die Dezentralisierung der einzelnen Komponenten und deren Vernetzung über standardisierte Schnittstellen. Hierdurch ist eine Lastverteilung der verschiedenen Prozesse auf mehrere Netzwerkknoten möglich. Ebenfalls sind somit parallelisierte Prozessabarbeitungen realisierbar, womit beispielsweise die Objektdetektion in mehreren tausend Bildern auf mehrere Bearbeiter verteilt werden kann. Eine Übersicht über die EMSML-Umgebung ist in Abbildung 7.2 dargestellt und beinhaltet die nachfolgend aufgeführten Kernfunktionalitäten.

Speicherung und Verwaltung: Die EMSML-Umgebung ermöglicht die zentrale Speicherung und Bereitstellung aller benötigten Rohdaten

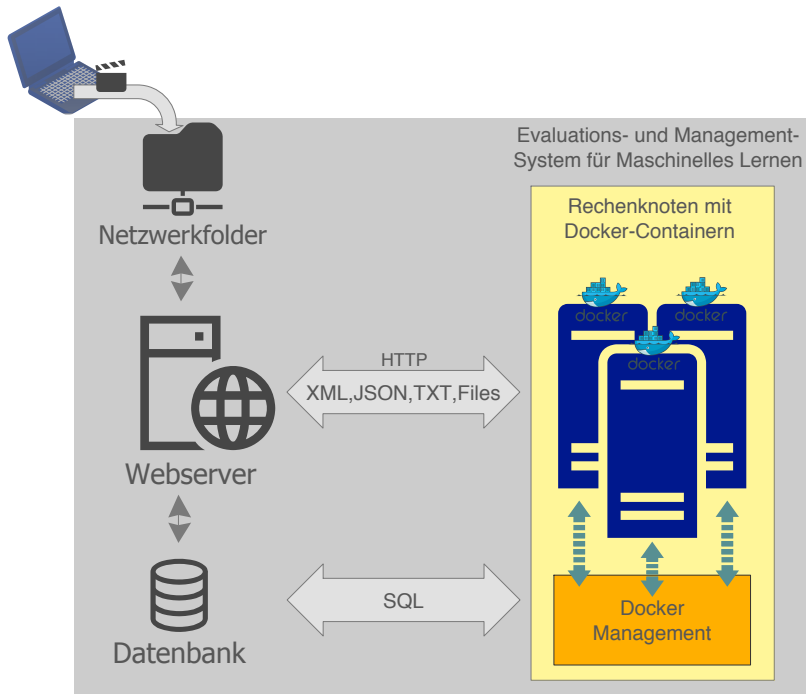


Abbildung 7.2: Systemumgebung von EMSML

(Bilder, Videos) über das LAN (Local Area Network). Für die Speicherung der Dateien stehen die Protokolle CIFS (Common Internet File System) und SMB (Server Message Block) zur Verfügung, wodurch ein systemunabhängiger, netzwerkbasierter Dateitransfer für alle gängigen Betriebssysteme zur Verfügung steht. Die verarbeitenden Frameworks haben darüber hinaus die Möglichkeit, per *HTTP* auf die gespeicherten Bilder und Videos zuzugreifen, wodurch keine Notwendigkeit besteht, die Rohdaten für die Verarbeitung auf den Rechenknoten zu hinterlegen. Hierfür muss ein *GET-Request* mit den Parametern *from* und *to* an den in EMSML integrierten Webserver übermittelt werden. Die Parameter *from* und *to* definieren die Anzahl der gewünschten Dateien, de-

ren Metadaten im zugehörigen Response als *XML* (Extensible markup Language) oder Text zur Verfügung gestellt werden.

Die durch die Frameworks extrahierten Ergebnisse können in Form von *JSON* (Javascript Object Notation) oder *XML* in EMSML importiert werden. Alternativ besteht die Möglichkeit, die Ergebnisse während der Verarbeitung mittels SQL (Structured Query Language) in der Datenbank zu speichern. Als Datenbankmanagement-System wird PostgreSQL eingesetzt.

Prozesssteuerung: Die benötigten Frameworks können in Form von Docker-Containern verwaltet und gesteuert werden. Docker-Container ermöglichen die Bereitstellung einer stabilen Ablaufumgebung jedes einzelnen Frameworks. Vorteilhaft ist hierbei die problemlose Verteilung auf neu hinzugefügte Rechenknoten. EMSML nutzt als Management-Umgebung die unter der zlib-Lizenz befindliche Software *Portainer*. Diese liefert ein webbasiertes Interface zur Verwaltung verschiedenster Rechenknoten und deren Docker-Container. Das User-Interface bietet u.a. die nachfolgend aufgeführten Kernfunktionalitäten.

- Import, Export und Erstellen von Docker-Images
- Container starten, beenden, löschen, hinzufügen und verwalten
- Hinzufügen oder Löschen von Services auf einem Cluster
- Hinzufügen oder Löschen von Service-Gruppierungen
- Nutzerverwaltung mit Teamzuordnung
- Verwaltung von Docker-Endpoints

Auf die verwalteten Docker-Container kann darüber hinaus in *Portainer* direkt eine SSH-Verbindung zu den verwalteten Containern hergestellt werden.

Vor- und Weiterverarbeitung: Videos können mithilfe eines Docker-Containers in Frames zerlegt werden. Hierbei werden alle Einzelbilder gespeichert, in der Datenbank referenziert und stehen per *HTTP*-Download den Frameworks für weitere Verarbeitungen zur Verfügung.

Neben den Rohdaten können auch bereits gespeicherte Framework-Ergebnisse zur Weiterverarbeitung in Form von *JSON* und *XML* abgerufen werden. Somit kann beispielsweise ein Framework zur Objektverfolgung die zuvor detektierten Objekte und deren Positionsinformation als Verarbeitungsbasis verwenden.

Analyse und Evaluation: Die in EMSML gespeicherten Frameworkdaten aus der Domäne des Maschinellen Lernens, können wiederum in den standardisierten Datenaustauschformaten *JSON* und *XML* für deren Analyse bzw. deren Evaluation bezogen werden. Alternativ ist ein direkter Datenbankzugriff zur Realisierung einer personalisierten SQL-Abfrage möglich.

2.2 Verwendete Frameworks

Für die Bestimmung einfacher Aktivitäten sollen die nachfolgend dargestellten, frei verfügbaren bzw. selbst implementierten Frameworks mit der EMSML-Umgebung in Kombination gesetzt werden. Dies erfordert teilweise die Anpassung des vorhandenen Quellcodes.

Detectron ist ein von Facebook entwickeltes Framework, basierend auf dem Deep Learning Framework *Caffe2* und bietet eine hochwertige sowie eine leistungsstarke Codebasis für die Objekterkennung. Verschiedene Algorithmen wie Mask R-CNN (Region-based Convolutional Neural Network), Faster R-CNN (Faster Region-based Convolutional Network), RPN (Region-based Proposal Network), Fast R-CNN (Fast Region-based Convolutional Network) und R-FCN (Region-based Fully Convolutional Network) sind implementiert. Darüber hinaus können mehrere Backbone-Netzwerkarchitekturen wie ResNet50, ResNet101, ResNet152, VGG16 oder Feature Pyramid Networks verwendet werden. Facebook bietet auch eine große Auswahl an trainierten Modellen zur weiteren Verwendung mit dem *Detectron* "Model Zoo Datensatz". *Detectron* soll für die Detektion von Objekten und deren Begrenzungsrahmen eingesetzt werden. Zusätzlich zu den x - y -Koordinaten des Startpunktes werden neben der Objektklassifizierung auch die Höhe und Breite der Box gespeichert. Die jedem Bild zugeordneten Klassen stammen aus dem Coco-Datensatz. [8]

YOLO9000 ist ein Netzwerk zur Echtzeitklassifizierung und -

lokalisierung von Objekten innerhalb eines Bildes und wurde mit über 9000 verschiedenen Objektkategorien trainiert. Im Gegensatz zu einem stapelbasierten und schrittweisen Ansatz zur Objekterkennung wird die Objekterkennung bei *YOLO9000* als ein einziges Regressionsproblem verarbeitet. Zur Erhöhung der Objektdetektionsleistung wird auch *YOLO9000*, neben *Detectron*, zur Objektdetektion verwendet. Auch hier werden die Objektklasse und die Werte des Begrenzungsrahmens gespeichert. [9]

Zur **Objektverfolgung** wird ein eigen implementiertes Framework verwendet. [10] Es stellt die durch *YOLO9000* und *Detectron* fusionierten Objekte in eine zeitliche Relation, basierend auf der euklidischen Distanz und der zugehörigen Objekt-Historie in vorherigen Frames. Als Ergebnis wird jedes getrackte Objekt in der Datenbank mit einer Objekt-ID versehen.

Für die **Aktivitätserkennung** sollen die im Rahmen der internationalen Evaluationskampagne TRECVID (Maryland USA) im Task „ActEV“ spezifizierten Aktivitäten „Opening“, „Closing“, „Entering“, „Exiting“, „Vehicle_turning_left“, „Vehicle_turning_right“ und „Vehicle_u_turn“ detektiert werden. Für deren Bestimmung wurden einfache Heuristiken implementiert.

Obwohl die EMSML-Umgebung wesentliche API's für die System-Kommunikation zur Verfügung stellt, bieten die Frameworks *Detectron* und *YOLO9000* keine entsprechenden Schnittstellen an. Eine Anpassung des Quellcodes ist somit unumgänglich.

Integration von *Detectron*

Detectron ist ein in Python implementiertes Framework. Somit stehen Klassen für *HTTP*, zur *XML*- und *JSON*-Verarbeitung sowie zur Datenbankkommunikation zur Verfügung. Für die Integration in EMSML wurden die nachfolgend beschriebenen Funktionalitäten hinzugefügt.

Dynamischer Bild-Download. Standardmäßig bietet *Detectron* die Möglichkeit, alle Bilder eines lokalen Ordners im Stapel zu verarbeiten. Der Bild-Download vom Webservice und die anschließende Klassifizierung ist jedoch nicht Bestandteil der Implementierung. Diesbezüglich wurde eine *HTTP*-Komponente hinzugefügt. Durch Hin-

zufügen weiterer Startparameter kann der Download-Modus aktiviert und ein URL (Uniform Resource Locator) bei der Ausführung von *Detectron* übergeben werden. Der übergebene URL liefert als Ressource ein Text-Dokument, in dem die URLs der zu verarbeitenden Bilder aufgeführt sind, welche anschließend mittels der *HTTP*-Komponente geladen und klassifiziert werden. Die Anzahl der Bilder kann dabei im *HTTP*-Request festgelegt werden. Standardmäßig werden 100 Bilder angefordert.

Export der Ergebnisse. Für die sofortige Speicherung der Ergebnisse wurde eine direkte SQL-Kommunikation mit der EMSML-Datenbank implementiert. Zur Reduzierung des Protokoll-Overheads werden die Ergebnisse von 100 klassifizierten Bildern im Block an die Datenbank übertragen.

Integration von *YOLO9000*

YOLO9000 wurde in *C* implementiert. Basierend auf dieser Grundlage gestaltet sich die Integration von Parsern für *JSON* oder *XML* sowie die Datenbank-Kommunikation schwieriger. Auch eine *HTTP*-Kommunikation ist aufwändiger zu realisieren.

Dynamischer Bilder-Download. Bezugnehmend auf die erwähnten Gründe wurde der *YOLO9000*-Quellcode nur um die Funktion des Bilder-Downloads erweitert. Hierfür muss beim Start von *YOLO9000* das Text-Dokument mit den zugehörigen Bilder-URLs übergeben werden. Der Download des Text-Dokumentes wurde hierbei durch das Linux-Kommandozeilenprogramm „*wget*“, außerhalb von *YOLO9000*, realisiert. *YOLO9000* interpretiert anschließend diese Datei, lädt die Bilder und klassifiziert sie. Anschließend wird *YOLO9000* beendet. Zur Realisierung der automatischen Klassifizierung aller Bilder kann mittels EMSML ein Shell-Skript für die Ausführung von *YOLO9000* erstellt werden. Neben der Aufteilung der zu bearbeitenden Bilder kann weiterhin die Anzahl der *YOLO9000*-Instanzen übergeben werden. Dies ermöglicht eine parallelisierte Verarbeitung.

Export der Ergebnisse. EMSML bietet neben der SQL-Schnittstelle auch den Daten-Import als *JSON* oder *XML*. Die Ergebnisse wurden in Form von *JSON*-Dateien im lokalen Filesystem gespeichert. Anschlie-

ßend konnten diese Dateien in EMSML importiert werden.

Integration der Objektverfolgung

Das selbst implementierte Framework zur Verfolgung der klassifizierten Objekte über mehrere Frames, bietet bereits eine Schnittstelle zum JSON-Import. Das hierfür benötigte Dokument kann in EMSML mittels *HTTP-GET*-Request geladen werden.

Durch Angabe der Video-ID und der Objektklasse wird ein JSON-Dokument erzeugt. Es enthält alle erkannten Objekte, inklusive ihrer Koordinaten, Größe und dem URL zum zugehörigen Frame. Eine automatische Verarbeitung aller Videos kann ebenfalls mittels Shell-Skript realisiert werden, welches durch EMSML für die zu verarbeitenden Videos und zu verfolgende Objektklasse erstellt wird. Auch hier kann die Anzahl der parallel verarbeitenden Instanzen berücksichtigt werden.

Aktivitätserkennung

Die Heuristiken zur Szenenerkennung werden für die Aktivitäten „Opening“, „Closing“, „Entering“ und „Exiting“ direkt in der EMSML-Datenbank ausgeführt, was zu sehr kurzen Verarbeitungszeiten führt. Hierbei werden die Begrenzungsrahmen bezüglich ihrer Größenänderung und Überschneidung mit anderen Objekten analysiert. In Abhängigkeit aufeinander folgender Bedingungen wird anschließend auf die Aktivität geschlossen. Am Beispiel der „Entering“-Aktivität werden die detektierten Objekt-Positionen, entsprechend ihres zeitlichen Verlaufs, bezüglich der Überschneidung mit anderen Objekten (z.B. Tür, Fahrzeug) verglichen. War das letzte Vorkommen eines Objektes mit einer Überschneidung verbunden, so wird auf die Aktivität „Entering“ geschlossen.

Für die Aktivitäten „Vehicle_turning_left“, „Vehicle_turning_right“ und „Vehicle_u_turn“ wurden die Richtungsvektoren der zugehörigen Objekte mittels Python ermittelt. Als Kommunikationsschnittstelle wurde SQL verwendet, sodass die Objektpositionen geladen, Richtungsvektoren berechnet und direkt in der Datenbank gespeichert werden. Zur anschließenden Aktivitätsbestimmung wurden die Richtungsvektoren über aufeinanderfolgende Frames von einem Startpunkt aus akkumuliert. Bei Überschreitung eines definierten Schwellwertes liegt

ein entsprechendes Bewegungsmuster vor. Eine Aktivität gilt immer als abgeschlossen, sobald sich der Richtungswinkel des Vektors entgegen der aktuellen Richtung bewegt oder das letzte Bild des zu verfolgenden Objektes erreicht ist. Der minimale Winkel, der für die Links- und Rechtsbewegung definiert wurde, ist $\pm 40^\circ$. Um eine Links- oder Rechtskurve zu erkennen, müssen mindestens zwei aufeinanderfolgende als links- oder rechts-klassifizierte Vektoren vorhanden sein. Für eine Drehbewegung wurde der minimale kumulierte Winkelbetrag auf $\pm 170^\circ$ eingestellt. Ein „Vehicle_u_turn“ kann nur ausgelöst werden, wenn diese minimale Summe der Winkel überschritten wird und zuvor mindestens zwei links- oder rechtsorientierte Vektoren aufgetreten sind.

Die ermittelten Aktivitäten wurden ebenfalls mittels SQL in der Datenbank gespeichert. Mittels *HTTP* können anschließend die erkannten Aktivitäten in Form von *JSON* aus der EMSML-Umgebung exportiert werden.

2.3 Framework-Management

Alle eingesetzten Frameworks wurden als eigenständige Docker-Images realisiert und werden über *Portainer* verwaltet. Voraussetzung für das Management eines Rechenknotens durch *Portainer* ist die Installation von *Docker* und die Aktivierung der *Docker-Remote-API*. Standardmäßig erfolgt die Kommunikation mit dem *Docker-Daemon* über einen nicht vernetzten *UNIX-Socket*. Optional ist jedoch auch das Anwendungsprotokoll *HTTP* nutzbar. Die Aktivierung erfolgt über eine *Docker-Konfigurationsdatei*. Standardmäßig verfügt die Remote-API über keine Authentifizierung. Eine Sicherung über TLS und Client-Zertifikatsverifizierung ist jedoch einstellbar und wird laut *Docker-Dokumentation* empfohlen. *Portainer* ermöglicht anschließend den Import des Client-Zertifikats.

Die verfügbaren *Docker-Images* können auf allen verwalteten Rechenknoten installiert und gestartet werden. Zur Lastverteilung besteht weiterhin die Möglichkeit, ressourcenintensive Frameworks gleichzeitig auf verschiedenen Rechenknoten auszuführen. Das in EMSML integrierte Session-Management stellt daraufhin jedem Rechenknoten nur einen Teil der Rohdaten zur Verfügung. Die Aufteilung der Rohdaten erfolgt linear anhand der Anzahl der verarbeitenden Framework-Instanzen. Die Verarbeitungszeit kann dadurch um die Anzahl der In-

stanzen verkürzt werden.

Über den in *Portainer* integrierten SSH-Zugang kann die Ausführung der Frameworks überwacht werden.

3 Ergebnisse und Diskussion

Die in Abbildung 7.1 dargestellte Prozesskette konnte mit der EMSML-Umgebung erfolgreich umgesetzt werden. Für alle verwendeten Frameworks wurden eigenständige Docker-Images erzeugt. Der verwendete Datensatz ist VIRAT V1 [11], der aus 455 verschiedenen Videosequenzen aus dem Bereich der Videoüberwachung besteht.

Trotz der verfügbaren standardisierten Schnittstellen zur Interprozesskommunikation war eine programmatische Anpassung der Frameworks *YOLO9000* und *Detectron* notwendig. Standardmäßig ermöglichen diese Frameworks nur die Bearbeitung von lokal verfügbaren Bildern. Auch der Datenexport in standardisierte Formate ist nicht Bestandteil der Frameworks. Eine Einarbeitung in den Quellcode der einzusetzenden Frameworks und dessen Anpassung ist demzufolge unumgänglich. Der Implementationsaufwand wird dabei durch die verwendete Programmiersprache bestimmt. Viele aktuelle Frameworks nutzen Python zur Implementierung, was die Adaption der benötigten Schnittstellen, aufgrund vieler verfügbarer Klassen, entsprechend einfach gestaltet.

EMSML bietet neben dem Datenimport in Form von *JSON* und *XML* auch die direkte Datenspeicherung mittels *SQL*. Der Vorteil liegt hierbei in der sofortigen Bereitstellung der Ergebnisse, welche wiederum von weiteren Frameworks sofort verarbeitet werden können. Gerade für die Echtzeit-Aktivitätserkennung ist diese Vorgehensweise wesentlich. Nachteilig ist hierbei jedoch die zusätzliche Verwaltung der Zugriffsrechte auf Datenbanktabellen, um keine ungewollten Datenbankänderungen zu riskieren. Die Einführung einer Datenbankabstraktionsschicht in die EMSML-Umgebung wäre somit vorteilhaft.

Die Ableitung der Aktivitäten erfolgt ausschließlich anhand der in der Datenbank gespeicherten, zuvor ermittelten, Frameworkergebnisse, was zu sehr kurzen Detektionszeiten führt. Zeitintensivere Vorverarbeitungen können durch die Architektur von EMSML parallelisiert werden. *Detectron* benötigt beispielsweise ca. 0.2 Sekunden für die Ob-

jektklassifizierung eines Bildes. Für die Echtzeit-Objektklassifizierung wären demzufolge 6 *Detectron*-Instanzen notwendig, welche mittels Dockerizing und die verteilte Systemarchitektur von EMSML schnell und einfach realisiert werden kann.

4 Ausblick

Die Aktivitätserkennung basiert aktuell auf einfachen Heuristiken, die sich auf die Auswertung räumlicher Objektbewegungen beschränken. Daher sollen zukünftig weitere Frameworks in die EMSML-Umgebung integriert werden. Hierzu zählt zum Beispiel *OpenPose* [12], welches es ermöglicht, die Gliedmaßen von Personen und deren Position zu extrahieren. Aus diesen gewonnenen Daten sollen anschließend zeitliche Merkmale extrahiert werden, um eine markante Bewegung und die damit verbundene Aktivität zu bestimmen.

Ziel ist weiterhin die echtzeitfähige Verschaltung der Frameworks unter Verwendung der EMSML-Umgebung. Hierfür muss der in EMSML rudimentär verfügbare Session-Layer um Funktionalitäten zur Sitzungssteuerung, Parameteraushandlung und Fehlerbehandlung erweitert werden.

Danksagung

Diese Forschung wurde aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01PL17062 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor_innen.

Literatur

1. D. J. Simons and C. F. Chabris, "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events," *Perception*, vol. 28, no. 9, pp. 1059–1074, Sep. 1999.
2. G. Awad, A. Butt, K. Curtis, Y. L. Lee, J. Fiscus, A. Godil, A. Delgado, A. F. Smeaton, Y. Graham, W. Kraaij, G. Quénot, J. Magalhaes, D. Semedo, and S. Blasi, "Trecvid 2018: Benchmarking video activity detection, video captioning and matching, video storytelling linking and video search," in *Proceedings of TRECVID 2018*. NIST, USA, 2018.
3. R. Thomanek, C. Roschke, R. Manthey, B. Platte, T. Rolletschke, M. Heinzig, M. Vodel, D. Kowerko, S. Kahl, F. Zimmer, M. Eibl, and M. Ritter, "University of applied sciences mittweida and chemnitz university of technology at trecvid 2018," Gaithersburg, Maryland, USA, Nov. 2018.
4. J. C. San Miguel, J. Bescós, J. M. Martínez, and Á. García, "DiVA: A Distributed Video Analysis Framework Applied to Video-Surveillance Systems," in *2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services*, 2008.
5. Y. Song and I. Kim, "DeepAct - A Deep Neural Network Model for Activity Detection in Untrimmed Videos." *JIPS*, 2018.
6. A. Jalal, S. Kamal, and D. Kim, "A Depth Video-based Human Detection and Activity Recognition using Multi-features and Embedded Hidden Markov Models for Health Care Monitoring Systems." *IJIMAI*, 2017.
7. W. Sultani, C. Chen, and M. Shah, "Real-world Anomaly Detection in Surveillance Videos," *arXiv.org*, 2018.
8. R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, and K. He, "Detectron," <https://github.com/facebookresearch/detectron>, 2018.
9. J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," *arXiv.org*, p. arXiv:1612.08242, Dec. 2016. [Online]. Available: <http://arxiv.org/abs/1612.08242v1>
10. R. Thomanek, C. Roschke, B. Platte, R. Manthey, T. Rolletschke, M. Heinzig, M. Vodel, F. Zimmer, M. Eibl, and M. Ritter, "A Scalable System Architecture for Activity Detection with Simple Heuristics," *IEEE WACV*, 2019.
11. S. Oh, A. Hoogs, A. Perera, N. Cuntoor, C.-C. Chen, J. T. Lee, S. Mukherjee, J. K. Aggarwal, H. Lee, L. Davis, E. Swears, X. Wang, Q. Ji, K. Reddy, M. Shah, C. Vondrick, H. Pirsiavash, D. Ramanan, J. Yuen, A. Torralba, B. Song, A. Fong, A. Roy-Chowdhury, and M. Desai, "A large-scale benchmark dataset for event recognition in

- surveillance video," in *CVPR 2011*. IEEE, jun 2011. [Online]. Available: <https://doi.org/10.1109/cvpr.2011.5995586>
12. Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," *CoRR*, vol. abs/1611.08050, 2016. [Online]. Available: <http://arxiv.org/abs/1611.08050>

Lexikonbasierte Sentimentanalyse von Online-Kundenbewertungen

Claudia Hösel

Hochschule Mittweida, D-09648 Mittweida, hoesel@hs-mittweida.de

Zusammenfassung Online-Kundenbewertungen stellen nicht nur eine wichtige Entscheidungsgrundlage für Kunden dar, sie werden von E-Commerce-Anbietern auch als Informationsquelle genutzt, um die Zufriedenheit ihrer Kunden zu analysieren. Um der Unübersichtlichkeit von Bewertungskommentaren zu begegnen, werden schriftliche Reviews in vielen Bewertungssystemen zusätzlich durch Bewertungssterne dargestellt. Die automatisierte Extraktion von Meinungen und Stimmungen aus schriftlichen Bewertungskommentaren ist Gegenstand vieler Studien. Die vorliegende Studie untersucht, inwieweit Bewertungssterne den Inhalt von Bewertungskommentaren widerspiegeln und stellt einen ersten Arbeitsstand vor. Die Datengrundlage bildeten 5.000 Kundenbewertungen, die mit einer lexikonbasierten Sentimentanalyse ausgewertet wurden, um auf Basis der ermittelten Polarität die Bewertungspunkte vorherzusagen. Die Klassifikation erfolgte mit traditionellen Standardverfahren wie k-Nearest Neighbor, Naïve Bayes und Random Forest, wobei die beste Vorhersagegenauigkeit mit Random Forest erzielt wurde.

Schlagwörter Sentimentanalyse, Opinion-Mining, Klassifikation, Natural Language Processing

1 Einleitung

Kundenbewertungen stellen im Business-to-Consumer E-Commerce eine wichtige Informationsquelle – sowohl für Kunden als auch für E-Commerce-Anbieter – dar. Auf Ebene der Kunden ist der Social Proof von Kundenbewertungen eine wichtige Entscheidungshilfe bei Kaufentscheidungen [1]. Rund zwei Drittel der Online-Käufer lesen Kundenbewertungen vor dem Kauf von Produkten in Online-Shops [2]. Für E-Commerce-Anbieter sind Kundenbewertungen ein wichtiger Bestandteil beim Verkauf von Produkten und Dienstleistungen, da sich hieraus unter anderem Rückschlüsse auf die Kundenzufriedenheit ziehen lassen. E-Commerce-Anbieter, wie *Amazon.com, Inc.* [3], betreiben daher – zumeist plattformabhängige – Bewertungssysteme, um Stimmungen und Emotionen aus den Rezensionen zu extrahieren. Eine intellektuelle Extraktion dieser Informationen ist insbesondere bei großen E-Commerce-Anbietern aufgrund der Vielzahl von Bewertungen kaum möglich. Es bedarf folglich einer automatisierten Auswertung, die mittels Opinion-Mining realisiert wird. Opinion-Mining, ein Teilgebiet des Text Mining, beschäftigt sich mit der automatisierten Extraktion und Bewertung von Meinungen aus Texten und bedient sich bei der Stimmungserkennung verschiedener Techniken [4]. Eine Opinion-Mining-Technik ist die Sentimentanalyse. Die Sentimentanalyse greift auf verschiedene Natural Language Processing (NLP)-Methoden, wie Tokenisierung oder Stemming, zurück, um die Stimmung eines Textes oder die emotionale Einstellung zu einem im Text enthaltenen Objekt zu analysieren [5].

2 Forschungsstand

Ein grundlegendes Problem der Sentimentanalyse besteht in der Kategorisierung der Stimmungspolarität. Natürlich-sprachliche Texte können beispielsweise Valenzverschiebungen – etwa durch Verneinung oder Intensivierung – enthalten, die von Menschen zumeist problemlos verstanden werden, von Computersystemen jedoch nicht [6]. Neuere Forschungsarbeiten versuchen diesem Problem durch eine lexikonbasierte Vorverarbeitung, zumeist unter Nutzung bestehender Sentimentwörterbücher, zu begegnen.

Prakoso et al. (2018) untersuchen in ihrer Studie die Auswirkungen auf die Genauigkeit der Sentimentklassifikation, indem sie Sentimentanalysen mit und ohne lexikonbasierte Vorverarbeitung durchführen und vergleichen. Für die lexikonbasierte Vorverarbeitung verwenden sie das WordNet-Dictionary. Für die Klassifizierung der Polarität, die Autoren unterscheiden positiv und negativ, verwenden sie die überwachten maschinellen Lernalgorithmen Decision Tree, k-Nearest Neighbor und Support Vector Machine. Die Autoren stellen fest, dass die Sentimentanalyse mit lexikonbasierter Vorverarbeitung bei allen Klassifikationsmodellen eine höhere Genauigkeit erzielt. [7]

Alkalbani et al. (2017) untersuchen in ihrer Arbeit Kundenbewertungen zu Cloud Services und nutzen hierfür ebenfalls einen lexikonbasierten Ansatz. Die Sentimentanalyse realisieren die Autoren mit dem Data-Mining-Tool RapidMiner, wobei sie für die lexikonbasierte Vorverarbeitung die Extension AYLIEN Text Analysis und das SentiWordNet Dictionary nutzen. Bei den Polaritätsgraden unterscheiden die Autoren zwischen positiv, negativ und neutral. Für die Klassifikation nutzen sie k-Nearest Neighbor, Naïve Bayes, Random Forest Model sowie Random Tree. Mit dem Random Forest Model erzielen die Autoren bei den drei Polaritätsgraden die höchste Genauigkeit (accuracy: 97.06%), gefolgt von k-Nearest Neighbor (accuracy: 95.38%). [8]

Um die Polarität von Online-Produktbewertungen zu klassifizieren kombinieren Fang & Zahn (2015) eine lexikonbasierte Sentimentanalyse mit ganzzahligen Ratinginformationen. Die von den Kunden im Rahmen der Bewertungen zusätzlich als Sterne bzw. Punkte vergebene Informationen überführen sie in ein ternäres Klassifikationsproblem mit den Polaritätsgraden positiv, negativ und neutral. Die schriftlichen Reviews betrachten die Autoren als binäres Klassifikationsproblem und unterscheiden positive und negative Polarität. Für die Klassifikation nutzen sie sowohl für die schriftlichen Reviews als auch für die sternenskalierten Bewertungen die überwachten Lernverfahren Support Vector Machine, Naïve Bayes und Random Forest Model. Während bei den schriftlichen Reviews, dem binären Klassifikationsproblem, das Random Forest Model die höchste Genauigkeit liefert, erzielen die Autoren bei dem ternären Klassifikationsproblem mit den Klassifikatoren Support Vector Machine und Naïve Bayes zwar bessere Ergebnisse, stellen jedoch eine insgesamt schlechte Performance bei der Erkennung der neutralen Klasse fest. [9]

Auch Lin et al. (2018) konstatieren für die neutrale Klasse eine geringe Vorhersagegenauigkeit. In ihrer Studie erstellen sie ein für den Software-Engineering-Bereich angepasstes Sentimentanalyse-Tool und nutzen Rekursive Neuronale Netzwerke für die Polaritätserkennung. Die Autoren stellen fest, dass das von ihnen erstellte Stimmungsanalyse-Tool unter Nutzung von Deep Learning trotz aufwendigem Trainingsprozesses negative Ergebnisse liefert. [10]

Auch weitere bestehende Forschungsarbeiten betrachten die Sentimentanalyse zumeist als binäres oder ternäres Klassifikationsproblem. Eine stärkere Skalierung der Polarität wird in bestehenden Ansätzen nur vereinzelt vorgenommen und bezieht sich zumeist auf ausgewählte Social-Media-Kanäle. So stellen El Alaoui et al. (2018) einen lexikonbasierten Ansatz für die Sentimentanalyse von Tweets vor, welcher sieben Polaritätsklassen – sehr positiv, mäßig positiv, leicht positiv, neutral, leicht negativ, mäßig negativ, stark negativ – unterscheidet. Das Sentiment der Tweets bestimmen sie über einen Datensatz ausgewählter positiver und negativer Hashtags, welche sie in ein Sentimentlexikon überführen. Für die Erstellung des Lexikons nutzen sie eine zufällige Teilmenge von 600 Tweets (50 pro Klasse) aus dem politischen Twitter-Korpus, die sie manuell klassifizieren. In die Berechnung ihres Sentiment-Scores, welcher den Polaritätsgrad eines Tweets bestimmt, beziehen sie unter anderem spezifische Funktionen des Microbloggingdienstes, etwa Re-Tweets oder Likes, ein. Ihren Ansatz testeten sie mit Google cloud prediction (accuracy: 66,66%) und Naïve Bayes (accuracy: 34,45%). [11]

Bestehende Forschungsarbeiten, die das Problem der Kategorisierung der Stimmungspolarität in Sentimentanalysen aufgreifen, kombinieren vielfach lexikonbasierte Ansätze mit maschinellen – zumeist überwachten – Lernalgorithmen und betrachten die Stimmungserkennung in der Regel als binäres oder ternäres Klassifikationsproblem. Bei der Stimmungserkennung von Kundenbewertungen erscheint eine stärkere Skalierung der Polarität jedoch sinnvoll, da bestehende – insbesondere plattformabhängige – Bewertungssysteme zumeist Metriken, wie schriftliches Review und ganzzahliges Ratingsystem, kombinieren. In diesem Kontext stellt sich die Frage, inwieweit die durch Kunden über Ratingsysteme vergebenen Punkte bzw. Sterne die in den schriftlichen Bewertungen geäußerten Meinungen abbilden.

Die vorliegende Studie greift dieses Problem auf und stellt einen

ersten Arbeitsstand vor. Ausgehend von der Annahme, dass plattformabhängige Bewertungssysteme von E-Commerce-Anbietern Bewertungskommentare mit einem fünfstufigen sternenskalierten Bewertungssystem kombinieren, wird die Stimmungserkennung als quinäres Klassifikationsproblem betrachtet. Für die Sentimentanalyse wird das Data-Mining-Werkzeug RapidMiner [12] verwendet. Die Kategorisierung der Stimmungspolarität erfolgt durch den Einsatz verschiedener NLP-Techniken in Kombination mit dem Sentimentwörterbuch SentiWordNet [13]. Für die Klassifikation werden die maschinellen überwachten Modelle k-Nearest Neighbor, Naïve Bayes und Random Forest implementiert, evaluiert und die Güte der Klassifikationen zwischen den Modellen verglichen.

3 Methodik

In dieser Studie wird ein Datensatz von „Kaggle“ [14] mit rund 400.000 Bewertungen von freigeschalteten Mobiltelefonen verwendet. Jedes dieser Geräte wurde auf *Amazon.com, Inc.* [3] verkauft und von Kunden dort auch bewertet. Der Datensatz umfasst die folgenden Attribute: „Product Name“, „Brand Name“, „Price“, „Rating“, „Reviews“, „Review Votes“. Im Rahmen der Vorverarbeitung werden nicht benötigte Attribute sowie fehlende Werte aus dem Datensatz entfernt. Überdies werden die Attributtypen in eine für die Klassifikatoren nutzbare Form umgewandelt. Beispielsweise wird das Attribut „Review“ vom Typ „polynomial“ in den Typ „Text“ umgewandelt. Der Datensatz ist im Hinblick auf das Attribut „Rating“ nicht ausgeglichen, was darauf hindeutet, dass für jede Bewertungsstufe eine unterschiedliche Anzahl von Bewertungen gesammelt wurde. Das Attribut „Rating“ basiert auf einem ganzzahligen sternenskalierten System, bei dem die höchste Bewertung fünf Sterne und die niedrigste Bewertung einen Stern annehmen kann. Die Verteilung der sternenskalierten Bewertungen stellt sich im vorliegenden Datensatz wie folgt dar: 71.167 Daten für einen Stern, 24.233 für zwei Sterne, 31.189 für drei Sterne, 60.335 für vier Sterne und 221.049 für fünf Sterne. Da dieses Ungleichgewicht zu einer Verzerrung bei der Klassifikation führen kann, wird der Datensatz im Hinblick auf das Attribut „Rating“ normiert. Aufgrund technischer Einschränkungen werden je Ratingstufe 1.000 Elemente verwendet.

3.1 Vorverarbeitung der Daten

In dieser Phase werden verschiedene Techniken des Natural Language Processing eingesetzt, um den Text zu bereinigen, zu strukturieren und in eine maschinenlesbare Form zu bringen. In der vorliegenden Studie wird dies unter Verwendung von RapidMiner 9.1.000 mit dem Operator „Process Documents from Data“ realisiert. Für die Polaritätserkennung wird die Erweiterung SentiWordNet 3.0.0 in RapidMiner implementiert. SentiWordNet ist eine öffentlich zugängliche lexikalische Ressource die für Opinion-Mining-Anwendungen entwickelt wurde [15]. In SentiWordNet werden stimmungstragende englische Substantive, Verben, Adjektive und Adverbien in Synsets zusammengefasst und durch semantische sowie lexikalische Beziehungen miteinander verknüpft [16]. Die stimmungstragenden Ausdrücke im Text werden identifiziert und verhältnisskaliert kodiert. Die im Preprocessing-Prozess verwendeten Operationen werden nachfolgend beschrieben:

- *Transform Cases*: Dieser Operator transformiert alle Zeichen der Token in Klein- oder Großbuchstaben. Wörter, die bis auf Groß- bzw. Kleinschreibung identisch sind, werden gleich behandelt. In dieser Studie werden alle Buchstaben in Kleinbuchstaben umgewandelt (lower case).
- *Tokenize*: Der Text wird in eine Folge von Token aufgeteilt. In dieser Studie werden alle durch Leerzeichen abgetrennten Wörter als Einheit betrachtet.
- *Filter Stopwords*: Um die Datenmenge zu reduzieren, werden Wörter, die keine Informationen enthalten, z.B. „an“, „the“, mit dem Operator Filter Stopwords eliminiert.
- *Filter Token (by length)*: Die Token werden, basierend auf einer festgelegten minimalen und maximalen Zeichenanzahl, gefiltert. Der Operator wird genutzt, um unnötige Token zu entfernen und hierdurch die Datenmenge weiter zu reduzieren. Token, die weniger als drei bzw. mehr als 25 Zeichen enthalten, werden entfernt.
- *Stemming (WordNet)*: Um die Dimensionen der Wörter zu reduzieren, werden verschiedene morphologische Varianten eines

Wortes auf ihren Wortstamm zurückgeführt. Mit dem WordNet-Dictionary werden Verben auf ihre Infinitivform und Pluralformen in Singular gebracht.

- *Extract Sentiment (English)*: Die Extraktion des Sentiments erfolgt auf Dokumentenebene. Hierfür wird das durchschnittliche Sentiment aller Token auf Grundlage eines Wörterbuchs ermittelt und ein Sentimentwert berechnet. Der Ausgabewert liegt zwischen -1 (Text ist sehr negativ) und +1 (Text ist sehr positiv). In dieser Studie werden Nomen, Verben, Adjektiven und Adverbien als stimmungstragende Wörter betrachtet und bei der Extraktion des Sentiments entsprechend berücksichtigt.
- *Open WordNet Dictionary*: Dieser Operator sucht den Pfad des Wörterbuchs, sodass die Operatoren „Stemming“ und „Extract Sentiment“ bei der Ausführung ihrer Operationen darauf zugreifen können.

3.2 Vektorraum-Modell und Begriffsgewichtung

Die Token des Dokuments werden verwendet, um einen Vektor zu erzeugen, der das Dokument numerisch darstellt und somit für mathematische Operationen nutzbar macht. Die Gewichtung der Terme erfolgt über das kombinierte Verfahren „Term Frequency - Inverted Document Frequency“ (TF-IDF). Dieses Verfahren berücksichtigt die Häufigkeitsverteilung von Termen im Korpus und gewichtet Terme anhand der Häufigkeit und Unterscheidungen [17].

3.3 Klassifikation

Die Kategorisierung von Sentimenten stellt im Wesentlichen ein Klassifizierungsproblem dar, bei dem die aus den Texten extrahierten Features für die Vorhersage des Sentiments genutzt werden. Um die Genauigkeit und Präzision der Vorhersage zu erhöhen, werden in dieser Studie drei nachfolgend beschriebene Klassifikationsmodelle verwendet und miteinander verglichen.

1. *k-Nearest Neighbor (k-NN)*: Der k-NN-Algorithmus bestimmt die Klasse von Trainingsbeispielen auf Grundlage

der k nächstgelegenen Nachbarn [18]. Die Nähe zu den k nächstgelegenen Nachbarn wird über verschiedene Abstandsmaße, z.B. euklidische Distanz, berechnet. Um Gleichstände zu vermeiden, sollte k vorzugsweise eine ungerade positive Zahl sein und zudem möglichst klein gehalten werden, da die Betrachtung einer großen Anzahl nächstgelegener Nachbarn - bei schlecht getrennten Klassen - zu einer schlechten Klassifikation führen kann. In dieser Studie wird der „ k -NN“-Operator von RapidMiner genutzt, um ein k -NN-Modell zu erzeugen. Es wird $k = 3$ verwendet; d.h. das Modell findet die drei nächstgelegenen Nachbarn und bestimmt das Sentiment des neuen Beispiels auf Grundlage der Ähnlichkeit dieser drei Nachbarn. Der Abstand zu den nächstgelegenen Nachbarn wird über die euklidische Distanz ermittelt.

2. *Naïve Bayes (NB)*: Der naïve Bayes-Algorithmus ist ein probabilistischer Klassifikator, der das Bayes-Theorem verwendet, um Objekte zu klassifizieren [19]. Der Algorithmus geht davon aus, dass zwischen den Prädiktoren eine starke (naive) Unabhängigkeit besteht, d.h. das Vorhandensein eines bestimmten Merkmals hat keinen Einfluss auf das andere. Dementsprechend wird nur eine geringe Menge an Trainingsdaten benötigt, um Abweichungen der für die Klassifizierung notwendigen Variablen abzuschätzen. In dieser Studie wird der „Naive Bayes (Kernel)“-Operator für die Modellerstellung genutzt, da dieser auch auf numerische Attribute anwendbar ist. Ein Kernel ist eine Gewichtungsfunktion, die nicht-parametrische Schätzverfahren verwendet. Als Schätzmodus für die Kerneldichte wird eine minimale Bandbreite von 0.1 sowie eine Kernelanzahl von zehn festgelegt.
3. *Random Forest (RF)*: Dieses Verfahren stellt eine Erweiterung von Entscheidungsbäumen dar und baut ein Klassifizierungsmodell in Form mehrerer Baumstrukturen auf. Eine zuvor definierte Anzahl von Bäumen wird dabei nach dem Prinzip des Bagging erstellt; d.h. die Vorhersagen aus den Bäumen werden miteinander kombiniert [20]. Die Klassifizierung erfolgt also durch wiederholte Anwendung von Entscheidungsbäumen, wobei nur eine zufällige Teilmenge von Attributen für jeden Durchlauf verfügbar

ist. In dieser Studie wird der Operator „Random Forest“ genutzt, um ein Random Forest Model zu erzeugen. Dieser Operator kann aus nominalen sowie aus numerischen Daten Entscheidungsbäume erstellen. Die Anzahl der Bäume wird in der Untersuchung auf 100 festgelegt und die Tiefe für jeden Baum auf zehn begrenzt.

3.4 Evaluation

Die Evaluation des Modells erfolgt durch eine zehnfache Kreuzvalidierung, die mit dem Operator „Cross Validation“ realisiert wird. Bei diesem Operator handelt es sich um einen verschachtelten Operator, der die zwei Teilprozesse Training und Testing durchführt. Der Datensatz wird hierfür in zehn Teilmengen gleicher Größe aufgeteilt, von denen eine für das Testing erhalten bleibt, die anderen neun werden für das Training verwendet. Der Prozess der Kreuzvalidierung wird zehnmal wiederholt, wobei jede der zehn Teilmengen genau einmal als Testdaten verwendet wird. Aufgrund des iterativen Trainings wird die Leistung des Modells unabhängig von einem bestimmten Testdatensatz bewertet. Die Leistung des Modells ergibt sich aus dem Mittel der zehn Ergebnisse.

4 Ergebnisse und Diskussion

In dieser Studie wird eine Sentimentanalyse mit lexikonbasierter Vorverarbeitung von Produktbewertungen mit dem Ziel durchgeführt, die von den Kunden abgegebenen ganzzahligen sternenskalierten Bewertungen auf Basis der schriftlichen Reviews vorherzusagen. Für das quinäre Klassifikationsproblem werden die Klassifikatoren k-Nearest Neighbor, Naïve Bayes und Random Forest eingesetzt und deren Genauigkeit verglichen. Der gesamte Prozess wird mit dem Data-Mining-Tool RapidMiner realisiert. Die Modelle werden über eine zehnfache Leave-one-out-Kreuzvalidierung evaluiert. Die Tabellen 8.1, 8.2 und 8.3 zeigen die Confusionmatrizen für alle Modelle.

4.1 k-Nearest Neighbor

Der Klassifikator k-Nearest Neighbor ordnet die Kundenbewertungen anhand des Sentiments mit einer Genauigkeit von 32,56% in die richtige Klasse ein. Die Vorhersagegenauigkeit der einzelnen Klassen weist für alle Klassen hohe Fehlerraten auf, was an den niedrigen Precision-Werten der einzelnen Klassen deutlich wird. Die höchste Fehlerrate und damit die geringste Vorhersagegenauigkeit weist Klasse 2 mit einer Precision von 25,29% auf. Der Recall-Wert von 67,70% zeigt jedoch, dass eine zu dieser Klasse gehörende Bewertung vom Klassifikator mit einer Wahrscheinlichkeit von 67,70% erkannt und richtig klassifiziert wird. Die geringste Fehlerrate weist Klasse 1 mit einer Precision von 49,89% auf. Allerdings liegt die Wahrscheinlichkeit, dass eine tatsächlich zu Klasse 1 gehörende Bewertung erkannt und richtig klassifiziert wird, bei nur 23,00%. Ähnliches gilt für Klasse 5: Diese wird zwar mit einer Genauigkeit von 40,95% vorhergesagt, entsprechende Bewertungen jedoch nur mit einer Wahrscheinlichkeit von 38,70% erkannt und richtig klassifiziert.

Tabelle 8.1: Confusionmatrix k-NN (accuracy: 32,56%)

	class 1	class 2	class 3	class 4	class 5	precision
pred. 1	230	101	62	37	31	49,89%
pred. 2	577	677	569	490	364	25,29%
pred. 3	49	65	179	97	92	37,14%
pred. 4	39	52	63	155	126	35,63%
pred. 5	105	105	127	221	387	40,95%
recall	23,00%	67,70%	17,90%	15,50%	38,70%	

4.2 Naïve Bayes

Naïve Bayes erzielt bei dem quinären Klassifikationsproblem eine Genauigkeit von 35,49%. Die niedrigen Precision-Werte der individuellen Klassen zeigen, dass die einzelnen Klassen nicht genau vorhergesagt

werden können. Jede Klasse weist eine hohe Fehlerrate bei der Vorhersagegenauigkeit auf. Die höchste Fehlerrate weist Klasse 5 mit einer Precision von 28,40% auf. Der zugehörige Recall-Wert von 84,10% zeigt jedoch, dass eine Bewertung, die zu Klasse 5 gehört, mit hoher Wahrscheinlichkeit erkannt und richtig klassifiziert wird. Die höchste Precision mit 49,54% wird bei Klasse 2 erzielt. Die Wahrscheinlichkeit, dass eine tatsächlich zu dieser Klasse gehörende Bewertung erkannt und richtig klassifiziert wird, liegt jedoch nur bei 16,72%. Bei Klasse 1 wird eine Precision von 49,12% erzielt. Die Wahrscheinlichkeit, dass eine in Klasse 1 gehörende Bewertung als solche erkannt und klassifiziert wird, beträgt 48,27%.

Tabelle 8.2: Confusionmatrix Naïve Bayes (accuracy: 35,49%)

	class 1	class 2	class 3	class 4	class 5	precision
pred. 1	473	246	137	64	43	49,12%
pred. 2	69	163	54	26	17	49,54%
pred. 3	55	58	161	54	29	45,10%
pred. 4	45	46	74	119	67	33,90%
pred. 5	338	462	559	721	825	28,40%
recall	48,27%	16,72%	16,35%	12,09%	84,10%	

4.3 Random Forest

Der Klassifikator Random Forest erzielt insgesamt eine Genauigkeit von 38,27%. Die einzelnen Klassen werden mit geringer Genauigkeit vorhergesagt. Die geringste Vorhersagegenauigkeit wird in diesem Modell bei Klasse 4 erzielt. Die Precision für diese Klasse liegt bei 31,75%. Überdies zeigt der Recall-Wert von 20,33%, dass eine zu Klasse 4 gehörende Bewertung nur mit geringer Wahrscheinlichkeit erkannt und richtig klassifiziert wird. Die höchste Vorhersagegenauigkeit in diesem Modell weist Klasse 1 mit einer Precision von 43,48% auf. Eine zu Klasse 1 gehörende Bewertung wird mit einer Wahrscheinlichkeit von 52,76% erkannt und richtig klassifiziert. Auffallend ist das Ergebnis von Klasse

5. Obgleich die Precision für diese Klasse mit 39,01% relativ gering ist, wird eine zu dieser Klasse gehörende Bewertung mit einer Wahrscheinlichkeit von 71,46% richtig erkannt und klassifiziert.

Tabelle 8.3: Confusionmatrix Random Forest (accuracy: 38,27%)

	class 1	class 2	class 3	class 4	class 5	precision
pred. 1	517	316	202	88	66	43,48%
pred. 2	140	232	151	80	36	36,31%
pred. 3	77	151	227	143	52	34,92%
pred. 4	56	103	145	200	126	31,75%
pred. 5	190	173	260	473	701	39,01%
recall	52,76%	23,79%	23,05%	20,33%	71,46%	

4.4 Vergleich der Klassifikatoren

Bei dem in dieser Studie vorliegenden quinären Klassifikationsproblem erzielt Random Forest mit 38,27% die höchste Genauigkeit, gefolgt von Naïve Bayes mit 35,49%. Die geringste Genauigkeit liefert k-Nearest Neighbor mit 32,56%. Abbildung 8.1 stellt die Vorhersagegenauigkeit der verwendeten Klassifikatoren auf Ebene der einzelnen Klassen differenziert dar. Die höchste Vorhersagegenauigkeit in Klasse 1 erzielt k-NN mit 49,89%. Dieser Klassifikator erkennt Bewertungen dieser Klasse um 6,41% häufiger, als Random Forest. In Klasse 2 erzielt NB mit 49,54% im Hinblick auf die Vorhersagegenauigkeit die besten Ergebnisse, d.h. nahezu jede zweite Bewertung wird korrekt vorhergesagt. Demgegenüber ordnet k-NN nur rund jede vierte Bewertung dieser Klasse richtig zu. In Klasse 3 erzielt NB mit 45,10% die höchste Vorhersagegenauigkeit, wohingegen k-NN Bewertungen dieser Klasse am schlechtesten voraussagt; verglichen mit NB ist die Vorhersagegenauigkeit von k-NN um 10,18% schlechter. In Klasse 4 erzielt k-NN jedoch die besten Ergebnisse. Bewertungen dieser Klasse werden von k-NN mit einer Genauigkeit von 35,63% vorhergesagt. Die Vorhersagegenauigkeit von NB und Random Forest unterscheidet sich nicht signifikant von k-NN,

weshalb anzunehmen ist, dass alle drei Klassifikatoren Bewertungen der Klasse 4 nicht eindeutig identifizieren können. In Klasse 5 erzielt k-NN mit 40,95% die höchste Vorhersagegenauigkeit. Random Forest kann Bewertungen dieser Klasse ähnlich gut vorhersagen; die Differenz zu k-NN beträgt 1,94%. Beim Vergleich der Klassifikatoren fällt auf, dass k-NN zwar insgesamt die geringste Genauigkeit erzielt, auf Ebene der Klassen jedoch bei drei von fünf Klassen die größte Vorhersagegenauigkeit liefert. NB erzielt bei zwei von fünf Klassen die beste Vorhersagegenauigkeit. Der Klassifikator Random Forest liefert zwar insgesamt eine bessere Genauigkeit, erzielt jedoch in keiner Klasse die beste Vorhersagegenauigkeit.

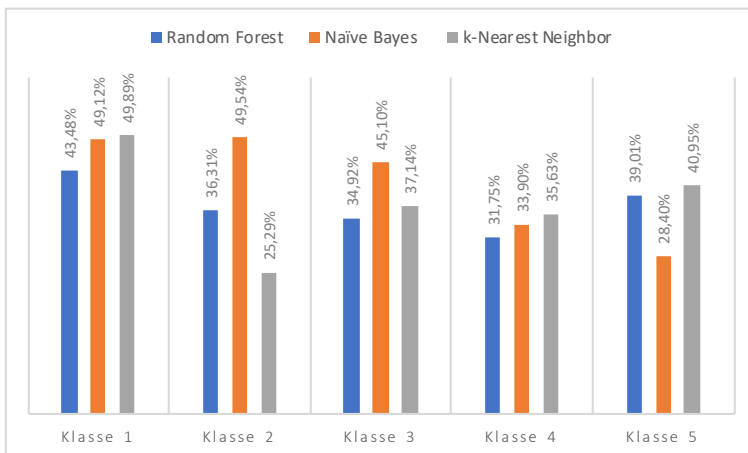


Abbildung 8.1: Vergleich der Precision der einzelnen Klassen von Random Forest, Naïve Bayes und k-Nearest Neighbor.

5 Fazit und Ausblick

Ziel dieser Studie war es, erste Aussagen zu treffen, inwieweit die durch Kunden über Ratingsysteme vergebenen Punkte bzw. Sterne die in den schriftlichen Reviews geäußerten Meinungen abbilden. Die Stimmungserkennung von Produktbewertungen wurde als quinäres Klassifikationsproblem betrachtet. Um erste Erkenntnisse zu gewinnen, wur-

de eine lexikonbasierte Sentimentanalyse mit den maschinellen Lernalgorithmen k-Nearest Neighbor, Naïve Bayes und Random Forest kombiniert. Die Ergebnisse der Klassifikatoren wurden mit zehnfacher Kreuzvalidierung evaluiert und anschließend verglichen. Random Forest erzielte mit 38,27% die höchste Genauigkeit, gefolgt von Naïve Bayes mit 35,49%. K-Nearest Neighbor lieferte mit 32,56% zwar insgesamt die niedrigste Genauigkeit, erzielte hinsichtlich der Vorhersagegenauigkeit bei drei von fünf Klassen jedoch die besten Ergebnisse. Naïve Bayes erzielte bei zwei von fünf Klassen die höchste Genauigkeit.

Die geringe Vorhersagegenauigkeit der eingesetzten Klassifikatoren lässt darauf schließen, dass sich die einzelnen Klassen nicht trennscharf voneinander abgrenzen ließen. Ursächlich hierfür könnte die geringe Menge der verwendeten Trainingsdaten sein. Der Datensatz wurde aufgrund technischer Beschränkungen auf 5.000 Elemente begrenzt. Jede Klasse enthielt 1.000 Elemente und damit für das Training eine verhältnismäßig geringe Anzahl. Überdies erfolgte die lexikonbasierte Vorverarbeitung mit einem domänenübergreifenden Sentimentwörterbuch, was dazu geführt haben könnte, dass das Sentiment domänenabhängiger Begriffe nicht korrekt erfasst und klassifiziert wurde. In der Fortführung dieser Studie soll der Fokus daher auf der Abgrenzung der einzelnen Klassen liegen. Ein erster Schritt kann dabei in der Anpassung des Sentimentwörterbuchs liegen, welches an die spezifische Domäne anzupassen ist. Überdies könnten Sentimentanalysen auf Satz- und Aspektebene genauere Ergebnisse liefern.

Danksagung

Diese Forschung wurde aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01PL17062 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor_innen.

Literatur

1. K. Plottek and C. Herold, "Micro Moments als entscheidender Moment im Rahmen einer zunehmend fragmentierteren Customer Journey," in *Customer Experience im Zeitalter des Kunden*, A. Rusnjak and D. R. A. Schallmo, Eds. Wiesbaden: Springer, 2018, pp. 143–176.
2. Bitkom, "Shopping digital – Wie die Digitalisierung den Handel tiefgreifend verändert," Tech. Rep., 2017. [Online]. Available: <https://www.bitkom.org/sites/default/files/file/import/171124-Studienbericht-Handel-Web.pdf>
3. "Amazon." [Online]. Available: <https://www.amazon.com>
4. T. Scholz, "Opinion Mining für verschiedene Webinhalte," in *Methoden der Webwissenschaft*, ser. Schriftenreihe Webwissenschaft, K. Scherfer and H. Volpers, Eds. Berlin: Lit, 2013, pp. 63–81.
5. W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014.
6. C.-N. Ziegler, "Automated Capture of Strategic Knowledge on the Web." Freiburg i. Br.: Albert-Ludwigs-Universität Freiburg i. Br., 2010, p. 160. [Online]. Available: <http://www2.informatik.uni-freiburg.de/~cziegler/papers/Habil-Thesis.pdf>
7. A. Aryo Prakoso, B. Winantesa Yananta, A. Fitra Setyawan, and Muljono, "A Lexicon-Based Sentiment Analysis for Amazon Web Review," in *2018 International Seminar on Application for Technology of Information and Communication*. IEEE, 2018, pp. 503–508.
8. A. M. Alkalbani, L. Gadhvi, B. Patel, F. K. Hussain, A. M. Ghamry, and O. K. Hussain, "Analysing Cloud Services Reviews Using Opining Mining," in *IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*. IEEE, 2017, pp. 1124–1129.
9. X. Fang and J. Zhan, "Sentiment analysis using product review data," *Journal of Big Data*, vol. 2, no. 1, pp. 1–14, 2015.
10. B. Lin, F. Zampetti, G. Bavota, M. Di Penta, M. Lanza, and R. Oliveto, "Sentiment analysis for software engineering: how far can we go?" in *the 40th International Conference*. ACM Press, 2018, pp. 94–104.
11. I. El Alaoui, Y. Gahi, R. Messoussi, Y. Chaabi, A. Todoskoff, and A. Kobi, "A novel adaptable approach for sentiment analysis on big social data," *Journal of Big Data*, vol. 5, no. 1, pp. 1–18, 2018.
12. "Rapidminer." [Online]. Available: <https://rapidminer.com>

13. "SentiWordNet." [Online]. Available: <http://sentiwordnet.isti.cnr.it/>
14. "Kaggle." [Online]. Available: <https://bit.ly/2PekQtP>
15. B. Pang and L. Lee, "Briefly Noted," *Computational Linguistics*, vol. 35, no. 2, pp. 311–312, 2009.
16. S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining." in *International Conference on Language Resources and Evaluation*, 2010, pp. 2200–2204.
17. R. Baeza-Yates and B. Ribeiro-Neto, "Modern Information Retrieval." Addison Wesley, 1999, pp. 29–30.
18. D. Li, B. Zhang, and C. Li, "A Feature-Scaling-Based k-Nearest Neighbor Algorithm for Indoor Positioning Systems," *IEEE Internet of Things Journal*, vol. 3, no. 4, pp. 590–597, 2016.
19. C. D. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval." New York: Cambridge University Press, 2008, p. 506.
20. L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

Entwurf einer interaktiven, dynamischen Logging-Schnittstelle in der Videospiel-Engine Unity

Design of an Interactive, Dynamic Logging Interface Using the Videogame Engine Unity

Manuel Heinzig, Marc Ritter

Hochschule Mittweida, Fakultät Angewandte Computer- und
Biowissenschaften, D-09648 Mittweida, heinzig@hs-mittweida.de

Zusammenfassung Dieses Paper legt die softwaretechnischen Grundlagen für eine Untersuchung der Tauglichkeit von Logging-Frameworks zum Zwecke des performanten Datenaustauschs zwischen der Videospiel-Engine Unity und einem komplexen Algorithmus. Hierfür wird mit Methoden der Objektorientierten Programmierung eine dynamische Schnittstelle zur einfachen Einbindung und Benutzung der Frameworks an Unity geschaffen. Anhand der beispielhaften Integration neuer Frameworks (nlog, log4net, serilog) erfolgt der Nachweis der korrekten Funktion. Das entstandene System stellt den Softwarekern für einen empirischen Test der Logging-Performanz auf Echtzeitfähigkeit und somit einem späteren Einsatz in einem entsprechenden Spiel zur Verfügung.

Schlagwörter Unity, performance, logging, framework

Abstract In this paper, we design a system, that can later be used to evaluate the capability of Logging Frameworks to serve as a performant data transmitter between the Engine Unity and a complex algorithm. With the use of object oriented programming, we develop a dynamic interface for easy integration and usage of new frameworks. The functional test is performed by the exemplary incorporation of commonly used frameworks. The resulting system serves as software core for an iterative study

on real-time performance, which successful evaluation proves its suitability for later usage in a scientific video game.

Keywords performance, logging, framework, data exchange,

1 Einleitung

Das Wort *Algorithmus* wird von Personen außerhalb der Informatik häufig als ein schwer zu definierendes, abstraktes Konstrukt aufgefasst. Im Kontext von Software und Code, so erhöht sich die induzierte Verständnisschwelle für Fachfremde noch weiter. Für in den Computerwissenschaften beheimatete Menschen stellt die Arbeit mit Algorithmen zwar einen täglichen Vorgang dar, jedoch hängt auch hier der Arbeitseifer stark von den Ansprüchen des umzusetzenden Vorgangs und den eigenen Handlungspräferenzen ab. Durch die bei der Implementierung häufig notwendige Vermischung von eigens geschriebenem Code mit dem eines anderen Entwicklers verstärkt sich die Verständnishürde auch für versiertes Personal noch weiter [1].

Für den meist auf iterativen Tests und deren statistischer Auswertung basierenden Vorgang der Algorithmenoptimierung bedeutet dies, dass hochspezialisiertes Personal eine verhältnismäßig einfachere Fleißaufgabe übernehmen muss. Erste Ansätze zur Verbesserung sehen daher Experimentalaufbauten vor, in denen die Variation von Eingabewerten und -daten eines Algorithmus nicht mehr auf Codezeilenebene geschehen muss, sondern durch eine deutlich einfacher zu bedienende grafische Oberfläche ausgelöst wird. [2, S.112ff]

Klassische Oberflächen benötigen meist dennoch eine recht große Einarbeitungszeit. Oft mangelt es der optischen Darstellung zudem an visueller Elaboriertheit und optimierter Nutzerführung, was beim Bedienenden ein gewisses Sachverständnis zum Tätigen der korrekten Eingaben notwendig macht. Als Weiterentwicklung dieser rudimentären Oberflächen sollen dem Anwender zukünftig kleine Videospiele zur Verfügung stehen, die die Optimierung hinter Gamification-Aspekten verschleiern. Auf diese Art und Weise können selbst fachfremde Testpersonen an der Evaluation mitwirken [3]. Auf den folgenden Seiten werden daher die softwaretechnischen Grundlagen für

eine einfache performante Nutzung variabler Frameworks aus der Videospiel-Engine Unity heraus erläutert.

2 Grundlagen

Das *Loggen*, also das Herausführen von Informationen aus einem System zu seiner Laufzeit, besitzt in vielen Anwendungsdomänen einen hohen Stellenwert. So ist das Führen eines Logbuchs in der Seefahrt schon seit vielen hundert Jahren Tradition, um zum einen die vom Festland nicht beobachtbaren Handlungen der Matrosen zu dokumentieren, zum anderen ermöglichen Niederschriften über Kursverlauf u.Ä. im Nachhinein wertvolle Analysen über den Reiseverlauf. [4]

Analog dazu ist es auch in Software hilfreich, detaillierte Informationen über die Geschehnisse zur Laufzeit im Nachgang einer Untersuchung zu unterziehen. Entsprechende Meldungen sind in strukturierter Art aus einem laufenden Programm herauszuführen, ohne die Performanz der Hauptsoftware dabei spürbar zu beeinflussen [5].

Durch die vielfältigen Verwendungsmöglichkeiten von *Logging* in Softwareprojekten [6] existiert eine große Fülle einsatzbereiter Implementationen. Beispielhaft seien *NLog*¹, *log4net*², *Logazmic*³, *Serilog*⁴, *Foundatio*⁵, *Loupe*⁶ und *BugSnag*⁷ genannt. Die aufgeführten Frameworks unterscheiden sich in der Art ihrer Einbindung in den Code eines Programms, weisen aber im Kern typische konzeptuelle Gemeinsamkeiten auf [7]. Es lassen sich daher praktisch allen Frameworks die im Folgenden kurz erläuterten Grundfunktionalitäten zuschreiben.

2.1 Log-Level

Für die häufig in großer Anzahl auftretenden Log-Events hat sich eine Kategorisierung anhand ihrer Wichtigkeit für die erfolgreiche Weiterführung des Programmablaufs oder die anschließende Analyse ent-

¹ <https://github.com/nlog/NLog/>

² <https://logging.apache.org/log4net/index.html>

³ <https://github.com/ihtfw/Logazmic>

⁴ <https://github.com/serilog/serilog>

⁵ <https://github.com/FoundatioFx/Foundatio>

⁶ <https://onloupe.com/>

⁷ <https://docs.bugsnag.com/platforms/dotnet/>; alle Links geöffnet am 19.02.2019

wickelt. Obgleich Logging-Frameworks auch eigene Bezeichnungen und feinere Stufeneinteilungen implementieren, haben sich die in Tabelle 9.1 aufgeführten Level als Quasistandard etabliert.

Trace	Feingranular detaillierte Informationen des Programmablaufs
Debug	Funktionale Fehlerbehebung (ergebnisbeeinflussende Ereignisse)
Info	Allgemeine Informationen des Programmablaufs
Warn	Abweichendes, potentiell fehlerverursachendes Programmverhalten
Error	Unerwünschtes Verhalten jeglicher Art (<i>Ausnahmefehler</i>)

Tabelle 9.1: De-facto-Standard-Stufen des Loggings und deren Einsatzzweck.

2.2 Ziele/Targets

Die übliche Ausgabe von Zeichen auf die Kommandozeile eignet sich aufgrund der temporär begrenzten Verfügbarkeit nur bedingt für nachträgliche Analysen. Im Sinne einer persistenteren Sammlung von Informationen bieten Frameworks in der Regel eine Funktionalität zum Senden der Log-Nachrichten an diverse alternative *Speicherziele* an⁸. Dies können bspw. Datenbanken verschiedenster Art sein, aber auch elaboriertere Methoden wie das direkte Versenden der Nachrichten über Netzwerkprotokolle gehören oft zum Funktionsumfang.

2.3 Filter

Als Resultat von Logging-Aktivitäten entstehen sehr viele, klar schematisierte, mit Laufzeitinformationen angereicherte Zeichenketten. Für eine effiziente Weiterverwendung können diese auf das Auftreten bestimmter Muster untersucht werden. Das ermöglicht die Umleitung von Nachrichten anhand ihrer Kriterien oder die Anpassung von Speicherparametern eines Ziels.

3 Softwarestruktureller Rahmen

Die verwendete Videospiel-Engine Unity besitzt eine Menge von Softwaremodulen, die dem Entwickler grundlegende Arbeitsschritte er-

⁸ <https://logging.apache.org/log4net/log4net-1.2.13/release/sdk/log4net.Appender.html>

leichtern sollen. Dazu zählen bspw. die Ansteuerung von Ein- und Ausgabehardware sowie das Arrangieren und Interagieren von Objekten in einer Szene. Softwareseitig hält dieses Konstrukt ein zentraler Verarbeitungskern zusammen, in welchem einzelne Schritte nach einem vordefinierten Muster ablaufen. Als Folge dieser Charakteristik ist es nur schwer möglich, mit den vorhandenen Mitteln performant zu loggen. Die auf das effiziente Herausführen und Abspeichern von Informationen ausgelegten Logging-Frameworks erscheinen als geeignete Werkzeuge zur Einschränkung und potentiellen Verbesserung dieses Missstandes [8].

Zum Test dieser Überlegung sollten möglichst viele Frameworks an die Grundroutinen der Engine angekoppelt und deren Performanzauswirkung beobachtet werden. Hierbei ist auf die Minimierung des Erweiterungsaufwands für den umsetzenden Entwickler zu achten. Mit diesem Ziel entstand eine dynamische Softwarestruktur, welche die einzelnen Frameworks über eine einheitliche Schnittstelle mit den Abläufen der Engine zusammenführt.

4 Softwarearchitektur

Logging-Frameworks besitzen jeweils eine eigene Syntax für ihre Verankerung im Host-Programm, haben dabei aber einen leicht divergierenden Funktionsumfang. Da diese Heterogenität nicht automatisch erfassbar ist, müssen derartige Metadaten daher initial vom Entwickler in einer Datei (XML-Format, durch eine XSD definiert) konfiguriert werden, was im späteren Verlauf z.B. das korrekte Anzeigen von Optionen in der GUI ermöglicht.

Die exakte Code-Syntax und -Semantik divergiert für jedes Framework, weshalb der Entwickler durch Ableitung vom Interface [9, S.7-15] eine Klasse implementieren muss, die die grundlegenden Funktionen eines Logging-Frameworks abbildet. So entsteht für jedes Framework eine, von einem übergeordneten Logging-Manager verwaltete, Kombination aus Anbindungsklasse und Metadatei. Der Manager erstellt daraus neue Instanzen von Loggern, sendet Nachrichten an die richtige Instanz und interagiert mit der Nutzeroberfläche.

Die zu loggenden Signale und Nachrichten entspringen schließlich einer Reihe von dynamischen Senderobjekten. Ihre Konfiguration ge-

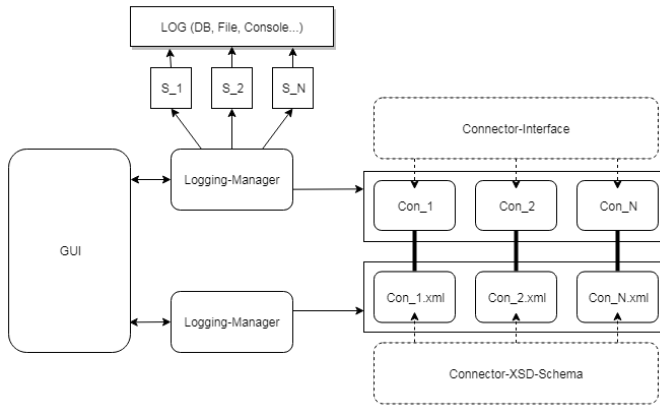


Abbildung 9.1: Schematische Darstellung der entstandenen Komponenten des Softwarekerns zur dynamischen Anbindung variierender Logging-Frameworks an eine steuernde GUI oder Spiellogik.

schieht über die GUI. Sie entsenden Nachrichten mit einer variierbaren Senderate an verfügbare Ziele und besitzen einige statische und variable Eigenschaften, welche die Log-Nachrichten ggf. als Nutzdaten anreichern können. Die entstandene Grundstruktur der beteiligten Softwaremodule ist in Abbildung 9.1 grafisch dargestellt.

Ausgehend von einer Anzahl, durch eine XML Schema Definition (XSL) beschriebener, konsistent strukturierter XML-Dateien (unten rechts, eine pro inkludiertem Logging-Framework) erstellt der Session Manager das Datenmodell und die Konfiguration für die darzustellende Szene. Die bereitgestellten Informationen befüllen diverse Elemente der GUI und bestimmen so deren Funktionsumfang. Beispielfür diesen Vorgang sind die Optionen in der Dropdown-Auswahl *Framework*, deren Einträge dynamisch aus den vorhandenen XML-Konfigurationsdateien generiert werden. Im Anschluss interagiert ein Tester mit der Oberfläche und erstellt durch entsprechende Tätigkeiten neue Senderobjekte(S_x), welche der Logging-Manager temporär speichert und verwaltet. Aufbau und genaue Funktionsweise der Sender werden durch die in getrennten Codemodulen implementierten Connectors (einer für jedes Framework) und der in der GUI eingestell-

ten Eigenschaften vorgegeben. Das Connector-Interface dient dabei in seiner Funktion als programmiertechnische Blaupause zur Wahrung einer einheitlichen, vom Manager konsistent ansprechbaren Schnittstelle für die diversen angebotenen Logging-Frameworks.

5 Visualisierung

Ein großer Vorteil beim Aufbau eines Testprogramms in einer Videospiel-Engine besteht in der einfachen Visualisierung und Nutzerinteraktion. Weniger optisch ansprechend und von geringerer Komplexität als im Anwendungsfall eines vollständigen Spiels bleibt die Darstellung aber auf einem funktionalen Niveau.

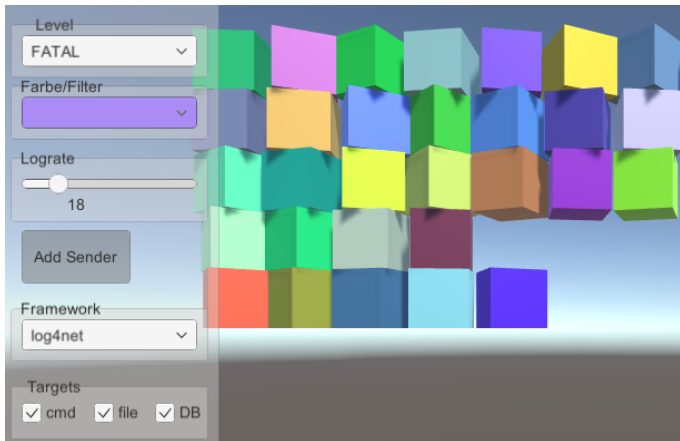


Abbildung 9.2: Aufbau der GUI. Optionsmenü (links) und Sender (rechts). Ein Würfel repräsentiert dabei ein Senderobjekt, welches in gewissen Zeitabschnitten (Pulsrate dargestellt durch Drehgeschwindigkeit) mittels des ihm zugewiesenen Frameworks an die zum Erstellungszeitpunkt aktivierten Targets loggt.

Die in Abbildung 9.2 gezeigte Oberfläche besteht vorrangig aus Interaktionselementen. Sie dienen der Einstellung von Eigenschaften der Senderobjekte. Nach Angabe des zu verwendenden Frameworks erfolgt die Auswahl des Log-Levels und die Eingabe der Senderate für Nachrichten. Nach optionaler Nutzung von Filteroptionen (exempla-

risch: Filtern der Blöcke nach Farben) und der Selektion mindestens eines Logging-Ziels wird ein neuer Sender über den Button hinzugefügt.

Der Sender ist ab diesem Zeitpunkt als bewegter Würfel visualisiert, dessen Rotationsgeschwindigkeit die Senderate repräsentiert. Die numerischen Werte seiner Farbe und die aktuelle (Dreh)position können als Nutzdaten mitgeloggt werden oder durch die optionale Filterung nach Farben oder Farbbereichen den Charakter der Log-Nachrichten beeinflussen. Die Auslenkung auf der vertikalen Achse bestimmt das angegebene Log-Level. Sieben langsam rotierende Blöcke in der obersten Zeile entsprechen folglich dem visualisierten Äquivalent von sieben geöffneten Sendekanälen, über die in langsamen Zeitintervallen Log-Nachrichten mit der Wichtigkeit *INFO* gesendet werden. Ein Filter kann hierbei bestimmen, nur das Senden von Blöcken eines bestimmten Grüntons zuzulassen. Alle in der GUI vorhandenen Optionen werden auf Grundlage der in der Metakonfigurationsdatei getätigten Einträge dynamisch befüllt.

6 Test des Prototyps und Leistungsbeurteilung

Zur Überprüfung des dynamischen Charakters der entstandenen Software werden einige Logging-Frameworks auf vorgesehene Art in das System integriert. Dies beinhaltet das Erstellen einer XML-Konfigurationsdatei sowie die korrekte Implementierung einer Connector-Klasse für das korrekte Ansprechen des gewünschten Frameworks auf Codeebene. Bei beiden Teilschritten kann sich der Entwickler an den vom Framework gegebenen Vorgaben orientieren, um eine gültige Programmstruktur zu erhalten. Anschließend erfolgt eine kurze Exploration der ableitbaren Leistungsindikatoren.

Aus der großen Anzahl verfügbarer C#-Logging-Frameworks werden stellvertretend *log4net*⁹, *nlog*¹⁰ und *serilog*¹¹ für den Test ausgewählt. Diese ermöglichen durch ihre Implementationsanforderungen repräsentative Aussagen für viele andere Frameworks. Mit den gewählten werden komplex zu bedienende, praktisch allumfassende Softwarekonstrukte großer Entwicklerteams mit einer enormen Anzahl

⁹ <https://logging.apache.org/log4net/index.html>

¹⁰ <https://nlog-project.org/>

¹¹ <https://serilog.net/>

von Ausstattungsmerkmalen ebenso repräsentiert wie kleine, aus einer einzigen Bibliotheksdatei bestehende und auf die notwendigsten Minimalfunktionen zugeschnittene Produkte. Wie vorgesehen, erhalten sie bei der Implementation jeweils einen Ordner für ihre Bibliotheksdateien, die XML-Datei mit Metadaten und eine auf der Interfaceklasse beruhende Schnittstellenimplementierung. Nach diesem einmaligen Aufwand sind alle notwendigen Optionen über die GUI auswählbar und die Frameworks ohne weitere Komplikationen durch Senderobjekte anzusteuern.

Eine Aussage über die Performanz des Softwareverbundes lässt sich anhand der auftretenden Darstellungsverzögerung des in Echtzeit ablaufenden Programms treffen. Die hierfür übliche Kenngröße ist die Angabe der FPS (Frames Per Second), also der Anzahl von Neuberechnungen der Visualisierung pro Sekunde [10]. Da die erstellte Szenerie durch die Verwendung einfacher Primitive und Objekte kaum Systemlast erzeugt, kann bei Schwankungen der Geschwindigkeit relativ sicher auf Verzögerungen durch die im Hintergrund ablaufenden Logging-Skripte geschlossen werden. Das Auslesen der FPS erfolgt dabei entweder direkt über die in Unity vorhanden Leistungsmessung, oder mittels einer wiederkehrenden Berechnung mittels der entsprechenden trivialen Formel an einer unkritischen Stelle im Programmcode.

7 Zusammenfassung

Die vorgestellte Variante der dynamischen Anbindung externer Logging-Werkzeuge in der Spiele-Engine Unity ermöglicht dem Anwender die einfache Nutzung und Parametrisierung einer abstrakten Softwarearchitektur. Auf diese Weise können auch fachfremde Personen mit wenig Aufwand an der Durchführung entsprechender Experimente mitwirken. Für Entwickler ist es möglich, weitere Frameworks mit überschaubarem Aufwand in das bestehende System zu integrieren. Das Ziel der dynamischen Anbindung und eingängigen Verwendung kann daher als grundsätzlich gelungen betrachtet werden und legt den Grundstein für weitere Tests in realitätsnäheren Anwendungen.

Danksagung

Diese Arbeit wurde aus Mitteln des Europäischen Sozialfonds und des Freistaates Sachsen gefördert.



Diese Maßnahme wird mitfinanziert durch Steuermittel auf Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Literatur

1. F. Zhang, F. Khomh, Y. Zou, and A. E. Hassan, "An empirical study of the effect of file editing patterns on software quality," *Journal of Software: Evolution and Process*, vol. 26, no. 11, pp. 996–1029, 2014.
2. A. Oram and G. Wilson, *Making Software: What Really Works, and Why We Believe It*. O'Reilly Media, 2010.
3. R. Bierig, J. Gwizdka, and M. J. Cole, "A user-centered experiment and logging framework for interactive information retrieval," in *UIIR@SIGIR*, 2009.
4. C. Ward and D. Wheeler, "Hudson's bay company ship's logbooks: a source of far north atlantic weather data," *Polar Record*, vol. 48, no. 2, pp. 165–176, 2012.
5. D. Yuan, S. Park, and Y. Zhou, "Characterizing logging practices in open-source software," in *Proceedings of the 34th International Conference on Software Engineering*, ser. ICSE '12. Piscataway, NJ, USA: IEEE Press, 2012, pp. 102–112.
6. W. Shang, M. Nagappan, and A. E. Hassan, "Studying the relationship between logging characteristics and the code quality of platform software," *Empirical Software Engineering*, vol. 20, no. 1, pp. 1–27, Feb 2015.
7. B. Camp, K. Ibitayo, P. Issa, and C. Xu, "Computer software framework and method for logging messages," Patent US6 802 067B1.
8. J. Gregory, *Game Engine Architecture, Second Edition*, 2nd ed. Natick, MA, USA: A. K. Peters, Ltd., 2014.
9. D. Poo, D. Kiong, and S. Ashok, *Object-Oriented Programming and Java*. Springer-Verlag London, 2008.
10. K. T. Claypool and M. Claypool, "On frame rate and player performance in first person shooter games," *Multimedia Systems*, vol. 13, no. 1, pp. 3–17, Sep 2007.

Optimization of a Hybrid Learning Approach for Power Electronics Course Using Virtual Laboratory

Muhamad Ali¹, Kurniawan D. Irianto², Didik Hariyanto¹, Thomas Koehler³, Djemari Mardapi¹, and Soenarto¹

¹ Universitas Negeri Yogyakarta, Technological and Vocational Education, Jalan Colombo No. 1, 55281 Yogyakarta, Indonesia

² Technische Universität Dresden, Chair of Communication Networks, 01187 Dresden, Germany

³ Technische Universität Dresden, Chair of Educational Technology, 01062 Dresden, Germany
muhal@uny.ac.id

Abstract Many challenges need to be addressed in facing the Industry 4.0 (I4.0), especially for teachers in schools or colleges. One of them is how to adapt a learning process with I4.0 where computers and technologies play an important role. A hybrid learning approach using a virtual laboratory could be the answer to this problem. Hybrid learning combines face-to-face learning, e-learning, virtual laboratory and practice in a laboratory. It could optimize the advantages of all methods and introduce educational technology to students. Furthermore, it will prepare the vocational teacher candidates in facing the I4.0. This paper focuses on developing, implementing and optimizing a power electronic course for vocational students with a hybrid learning approach using a virtual laboratory in Universitas Negeri Yogyakarta, Indonesia and it is based on the author's experiences. The hybrid learning approach is validated by experts from electrical engineering education using an assessment method and analyzed by means statistically. The result shows that the hybrid learning approach by using virtual laboratory can be adopted for improving the analytical thinking in solving power electronics problems for students as a vocational teacher candidate.

Keywords hybrid learning approach, power electronics, e-learning, vocational teacher, industry 4.0

1 Introduction

Modern industry in the era of industrial revolution 4.0 (I4.0) has led to an increase in energy requirements, especially from renewable energy. Most of the energy that is used by a human being is electricity. To improve the quality and efficiency of electrical energy, many things can be done by converting energy. However, traditional converter systems are insufficient to increase the controllability, capacity, and the efficiency of power as well as efficiency for energy systems [1]. For these reasons, current studies are now more focused on electronics switching. Power electronics consist of electronic switching principles to convert from Alternating Current (AC) to Direct Current (DC), DC to DC, AC to AC and DC to AC [2]. It is used in a wide range of applications such as motor controls, industrial applications, vehicle systems, electromechanical controls, and power systems integration of renewable energy resources [1, 2]. It becomes a subject course in the electrical engineering education curriculum for preparing teachers in electrical engineering fields.

The most common teaching approach used in higher education is a face-to-face lecture, and it is known as the traditional teaching. However, in recent years, the lecture is described as passive teaching, because it discourages students to think critically and construct their learning. It focuses only on face-to-face interaction. It does not accommodate student opportunities for collaborative learning. Moreover, it does not give a chance to the instructors to implement learning with analytical and critical thinking skills [3, 4]. As a result, students may feel bored in a classroom with traditional teaching [5]. They usually do not get motivated by the face-to-face classroom. Besides, they might think it will not bring any benefit for their future careers. Therefore, they pay less attention to the lectures. They do not care and do not participate in the class. As a result, it can cause behavioral problems for them.

Nowadays, the learning and teaching process on power electronics course have changed greatly. Science and technology have become digital and an integral part of the educational system in many countries. Many institutional educations provide e-learning to support the learning process such as downloading lecture notes, learning web-based materials, attending web-based lectures and video lectures [6]. Many lecturers have investigated the scope of information and communication

technologies in education [3, 4, 7]. On the other side, some lecturers have developed the virtual laboratory to give the practical skill for students and implemented in many subject areas such as Chemistry [5], Electronic Engineering [8], Robotics [9], Science [10] and Power Electronic [11, 12]. The implementation of online learning and virtual laboratories has many advantages. They are easy to use, reduce time, and allow students to perform several experiments within a limited time and plan their future studies carefully [3, 5, 7, 10].

However, there are some problems with the way of teaching with e-learning or virtual laboratories, especially for the engineering subject like a power electronics course. E-learning or virtual laboratories cannot satisfy the needs of the participants in the educational process. Skill development must be obtained by experiment rather than only computer simulation through e-learning or by the virtual laboratory. As a teacher candidate in vocational school, students must have analytical and critical thinking skills to collect and analyze information, solve problems and make the best decision [13]. It can be improved by some learning methods such as causal thinking, creative thinking, knowledge-seeking, systematic problem solving, and decision-making. The analytical and critical thinking skills can be measured from several aspects like 1) inquisitiveness, 2) self-confidence, 3) analyticity and 5) maturity [13, 14].

2 The Objective of The Study

Considering the previously defined statement, the main objective of this study is:

- a. Developing a hybrid learning approach using a virtual laboratory to increase analytical and critical thinking skills on solving the problems in the power electronics course.
- b. Implementing the power electronics course using a hybrid learning approach to increase analytical and critical thinking skills on solving the problems in the power electronics course.
- c. Optimizing the power electronics course using a hybrid learning approach with the combination of face-to-face, e-learning, virtual

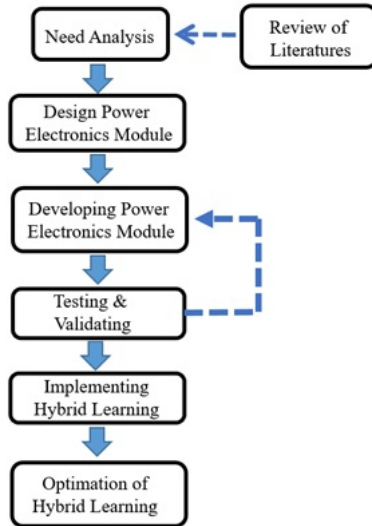


Figure 10.1: The procedure of research

laboratory and practice in the real laboratory to increase the analytical and critical thinking skills on solving the problems in the power electronics course.

3 Research Method

The research of the optimization of a hybrid learning approach using virtual laboratories in power electronics courses is carried out through a procedure as shown in Figure 10.1. The first step involves the research and information collecting for the need analysis. It is conducted by literature reviews, classroom observations, and report preparation for the state of the art. The second step is to design a Power Electronics module. The next step is developing the Power Electronics module that consists of a manual, e-learning, a virtual laboratory, and a laboratory module.

The fourth step is the validation of the power electronics module by the electrical engineering education expert. The fifth step is implementing a hybrid learning approach in the power electronics course with

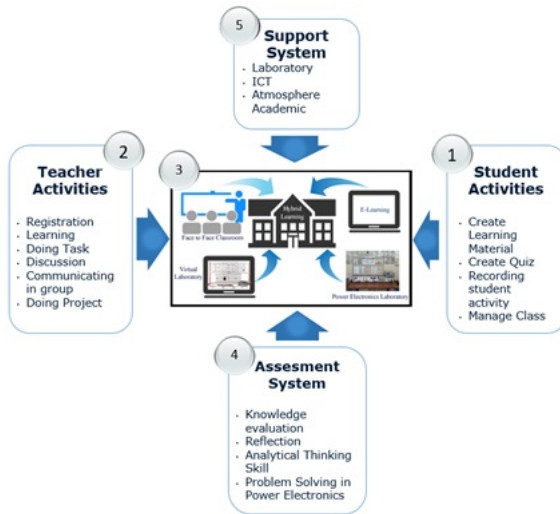


Figure 10.2: The hybrid learning approach using virtual laboratory

integration of a virtual laboratory. The implementation of the hybrid learning approach has carried out through classroom action research. The instrument that used to measure the module of hybrid learning on power electronics course was a questionnaire using a 4-point Likert scale ranging from 1 (very weak) to 4 (very good/strength). The competence of students will be measured by a test and the analytical and critical thinking skills were measured by observation of the students about inquisitiveness, self-confidence, analyticity, and maturity.

4 Result and Discussion

4.1 Developing Hybrid Learning Approach

The hybrid learning approach using the virtual laboratory on power electronics course is illustrated in Figure 10.2. The components of hybrid learning using a virtual laboratory on power electronics course are composed of five elements. There are 1) student activities, 2) teacher activities, 3) hybrid learning, 4) assessment system, and 5) support sys-

tem. The development of a hybrid learning module was based on the need of analysis.

The hybrid learning module was validated by three experts from electrical engineering education who have experience in e-learning, computer media development, and virtual laboratory. Based on the judgment from the experts, the hybrid learning approach applied in the power electronics course with virtual laboratory is highly recommended with an average score of 3.22.

Average scores of 3.22 can be interpreted as good level; it means all experts give a good or a very good interpretation in each aspect of the hybrid learning module on the power electronics course. They agree and give a good rate to the hybrid learning approach module with confidence. Meaning, it can be implemented to improve analytical thinking skills and evaluation ability for problem solving in the power electronics fields.

Table 10.1: The expert judgment to the hybrid learning approach on power electronics course

No	Aspects of Hybrid Learning Approach	Mean	Interpretation
1	Usefulness of Power Electronics module in hybrid learning	3,10	Good
2	Face to face activities	3,02	Good
3	E-Learning activities	3,42	Very Good
4	Virtual laboratory	3,36	Very Good
5	Laboratory activities	3,04	Good
6	Assesment system	3,38	Very Good
	Summary	3,22	Good

Average scores of 3.22 can be interpreted as good level; it means all experts give a good or a very good interpretation in each aspect of the hybrid learning module on the power electronics course. They agree and give a good rate to the hybrid learning approach module with confidence. Meaning, it can be implemented to improve analytical thinking skills and evaluation ability for problem solving in the power electronics fields.

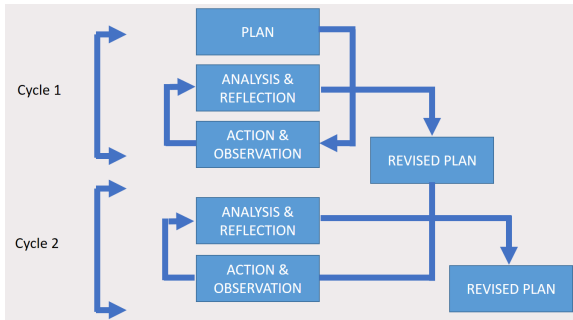


Figure 10.3: Class Action Research in Power Electronics

4.2 Implementation and Optimization

The implementation of hybrid learning, using a virtual laboratory on the power electronics course, is done in the odd semester from September until December 2018. It is implemented by class action research to obtain the best method for implementing a hybrid learning approach in the power electronics course. Active learning strategies are used to allow students to participate in the classroom, e-learning, virtual laboratory and practice in the laboratory. This method can enhance the skill such as analytical skill, communication skill, and social skills.

Table 10.2: The average score test and analytical thinking skill

No	Aspects	Cycle 1	Interpretation	Cycle 2	Interpretation
1	Test Score	6,22	Enough	7,96	Good
2	Inquisitiveness	2,42	Weak	2,92	Good
3	Self-Confidence	2,68	Good	3,28	Very Good
4	Analyticity	2,62	Good	3,32	Very Good
5	Systematicity	2,48	Weak	3,12	Good

Table 10.2 shows the result of test score and analytical thinking skills aspect through the implementation of the hybrid learning approach in cycle 1 and cycle 2. The learning process in cycle 1 was done by a hybrid learning approach without the virtual laboratory. Cycle 2 was done by

a hybrid learning approach using the virtual laboratory.

The combination of face-to-face, e-learning, virtual laboratory and practice in the laboratory was planned, done, analyzed and reflected in each cycle. Based on 2 cycles of hybrid learning using virtual laboratory, we identified the following strategies to achieve the best results:

- a) Theory in classroom: Power electronics learning starts from learning activity in the classroom with the face-to-face method organized by the teacher. Face-to-face learning in the classroom is meant to provide basic provisions for electronic concept theory. From this learning, students are expected to have knowledge as a solution to solve some problems in power electronics fields. The active learning method is implemented in the classroom with the face-to-face learning process.
- b) Simulation using virtual laboratory: The virtual laboratory is a laboratory clone obtained through the use of computers and software. The use of a virtual laboratory is aimed so that students could understand the procedures and do experiments in the real laboratory. Virtual learning is done by using power electronics computer-based learning media, Power Simulator (PSIM) and Personal Simulation Program with Integrated Circuit Emphasis (PSPICE) software to help students understand the procedures of designing and analyzing the power electronic circuits.
- c) Practice in real laboratory: Laboratory activities begin with the introduction of tools and materials, measurements, and testing devices, identifying tools and practice materials, assembling power electronic circuits (rectifiers, controlled rectifiers, AC regulators, DC chopper and inverters), measurement process and analyzing input and output of the circuit. Student activities in the laboratory are designed by the project-based learning which has a chance for the lecturer to give some assignments to the students, for example, designing power electronic circuits and solving real-world problems.
- d) Analysis: The next step is to analyze the results of laboratory activities in accordance with the theory that has been given. The analysis is carried out in the form of reports accompanied by data

on the results of lab works and projects. Every student must make a laboratory activity report for every given job.

- e) Presentation: The final step in hybrid learning is to present the results of the project that students are working on. Presentations are conducted in groups. It aims to train the students to work in a group.

5 Conclusions

The hybrid learning approach on the power electronics course integrated a variety of learning methods. Traditional face-to-face learning in the classroom gives some basics knowledge for students. E-learning adds the student's understanding of theory and self-paced learning. It contributes to the improvement of analytical thinking skill, and to the evaluation of ability to solve problems in power electronics fields. A presentation can give the students experience to prepare, design, practice and present the theoretical concepts in class.

The optimization of the hybrid learning approach for the power electronics course was done by integrating all of the resources related to conventional teaching, e-learning, virtual laboratory and practice in the laboratory. Students are more interested and they keep their attentions in class, e-learning, virtual laboratory and practice in the laboratory. Altogether, this research shows that hybrid learning can be adopted for improving analytical thinking skill in solving power electronics problems for students as a vocational teacher candidate.

References

1. M. H. Rashid, *Power Electronics: Devices, Circuits, and Applications*. New Jersey: Prentice-Hall, 2003.
2. M. Ali, *Aplikasi Elektronika Daya pada Sistem Tenaga Listrik*. Yogyakarta: Universitas Negeri Yogyakarta (UNY) Press, 2018.
3. A. V. Balastov and E. Y. Sokolova, "Adult Learners' Communicative Foreign Language Competence Development in Higher School Via Information Technology and Multimedia Implementation," *Mediterranean Journal of Social Sciences*, vol. 6, no. 1, pp. 537–543, 2015.

4. A. Buran, "How to Use Blogs in Creating Special Opportunities for Language Learning," *Mediterranean Journal of Social Sciences*, vol. 6, no. 1, pp. 532–536, 2015.
5. C. Tüysüz, "The Effect of the Virtual Laboratory on Students' Achievement and Attitude in Chemistry," *International Online Journal of Educational Sciences*, vol. 2, no. 1, pp. 37–53, 2010.
6. M. Ali, "Peningkatan Kualitas Pendidikan dan Pembelajaran Melalui Teknologi Informasi dan Komunikasi di Universitas Negeri Yogyakarta." Prosiding SENTIA Politeknik Negeri Malang, 2009, pp. C1–C5.
7. D. L. Matukhin and A. M. Evseeva, "Further Professional Training as a Constituent Part of Continuing Vocational Education," in *Lecture Notes in Management Science*, 2014, vol. 31, pp. 104–109.
8. M. Macias, V. Cazares, and E. Ramos, "A Virtual Laboratory For Introductory Electrical Engineering Courses To Increase The Student Performance." 31st ASEE/IEEE Frontiers in Education Conference, 2001.
9. F. A. Candelas, S. T. Puente, F. Torres, F. G. Ortiz, P. Gil, and J. Pomeraz, "A Virtual Laboratory for Teaching Robotics," *Int. J. Engng Ed.*, no. 3, pp. 363–370, 2003.
10. D. Liu, P. Valdiviezo-Diaz, G. Riofrio, Y.-M. Sun, and R. Barba, *Procedia Computer Science*, vol. 75.
11. R. Coteli and A. O. Gokcan, "Virtual Laboratory for Power Electronic Based Reactive Power Compensators," *Tehnički vjesnik*, vol. 25, no. 1, pp. 86–93, 2018.
12. K. Cheng, C. Chan, N. Cheung, and D. Sutanto, "Virtual Laboratory Development for Teaching Power Electronics," in *IEEE 33rd Annual Power Electronics Specialists Conference*, 2002, pp. 461–466.
13. A. Doyle, "Creative Thinking Definition, Skills, and Examples," <https://www.thebalancecareers.com/creative-thinking-definition-with-examples-2063744>, online (Accessed: 2019-04-05).
14. P. A. Facione and N. C. Facione, "The Holistic Critical Thinking Scoring Rubric: A Tool for Developing and Evaluating Critical Thinking," <http://www.affect-reason-utility.com/1301/Rubric.pdf>, 1994, online (Accessed: 2019-01-21).

Interaktives Storytelling - Mit Geschichten spielend online lernen Erklärvideos im wissensvermittelnden Kontext an Hochschulen

Rika Fleck

Hochschule Mittweida, D-09648 Mittweida, fleck@hs-mittweida.de

Zusammenfassung Die Studierenden, die jetzt mit dem Studium beginnen, leben und erleben eine völlig neue Art des Medien- und damit auch des Wissenskonsums. Das zeigt die Onlinestudie von ARD und ZDF: Die zwischen 1980 und 2000 geborenen „Millennials“ wachsen als Digital Natives auf. Und: Sie erwarten mobile Unterhaltung, On-Demand-Angebote, emotionale Ansprache und ständig neue innovative Formate. Die Informationsflut hat sie schon früh gelehrt, zu selektieren und effektiv das Wesentliche eines Sachverhaltes zu erfassen. Die 14- bis 29-jährigen sind täglich etwa 245 min lang online und mehr als ein Drittel von ihnen nutzt das Internet via Smartphone. [1] Auf dieses veränderte, durch Digitalisierung geprägte Konsumverhalten, werden sich auch die Hochschulen einstellen und auf die Zielgruppe zugeschnittene Lerninhalte anbieten müssen: Kurz, prägnant, mobil abrufbar und videobasiert ist die Devise. [2] Keine andere Art von Inhalten kommt dabei so gut an wie Videos. Zwar sind es die unterhaltsamen und lustigen Inhalte, die am meisten geschaut werden, noch weit vor Nachrichten und Social-Media-Posts. [3] Aber laut der JIM-Studie gehen Videos auch als die am meisten favorisierte digitale Anwendung für das Studium hervor. [4] Das vorliegende Paper setzt sich mit der Ausgangssituation der Wissensanbieter auseinander, fasst Theorien über das Lernverhalten zukünftiger Studierender zusammen und stellt ein Experiment zur Wissensvermittlung mittels interaktivem Storytelling vor, welches über zwei Semester mit Studierenden an der Hochschule Mittweida umgesetzt wurde. In diesem Experiment geht es darum, Wissen in einem interaktiven Erklärfilm zu vermitteln und interaktiv abzufragen. Dieses Projekt ist eine Kooperation mit dem Universitätsklinikum Carl Gustav Carus in Dresden und wird derzeit für eine spätere Auswertung getestet.

Schlagwörter digital learning, interactive storytelling, digital natives, e-learning, explainity film

1 Einleitung

Ausgangssituation an den Hochschulen

Vier von fünf Abiturienten entscheiden sich nach der Schule für ein Studium. [5] Das wirkt sich einerseits auf den Betreuungsschlüssel für Studierende aus. Eine individuelle Betreuung ist in vielen Modulen nur noch schwer möglich, auch in den Hochschulen für angewandte Wissenschaften. [5] Deshalb wird die Suche nach Möglichkeiten, die Lehre zu digitalisieren immer bedeutender. Aber auch die Palette an Lebensmodellen während des Studiums hat sich verändert und ist vielfältig geworden. Der „Normalstudierende“, der sich nach dem Abitur ausschließlich dem Studium widmet, macht einen geringen Teil unter Studierenden aus: 62 Prozent sind nebenbei erwerbstätig, 5 Prozent haben bereits Familie. Dass sich dadurch die Zahl der Studierenden in Fernstudiengängen zwischen 2003 und 2014 mehr als verdoppelt hat, zeigt, wie groß die Nachfrage nach flexiblen Studienmodellen ist. Für Hochschulabsolventen sind aber heute auch internationale und interkulturelle Erfahrungen während des Studiums ein Kriterium für den Erfolg auf dem Arbeitsmarkt. Digitale Lehr- und Lernelemente fördern die Auslandsmobilität, die bereits seit 1999 mit dem Bologna-Prozess angeschoben wurde.

Digital Turn an Hochschulen. Der Begriff „Digital Turn“ beschreibt die Änderungen und Auswirkungen, die sich aus der Nutzung der rasant entwickelnden Digitaltechnologie ergeben. Die „digitale Wende“ bezieht sich dabei zum einen auf die verschiedenen Prozesse bei der Umwandlung analoger Daten in digitale Form und zum anderen auf das Verständnis und die Verwendung der neuen Möglichkeiten. [6] Bei der Kernaufgabe Hochschullehre rückt die weitreichende Integration digitaler Medien und Technologien allmählich in den Bereich der Normalität. Institutionen erarbeiten schon seit Jahren Strategien, experimentieren mit digitalen Lehrszenarien und erforschen Lernprozesse mithilfe digitaler Datenerfassung und -auswertung. [7] Einige Hochschulleitungen fördern die Weiterentwicklung innovativer digitaler Lehr- und Lernformate besonders aktiv. Lehrende werden an diesen Hochschulen ermutigt, neue Wege zu gehen und erhalten Unterstützung und Ressourcen für die Umsetzung. [7]

Digitale Hochschule Mittweida. In Sachsen wurde mit der Bildungsportal Sachsen GmbH (BPS GmbH) schon 2004 der Grundstein für eine landesweite digitale Lehr- und Lernplattform gelegt, an der auch die Hochschule Mittweida als Projektpartner beteiligt ist. Ziel dieses Bündnisses ist es, eine Standardisierung bei der Entwicklung und Nutzung digitaler Lernanwendungen zu schaffen. Seitdem unterstützt die BPS GmbH das Lehrpersonal in Mittweida als e-learning-Dienstleister technologisch bei allen onlinebasierten Prozessen. [8] In Kooperation mit dem Sächsischen Staatsministerium für Wissenschaft und Kunst (SMWK), der Landesrektorenkonferenz Sachsen (LRK Sachsen) und dem Hochschuldidaktischen Zentrum Sachsen (HDS Sachsen) bietet dieses vernetzte Verbundprojekt aktuell ein Service-Angebot an, welches die Qualität der Lehre auch auf hochschuldidaktischer Ebene forciert und sichert.

Qualitätssicherung in der Lehre. Ebenfalls erklärtes Ziel der Bildungspolitik ist die Sicherung und Verbesserung der Lehrqualität. [9] Das BMBF stellt mit dem Qualitätspakt Lehre zwischen 2011 und 2020 insgesamt etwa zwei Milliarden Euro für Projekte zur Verbesserung der Lehrqualität an Hochschulen zur Verfügung. Innerhalb der digitalen Lehre heißt Lehrqualität, sich mit Hochschuldidaktik auseinander zu setzen. Gerade, wenn es um den Einsatz von Medien in der Lehre geht, ist es unumgänglich, pädagogische Konzepte zu entwickeln. Das bedeutet, dass die Weiterbildung zur Nutzung digitaler Medien in der Lehre grundlegend auf Didaktik und Pädagogik ausgelegt sein muss und erst im zweiten Schritt mit digitalen Technologien umgesetzt werden kann. Die Hochschule Mittweida versucht mit dem „Digitalpaket Lehre 2019“ zudem im Bereich der Lehre neue Zielgruppen zu generieren und bestehende zu begleiten. Finanzielle Förderungen sollen das digitale Lehrangebot erweitern und dazu beitragen, der steigenden Flexibilitätserwartung gerecht zu werden. Die Förderungen sollen aber auch die Qualität der Lehre und die akademische Konkurrenzfähigkeit erhöhen. [10] Darüber hinaus wurde Ende 2018 das Institut für Wissenstransfer und Digitale Transformation (IWD) gegründet. Mit dem IWD, welches das bisherige Institut für Technologie und Wissenstransfer (IT-WM) integriert, setzt die Hochschulleitung Ziele aus dem Hochschulentwicklungsplan 2025 zur Digitalisierung um und schafft ein Kompetenzzentrum, welches alle Bereiche der Hochschule unterstützt und als Träger für Pilotvorhaben fungieren soll. [10] Durch die Bologna-Reform

wurde die Qualitätsdiskussion auf Hochschulebene wiederbelebt. Seitdem tragen Akkreditierungen und Lehrveranstaltungsevaluationen zu deren Sicherstellung bei. Nicht zuletzt bedingt die zunehmende politische Relevanz von Hochschulrankings eine Vergleichbarkeit des Qualitätsverständnisses. Dabei spielen Forschung, internationale Ausrichtung und Studienbedingungen eine wesentliche Rolle, die in der Hochschule Mittweida über ein zentrales Informationsportal erfasst werden. [11] Für den Bereich Studium und Lehre sollen bis 2020 alle relevanten Strukturen und Prozesse aufgebaut sein, um die Voraussetzungen für eine Systemakkreditierung zu erfüllen. Die Arbeits- und Entwicklungsschwerpunkte liegen auch in den Bereichen der Digitalisierung.

1.1 Ausgangssituation der Studierenden

Zahlreiche Studien geben das Bild einer gehetzten Gesellschaft wieder. [12] Die Zeit spielt auch beim Lernen eine große Rolle. In möglichst kurzer Zeit sollen Informationen und Inhalte effektiv und auf das Wesentliche reduziert aufgenommen werden. Vielen Bereichen geht es da ähnlich. TED-Talks, Tutorials, Science Slams oder Erklärvideos sprechen für diese Entwicklung. Die deutsche Lese-App Blinkist zum Beispiel fasst das Wichtigste aus den bekanntesten und meist verkauften Fachbüchern in 15 min zusammen, als Hörbuch oder zum Lesen. [13] In Werbetexten für Blinkist ist vom Traum des täglichen Lernens nebenbei zu lesen: „Ob beim Warten in der Schlange, auf der Fahrt zur Arbeit oder vor dem Schlafengehen im Bett“. [14] Um nichts anderes geht es beim flexiblen Lernen.

Online lernen. Für die Generation Y müssen die Lehrinhalte aber auch unterhaltend sein. Deshalb sind digitale gamebasierte Lernmöglichkeiten (per App, oder über eine Webanwendung) beliebt geworden. Zwei Begriffe sind dabei essentiell, wenn über das Lernen mit digitalen Technologien gesprochen wird: e-learning und Edutainment. Mit E-Learning wird „die elektronische Unterstützung des Lernprozesses“ [15] beschrieben. Die Medien sind nur Mittel zum Zweck. Der entscheidende Vorteil bei e-learning-Inhalten ist, dass sie durch die Online-Veröffentlichung jederzeit abrufbar, also räumlich und zeitlich unabhängig sind und die Nutzer individuell entscheiden können, wann und wo sie lernen. Das ist auch das Ergebnis der Trendstudie „mmb Learning Delphi“. Mehr als die Hälfte der 65 Befragten ist

davon überzeugt, dass kleine Lerneinheiten, die on demand und mobilfähig zur Verfügung stehen, flexibler in den Arbeitsalltag integriert werden können. [16]

Spielerisch lernen. Als weiteren Trend der „mmb-Learning-Delphi-Studie“ nannten die Befragten interaktive Videos. Interaktive Videos werden den Begriffen edutainment, „computer based learning“ oder „serious games“ zugeordnet. Edutainment wird aus den zwei englischen Wörtern education und entertainment gebildet. [15] Bei Edutainment liegt der Schwerpunkt auf der Einbindung von unterhaltenden Elementen. Typische Gestaltungsmittel dafür sind Personalisierung, Visualisierung und erzählerische Elemente. Schon in den 1990er Jahren wurden digitale Spiele zur Wissensvermittlung eingesetzt. Es handelte sich damals um eher einfache Lernspiele, die hauptsächlich Vorschulwissen vermittelten. Digitale Spiele sind nach Wagner ein regelbasiertes, interaktives Medium, das Spielende emotional bindet und in einem von der Realität abgegrenzten Raum stattfindet. [17] Bei unterhaltenden Spielen können die Spieler die Zeit und ihr reales Umfeld regelrecht vergessen. Ein Grund dafür ist der schrittweise Aufbau der erforderlichen Kompetenzen in einem kontinuierlichen Zyklus aus Handlung und Rückmeldung. Solch eine intensive und selbstvergessene Auseinandersetzung mit dem Spielgegenstand wünschen sich Bildungsanbieter für ihre Lerninhalte. Deshalb sind sie daran interessiert, die Eigenschaften digitaler Spiele im Bildungskontext einzusetzen. [18] Durch die Möglichkeiten des technologiegestützten Lernens stellt Game-Based-Learning einen vielversprechenden Ansatz dar. Die Lernziele von Game-Based Learning gehen über das reine Verstehen und Speichern von Lerninhalten hinaus. Sie beinhalten auch den Erwerb von generischen und metakognitiven Fertigkeiten. [19]

Neue Lernkonzepte

Erklärfilme. Erklärfilme helfen dabei, komplexe Sachverhalte schneller zu erfassen und sich besser im Gedächtnis zu verankern, weil die Inhalte kurz und knapp, auf das Wesentliche beschränkt, erklärt sind – vergleichbar mit einem eingerahmten und farbig hinterlegten Merksatz in Lehrbüchern am Ende eines Themenkomplexes – nur eben in audiovisueller Form. Viele Begriffserklärungen im Netz beziehen sich auf die Marketingkommunikation, auf Unternehmen oder Organisationen, die

ihre Produkte in Erklärfilmen an ein breites Publikum verkaufen wollen. Eine Definition für den Erklärfilm im wissensvermittelnden Kontext gibt es nicht. Grundlegend besteht ein Erklärfilm aus Bewegtbild (real aufgenommen oder animiert), Geräusch, Text und Musik. Der Text in Erklärfilmen ist einfach und klar formuliert. Fremdwörter oder lange verschachtelte Sätze werden vermieden, damit sie schnell erfasst und verarbeitet werden können. Schlüsselbegriffe werden durch plakative Illustrationen visualisiert. Was die Länge der Videos angeht, ist die Devise: Je kürzer desto besser. Endlose 10-min-Filme finden in den sozialen Netzwerken keinen Anklang. Es ist also nicht verwunderlich, dass auch beim Lernen der Trend hin zum so genannten Micro-Learning geht. Studien zeigen, dass bei langen E-Learning-Einheiten, die ohne Unterbrechung absolviert werden, 60 Prozent des Gelernten nahezu sofort wieder vergessen werden. Diese "Vergessenskurve" [20] lässt sich mit kürzeren Lerneinheiten - auch learning nuggets genannt - unterbrechen. Die Lernenden zeigen dann deutlich bessere Ergebnisse in der Erinnerung der Inhalte. [21] Entscheidend für erfolgreiches Micro-Learning ist die didaktisch sinnvolle Aufbereitung des Lerninhalts. Didaktische Reduktion bezeichnet die Tätigkeit des Lehrenden, aus der Stofffülle den richtigen Stoff auszuwählen, sich auf das Wesentliche zu konzentrieren und Kompliziertes zu vereinfachen. Dennoch sind Erklärfilme längst nicht mehr die allgegenwärtige Option zur Wissensvermittlung für die Zukunft, weil ihnen der Aspekt der Unterhaltung, bzw. der Interaktion fehlt. Das reine passive Konsumieren eines Inhaltes rückt nach und nach in den Hintergrund und wird immer mehr durch aktives Mitgestalten des Inhaltes bestimmt. Mit modernen Technologien und interaktivem Storytelling ist es möglich, eine neue zeitgemäße Form des Lernens zu generieren.

Storytelling. Mit Storytelling bekommt der nüchterne Sachverhalt eine emotionale Bindung, z. B. eine aus dem Leben gegriffene Geschichte mit inszenierter Dramaturgie. Das Publikum wird so kognitiv und affektiv angesprochen, nimmt Informationen nicht nur rational auf, sondern auch emotional. In den Grundzügen geht es darum, durch die Verknüpfung von Informationen und Emotionen, Aufmerksamkeit zu erzielen. Storytelling existiert so lange es Menschen gibt. Am Lagerfeuer wurden existentielle Geschichten über die besten Jagdgründe oder über Auseinandersetzungen mit dem Nachbarland erzählt. [21] Der Begriff Storytelling setzt sich aus den englischen Wörtern für Geschichte (story)

und Erzählen (telling) zusammen, [22] bedeutet zunächst nichts anderes als Geschichten erzählen. Aber WIE ist die Frage. Viele Erklärfilme sind zwar nach dem Grundmuster der Dramaturgie aufgebaut, aber dennoch ohne emotionale Bindung. Es reicht nicht aus, die Geschichte an einer Hauptfigur aufzuziehen, aus deren Sicht dann der Sachverhalt erklärt wird. Beim Storytelling muss inszeniert werden, es braucht einen Spannungsbogen, Konflikte oder überraschende Momente, die humorvoll verpackt sind. Neben den Grundelementen Held, Handlung und Dramaturgie sorgen weitere Elemente für Emotionen. Schon im alten Griechenland wurde mit Witz und Humor die Gesellschaft kritisiert oder mit Kontrasten zwischen den Protagonisten wie Romeo und Julia geradezu ein Spannungsverhältnis provoziert. Ein Wissensvorsprung gegenüber den handelnden Personen lässt das Publikum mitfiebert und Überraschungsmomente oder plötzliche Wendungen lassen den Ausgang des Films nicht vorausahnen. Sogenannte Spiegelneuronen [23] sind nach Stefanie Reinberger auch ein Grund dafür, uns mitfiebert zu lassen. Diese Nervenzellen werden nämlich nicht nur aktiv, wenn wir selbst handeln oder empfinden, sondern auch dann, wenn wir eine Handlung oder eine Empfindung nur beobachten. Wir können uns also durch reines Beobachten emotional in die Lage anderer versetzen. Und diese Beobachtung wird schließlich auch als eigene Erfahrung im Gedächtnis abgespeichert. Unser Gehirn legt all diese Ereignisse, Erlebnisse und Informationen als sogenannte neuronale Muster im Unterbewusstsein ab. Das ist der Punkt, an dem Storytelling ansetzt, diese im Gehirn angelegte Wahrnehmung zu aktivieren, indem sie in eine Geschichte verpackt wird, die aus der eigenen Lebenserfahrung schöpft und dadurch für aktive Aufmerksamkeit sorgt.

2 Forschungsprojekt „Signale Frühgeborener erkennen und verstehen – Ein interaktives Trainingstool für Eltern frühgeborener Kinder (,SIGNAL-Train‘)“

Interactive Storytelling. Interactive Storytelling geht noch einen Schritt weiter. Im Gegensatz zum herkömmlichen Storytelling, bei dem der Zuschauer die Geschichte nur passiv erlebt, fordert interaktives Storytelling zum Mitmachen auf. Der Begriff wurde historisch durch die

Spieleindustrie geprägt, über seine Definition herrscht allerdings kein einheitlicher Konsens. Für Chris Crawford ist Storytelling dann interaktiv, wenn der Spieler (Lernende) selbst über den weiteren Verlauf der Handlung entscheidet. In der Wissensvermittlung kann mit Interaktion zum Beispiel das gerade gelernte Wissen abgefragt werden. Die Möglichkeiten, Wissen in einer interaktiven Struktur zu vermitteln, sollen in einem gemeinsamen Forschungsprojekt mit Eltern früh geborener Kinder am Universitätsklinikum Carl Gustav Carus in Dresden getestet und analysiert werden. Ziel ist es, zu untersuchen, inwieweit sich interaktives Storytelling in Erklärfilmen eignet, Wissen zu vermitteln.

Problemstellung. Ein Team aus Sozialpädagogen, Psychologen und Kinderkrankenschwestern auf der neonatologischen Station des Uniklinikums in Dresden begleitet und berät vor allem frühgeborene Kinder und ihre Eltern vom ersten Tag der Geburt an bis zur Entlassung aus dem Krankenhaus. Eine Frühgeburt kommt für die Eltern meist überraschend. Oft fehlen zehn oder mehr Wochen, in denen das Kind im Uterus noch wichtige Erfahrungen machen könnte. Wenn es zu früh geboren wird, bestehen neben der organischen Unreife auch große Anpassungsprobleme in Bezug auf die neue Umwelt – in der es sich nicht wie normal Geborene äußern kann. Diese Zeichen (Signale) der Kinder müssen die Eltern erkennen und richtig deuten lernen. Im Mittelpunkt der Begleitung steht deshalb auch der frühestmögliche Aufbau sicherer Eltern-Kind-Bindungen mit allen positiven Auswirkungen auf die gesundheitliche Entwicklung des Kindes und die familiäre Lebensqualität. Zu diesem Zweck bietet das Team den Eltern unter anderem ein vielfältiges Elternschulungsprogramm aus persönlichen Gesprächen, regelmäßigen Aufklärung- und Auswertungsgesprächen von Videoaufzeichnungen an. Dem Personal steht dafür aber immer weniger Zeit zur Verfügung, auch fehlen die Räumlichkeiten, um individuelle Gespräche mit den Eltern zu führen. So entstand die Initiative, in einem gemeinsamen Forschungsprojekt, Videosequenzen herzustellen, welche die verschiedenen Reaktionen der frühgeborenen Kinder beim Waschen, Füttern oder Windeln zeigen und erklären. Mit Hilfe eines Trainingstools soll die Eltern-Kind-Interaktion frühzeitig präventiv unterstützt, die Voraussetzungen für eine optimale Bindungsentwicklung gefördert und die neurokognitive Entwicklung des Kindes verbessert werden. Ein weiteres Ziel bei diesem Vorhaben besteht darin, die Videosequenzen softwaretechnisch so zu verknüpfen, dass der Editor inter-

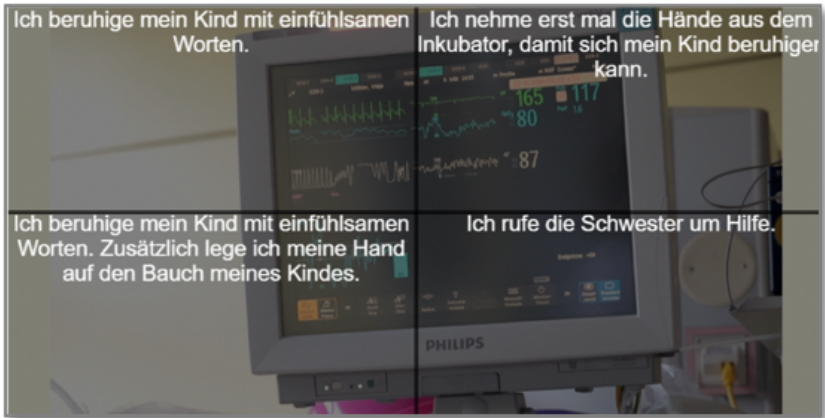


Abbildung 11.1: Videoeditor mit Abfrageinteraktion nach dem 1. Semester [24]

aktiv die Reihenfolge der Videosequenzen festlegen kann. Ein Beispiel zum Verständnis: Der Lernende schaut sich die Videoszene an, in der zu sehen ist wie ein frühgeborenes Kind im Inkubator gewickelt wird. Plötzlich ist ein Piep-Ton zu hören und eine erhöhte Herzfrequenz am angeschlossenen Herzfrequenzgerät zu sehen. Der Film bricht an dieser bestimmten Stelle ab. Es erscheint die Frage: "Wie würden Sie in dieser Situation mit Ihrem Kind umgehen?" Zur Beantwortung der Frage erscheinen vier verschiedene Möglichkeiten. Der Lernende muss sich nun für eine Antwort entscheiden und der Film läuft entsprechend anders weiter. Danach folgt in einer weiteren Videosequenz die Auswertung.

Aufgabenstellung. Dieses Forschungsprojekt wurde über zwei Semester 2018/2019 im Modul Wissenschaft und Wirtschaft (WW) des Studiengangs Medieninformatik und interaktives Entertainment in der Hochschule Mittweida von einem vierköpfigen Team umgesetzt und von Professoren und Mitarbeitern [25] der Fakultäten Medien und Angewandte Computer- und Biowissenschaften betreut. Das Videomaterial wurde im Vorfeld von Mitarbeitern der Fakultät Medien auf der neonatologischen Station abgedreht und lag den Studierenden zu Semesterbeginn vor. Nach der Erstellung eines Anforderungskatalogs, eines Technologiekonzeptes und einem Semesterzeitplan entschied sich das

Team dafür, die interaktive Verknüpfung der Videosequenzen nicht mit einer bestehenden Open-Source-Software zu generieren, sondern nahm sich der Herausforderung an, selbst ein Programm mit Editor zu entwickeln.

Bisherige Ergebnisse. In der ersten Phase entstand ein voll funktionsfähiger interaktiver Prototyp für die Evaluierung. Der Bearbeiter bzw. Lehrende konnte in dem Videoeditor eigene Videosequenzen einfügen und miteinander verknüpfen sowie eingegebene Daten der Lernenden (Zeiterfassung während der Interaktion und Nutzerbefragung) auswerten und evaluieren. Die Evaluation enthält zum einen die technische Funktionalität und zum anderen einen Fragebogen, indem das Lernen mit interaktivem Storytelling abgefragt wird. Kann der Lernende dem Ablauf des Filmes folgen? Ist der Inhalt verständlich? Auch das Wahrnehmen multisensorischer Gestaltebenen, die bei der Wissensaufnahme eine ausschlaggebende Rolle spielen, wird abgefragt. Während beim Hören einer Information etwa 20 Prozent des Inhaltes im Gedächtnis gespeichert werden, sind es beim Lesen 30 Prozent. Beim Hören und Lesen 50 Prozent und wenn Information beispielsweise durch Abfragemechanismen wiederholt werden, sogar 70 Prozent. [20] Es wird angenommen, dass je mehr Kanäle bei der Informationsaufnahme gleichzeitig angesprochen werden, desto wahrscheinlicher ist es, dass diese Informationen im Gehirn abgespeichert werden. Es wird weiterhin angenommen, dass durch das Erleben einer Situation von anderen Menschen, die wie im Forschungsprojekt ähnliche Situationen mit früh-geborenen Kindern kennen und gleiche Ziele wie die Lernenden verfolgen, sich die Lernenden besser mit den Figuren im Film identifizieren und sich in die Gefühle hineinversetzen kann. Diese Beobachtungen, die zugleich Lernstoff in der Wissensvermittlung sind, werden als eigene Erfahrung abgespeichert und langfristig im Gedächtnis verankert.

In der zweiten Phase wurden Änderungswünsche aus der internen Evaluation aus der Sicht des Bearbeiters bzw. Lehrenden eingearbeitet. Das Design der Anwendung wurde ansprechender und intuitiver gestaltet, der Editor um Aussehen und neuen benutzerfreundlichen Funktionen erweitert. Aus einer statischen Verknüpfung der Videos im Editor ist am Ende des zweiten Semesters eine dynamische entstanden. Der Lehrende kann die Videosequenzen nun per drag and drop intuitiv und beliebig miteinander verknüpfen. Die Auswertung des Fragebogens wird jetzt



Abbildung 11.2: Videoeditor mit Videosequenzen und Verknüpfung nach 2. Semester [26]

automatisch grafisch dargestellt. Der Lehrende kann sehen, für welche Antwortoption sich der Lernende entschieden hat und wie viel Zeit er dafür verwendet hat. Seit Anfang März 2019 liegt das Tool nun den Eltern auf der neonatologischen Station zur Evaluierung vor. Die Ergebnisse stehen in wenigen Monaten zur Auswertung zur Verfügung.

Literatur

1. "Kern-Ergebnisse der ARD/ZDF - Onlinestudie," 2016. [Online]. Available: <http://www.ard-zdf-onlinestudie.de/files/2016/Kern-Ergebnisse-ARDZDF-Onlinestudie.2016.pdf>
2. mmb-Institut Gesellschaft für Medien-und Kompetenzforschung mbH, "Erklärfilme als Umsatzbringer der Stunde Ergebnisse der 12. Trendstudie „mmb Learning Delphi“," 2017/2018. [Online]. Available: <https://www.mmb-institut.de>
3. T. Bauer, "Video Content: Wie die User Bewegtbild konsumieren wollen." [Online]. Available: <https://onlinemarketing.de/news/video-content-hubspot-report?ut>

4. S. Feierabend, T. Plankenhorn, and T. Rathgeb, "JIM 2017 Jugend, Information, (Multi-) Media, Basisstudie zum Medienumgang 12- bis 19-Jähriger in Deutschland," 2017.
5. S. Höfer, I. Schünemann, and O. Janoschka, "Hochschulreform Digitalisierung, The digital Turn, Hochschulbildung im digitalen Zeitalter, Arbeitspapier Nr. 27." 2016.
6. DGF, "The Digital Turn in the Sciences and Humanities," 2018. [Online]. Available: http://www.dfg.de/en/research_funding/principles_dfg_funding/digital_turn/index.html
7. S. Höfer, I. Schünemann, and O. Janoschka, "Hochschulreform digitalisierung, the digital turn, hochschulbildung im digitalen zeitalter," 2016.
8. B. S. GmbH, "BPS Bildungsportal Sachsen GmbH." [Online]. Available: www.bps-system.de/cms/produkte/
9. BMBF, "Qualitätspakt Lehre." [Online]. Available: <https://www.bmbf.de/de/qualitaetspakt-lehre-524.html>
10. "Die digitale Hochschule. Informationen zum Digitalpaket 2019," 2018. [Online]. Available: www.rektorat.hs-mittweida.de/prorektorat-hochschulentwicklung/digitale-hochschule.html
11. M. Klinger, "Merkmale guter Hochschullehre: Definitionsversuche und Operationalisierbarkeit," 2011.
12. R. Ball, "Die pausenlose Gesellschaft: Fluch und Segen der digitalen Permanenz," *Schattauer GmbH, Stuttgart*, 2014.
13. Blogkommentar, "E Mein Blinkist Erfahrungsbericht – Fragen, Antworten und ob es sich lohnt." [Online]. Available: <https://www.schreibsuchti.de/2014/05/15/mein-blinkist-erfahrungsbericht-fragen-antworten-und-ob-es-sich-lohnt/>
14. "Blinklist." [Online]. Available: <https://www.blinkist.com/de>
15. S. Korupp, "Möglichkeiten der digitalen interaktiven Aufbereitung musealer Inhalte und Entwicklung eines Prototyps am Beispiel der Valdivia- Expedition am Museum für Naturkunde Berlin," 2018.
16. mmb Branchenmonitor, "E-Learning-Wirtschaft 2017, mmb-Studie (explain it GmbH): Neues Arbeiten braucht neues E-Learning," 2017.
17. M. Ebener and S. Schön, *Faszination Computerspielen. Theorie - Kultur - Erleben. Interaktionstechnologie im gesellschaftlichen Spiel.* in K. Mitgutsch & H. Rosenstingl, 2013.
18. M. Bopp, "Immersive Didaktik: Verdeckte Lernhilfen und Framingprozesse in Computerspielen. In: kommunikation@gesellschaft," 2005. [Online]. Available: <http://www.soz.uni-frankfurt.de/K.G/B2.2005.Bopp.pdf>

19. M. Kerres, M. Bormann, and M. Vervenne, "Didaktische Konzeption von Serious Games. Zur Verknüpfung von Spiel- und Lernangeboten." 2009. [Online]. Available: <http://www.medienpaed.com/2009/kerres0908.pdf>
20. K. Schroer, "Die Psychologie des Lernens – Wie funktioniert Lernen?"
21. "Micro-learning, warum weniger Mehr ist." [Online]. Available: <https://www.explain-it.tv/de/blog/artikel/2017/12/11/micro-learning-warum-weniger-mehr-ist/>
22. P. Kleine Wieskamp, "Storytelling: Digital - Multimedial - Social," *Carl Hanser Verlag GmbH & Co. KG*, 2006.
23. S. Reinberger, "Spieglein, Spieglein im Gehirn," 2011. [Online]. Available: <https://www.dasgehirn.info/denken/im-kopf-der-anderen/spieglein-spieglein-im-gehirn>
24. A. Schmidt, S. Schmelzer, Giesa, and L. Ullrich, "Signal-Train: 1. Semester Abschlusspräsentation," 2019.
25. Holger Langner, Marc Ritter, Wilfried Schubert, Matthias Vodel, Klaus Liepelt, Rika Fleck, "Aufgabenstellung im Modul Wissenschaft & Wirtschaft 3 und 4 im Studiengang Medieninformatik & Interaktives Entertainment," *Hochschule Mittweida*, 2018.
26. A. Schmidt, S. Schmelzer, Giesa, and L. Ullrich, "Signal-Train: 2. Semester Abschlusspräsentation," 2019.

Eine netzwerkanalytische Aufarbeitung des Nahverkehrs in Mittelsachsen

Christoph Schulte

Hochschule Mittweida, Fakultät Medien, D-09648 Mittweida,
schulte@hs-mittweida.de

Zusammenfassung Der Zugang zu mobilen Bewegungsdaten durch Smartphone-Sensoren ermöglicht die Anwendung neuer Methoden zur Analyse des regionalen Nahverkehrs. Im Folgenden wird ein netzwerktheoretisches Konzept entwickelt, mit dessen Hilfe sich Verkehrsströme als Netzwerk strukturieren lassen können. Das Verkehrsnetz wird dabei im Sinne Whites als Identität betrachtet. Dieser theoretische Hintergrund bietet die Basis zur Entwicklung zukünftiger Verfahren zur Planung, Modellierung und Simulation des Nahverkehrs.

Schlagwörter Netzwerkanalyse, Nahverkehr, Blockmodelle

1 Einleitung

Neue Erhebungsmodi für mobile Sensordaten und Analysemethoden aus dem Bereich der Netzwerkforschung finden zahlreiche Anwendungen in der Verkehrsforschung. Das Ziel ist eine bestmögliche Modellierung des Nahverkehrs insbesondere in urbanen Räumen um die zukünftige Verkehrsplanung zu optimieren. Auffallend bei den bisherigen Ergebnissen ist, dass es an einer geeigneten interdisziplinären Annäherung an das Thema fehlt. Netzwerkanalytische Arbeiten bauen zwar häufig auf eine fundierte theoretische Grundlage auf, beschränken sich in der Analyse jedoch lediglich auf Netzwerke des öffentlichen Personennahverkehrs (ÖPNV) [1–3]. Arbeiten mit computergestützten mathematischen Verfahren sammeln über Geodaten aus Mobiltelefonen zwar Bewegungsprofile aus dem gesamten Nahverkehrsnetz, hier fehlt es allerdings häufig an einer fundierten theoretischen Grundlage

sowie einer validen Annotation des genutzten Verkehrsmittels (Modal Split) [4–6].

Im Folgenden soll daher ein Arbeitskonzept vorgestellt werden, dass die Vorteile beider Forschungsrichtungen verbindet. Nach einem Überblick über den aktuellen Forschungsstand wird ein netzwerktheoretisches Konzept vorgestellt, das auf den Anwendungsfall des Nahverkehrs übertragen wird. Die leitende Forschungsfrage, die sich mit dieser Grundlage beantworten lassen soll, lautet: Wie strukturiert sich das Netzwerk des Nahverkehrs in Mittelsachsen? Eine strukturelle Analyse kann mit Hilfe von Blockmodellen durchgeführt werden. Diese Ergebnisse können als Grundlage für neue Simulationsverfahren des Nahverkehrs dienen.

2 Forschungsstand

Die Entwicklung und Strukturierung des Verkehrs wird intensiv erforscht, wobei sich die meisten netzwerkanalytischen Arbeiten auf die Untersuchung des öffentlichen Personennahverkehrs (ÖPNV) beschränken. Diese Beschränkung ist zwar methodisch nachvollziehbar, vernachlässigt jedoch einen Großteil der Publikumsbewegungen im Kleinen.

Durch eine Analyse standardisierter Graphendarstellungen von Netzwerken des ÖPNV in 14 Großstädten identifizieren Von Ferber et al. [1] generalisierbare Eigenschaften der Verkehrswege, die als Grundlage für weitere Modellierungen im Bereich der Verkehrsforschung dienen können. Durch die zusätzliche Berücksichtigung der Gewichtung und Gerichtetheit der Graphen lassen sich zudem genaue Aussagen über die wichtigsten Verkehrslinien und Haltestellen einer Stadt tätigen [2].

Soh et al. [7] heben in ihrer Analyse des Bus- und Bahnnetzes von Singapur die Bedeutung der Dynamik hervor. Demnach genüge es nicht, sich auf rein topologische Analysen zu beschränken, da die dynamische Analyse des Verkehrs innerhalb einer Woche zusätzlich weitreichende Erkenntnisse über die Beschaffenheit des Netzes bringe. Eine netzwerkanalytische Aufarbeitung des ÖPNV-Netzes könne in Verbindung mit numerischen Simulationsverfahren zudem neue Möglichkeiten für Verkehrsplaner eröffnen, um die Netzanbindung zu optimieren und dem

Bedarf der Nutzer anzupassen [8].

Eine weitere Möglichkeit, Wegstrecken im Verkehrsnetz zu optimieren finden Zou et al. [9] mit einem agentenbasierten Modell, das auf Umfragedaten beruht. Ihre Arbeit ermöglicht die Berücksichtigung des Modal Split, scheitert allerdings an der Modellierung der Struktur des Verkehrsnetzes als solches.

Eine aussichtsreiche Alternative für neue Analysemethoden bieten anonymisierte Bewegungsprofile, die aus Mobilfunkdaten gewonnen werden können [5, 10]. Einigkeit besteht darin, dass die so gewonnenen Daten auch für die Verkehrsforschung eine wichtige Rolle spielen können, allerdings fehle es bislang an praktikablen Anwendungen [6]. Toole et al. [11] haben bereits erste Versuche unternommen, anonymisiert gewonnene Geodaten mit netzwerktheoretischen Konzepten zu analysieren und zu interpretieren. Eine weitere methodisch wichtige Entwicklung ist Alexander et al. [4] gelungen, die mit Hilfe von Inferenzen aus Bewegungsaktivitäten strukturell wichtige Positionen wie Wohnumgebungen und Arbeitsplätze identifizieren können.

Es zeigt sich insgesamt also, dass ein breites Interesse daran besteht, auf Grundlage innovativer analytischer Methoden und neuer Erhebungsformen besseres Verständnis über den Nahverkehr zu gewinnen. Netzwerkanalytische Arbeiten stützen sich meist auf Fahrgastdaten des ÖPNV. Um ein genaues Bild über den Nahverkehr als Gesamtes zu gewinnen fehlen allerdings wichtige Informationen: Zum einen die genauen Wegstrecken der Nutzer des ÖPNV, zum anderen ein großer Teil der Verkehrsteilnehmer generell, die nicht den ÖPNV nutzen. Aus den Geodaten von Mobilfunkgeräten lassen sich andererseits zwar genaue Wegstrecken modellieren, bislang fehlt allerdings eine Möglichkeit, Aussagen über die genutzten Verkehrsmittel zu treffen. Ebenso fehlt es den computergestützten mathematischen Verfahren häufig an einer fundierten theoretischen Grundlage, die Aussagen über die Struktur des Nahverkehrs zulassen.

Es scheint daher lohnenswert, die positiven Aspekte der bisherigen Forschungsansätze zu verbinden, um so neue Erkenntnisse über die Struktur des gesamten Nahverkehrsnetzwerks zu erhalten. Auf Grundlage eines fundierten netzwerktheoretischen Konzepts können kleinteilige Bewegungsdaten im Gesamtsystem analysiert und interpretiert werden.

3 Methodisches Konzept

Einen passenden Rahmen zur theoretischen Einordnung kleinteiliger Bewegungen im öffentlichen Raum bieten Harrison Whites Ausarbeitungen über „Identity and Control“ [12, 13]. Seine zentralsten Begriffe sind Identitäten, Kontrolle, Disziplin sowie die strukturelle Äquivalenz. Im Folgenden sollen die unterschiedlichen Konzepte zunächst erläutert werden und anschließend auf das Verkehrssystem übertragen werden, bevor dann ein geeignetes Analyseverfahren vorgestellt wird um das Netzwerk des Nahverkehrs zu strukturieren.

Der grundlegendste Begriff für die relationale Theorie Whites ist der der strukturellen Äquivalenz [14]. Mit Hilfe dieses Konzepts abstrahiert er vom Individuum auf eine höhere Struktur, die sich mit Hilfe der Blockmodellanalyse identifizieren lässt [15]. Als strukturell äquivalent bezeichnet White dabei Akteure eines Netzwerkes, die ein ähnliches Beziehungsprofil zu den anderen Akteuren des Netzwerkes haben [16]. Lorrain und White [14, 49] konzentrieren sich demnach auf „interrelations among relations within concrete groups“, wodurch sie äquivalente Positionen innerhalb einer Netzwerkstruktur aufdecken. Im Fokus stehen also nicht direkte Beziehungen zwischen bestimmten Individuen, sondern das Aufdecken ähnlicher Beziehungsmuster in einem Netzwerk. In Bezug auf das Netzwerk des Nahverkehrs sind dementsprechend Wege als strukturell äquivalent zu bezeichnen, die zur Erfüllung eines bestimmten Zweckes mit dem gleichen Verkehrsmittel zurückgelegt werden. Das Konzept lenkt also den Fokus weg von dem Handeln des einzelnen Verkehrsteilnehmers, sondern konzentriert sich im angewandten Fall auf die sinnbeladene Nutzung eines Weges im Gesamtnetzwerkes des Nahverkehrs.

Identitäten im Sinne Whites müssen nicht zwangsläufig personen- oder objektgebunden sein, sondern werden generalisierter als Quelle von Aktionen oder Entitäten verstanden, in die sinnhafte soziale Bedeutung interpretiert werden kann [13, 2]. In Bezug auf ein Verkehrssystem können also auf unterschiedlichen Ebenen viele Identitäten interpretiert werden. Je nach Situation kann ein einzelner Verkehrsteilnehmer als Identität wahrgenommen werden, zurückgelegte Wegstrecken zum Arbeitsplatz können eine weitere Identität bilden, ebenso können Passagiere in einem Bus oder einer Bahn als Identität zusammengefasst werden. In der umfassendsten Form kann das gesamte Verkehrsnetz

als eigenständige Identität erfasst werden, die tagtäglich musterhafte Abläufe wiederholt [17].

Der Begriff der Kontrolle hat bei Harrison White eine mehrschichtige Bedeutung und ist insgesamt sehr unscharf definiert:

“Control is both anticipation of and response to eruptions in enviroining process. Control projects participate in how identities array in social structures, with social order as a possible by-product. Social processes and structure are thus traces from successions of control efforts.” [13, 7]

Identitäten und Kontrolle bedingen sich demnach gegenseitig. Zum einen setze Kontrolle immer existierende Identitäten als Ziel und Ausgangspunkt voraus, zum anderen können sich diese Identitäten nur ausbilden, wenn sie durch Kontrolle in einer bestimmten Position gehalten werden [16, 70]. Kontrolle hat für Identitäten eine ordnende und stabilisierende Funktion.

In unserem Anwendungsfall erhält die Identität des Nahverkehrsnetzwerkes eine regelmäßige Ordnung durch bestimmte Kontrollmechanismen. So führen bestimmte Regeln dazu, dass das Verkehrsnetz bestimmten Mustern folgt, die es ohne Kontrolle nicht geben würde. Feste Arbeitszeiten führen dazu, dass sich Bewegungsflüsse zu bestimmten Zeiten ähneln. Fahrpläne im ÖPNV stabilisieren die Wege und Fahrtzeiten der Busse und Bahnen. Ampelschaltungen bestimmen den Fluss des Autoverkehrs. Viele kleine soziale Regeln, Prozesse und Ordnungen, die zunächst obsolet erscheinen, führen also dazu, dass sich das Verkehrsnetz erst als stabile Identität entwickeln kann. Ohne Ordnung würde ein chaotisches System entstehen, dass keine Struktur erkennen lässt.

Das letzte theoretische Konzept, das an dieser Stelle beschrieben werden soll, ist das der Disziplin. White beschreibt Disziplinen als Prozesse, die Transaktionen zwischen und innerhalb von Identitäten koordinieren und dadurch stabilisieren [13, 63]. Diese Stabilisierung ermöglicht die Ausbildung sozialer Moleküle, die selbst aus unterschiedlichen Identitäten bestehen und auf Grundlage ihrer verliehenen Stabilität zu einer neuen Identität verschmelzen können [16, 70 f.].

In unserem Anwendungsfall beschreibt die interface-Disziplin, wie sich das Verkehrsnetz hinsichtlich qualitativer Kriterien organisiert [13, 64]. Transaktionen, also das Zurücklegen von bestimmten Wegstrecken

zu einer bestimmten Zeit, werden permanent hinsichtlich ihrer Qualität kontrolliert. Falls sich eine andere mögliche Strecke oder ein anderes Verkehrsmittel als besser als das bisher genutzte zeigt, um von einem Punkt zu einem anderen Punkt zu gelangen, so wird diese Alternative gewählt. White orientiert sich in der Beschreibung der interface-Disziplin an Kontrollprozessen in wirtschaftlichen Produktionsordnungen.

Mit dieser theoretischen Grundlage ist es möglich, Bewegungsdaten strukturell zu analysieren. Ein geeignetes Analyseverfahren entwickelten White et al. mit der Blockmodellanalyse [15]. Mit dieser Methode wird die Soziomatrix eines Netzwerks durch Permutationen so verändert, dass sich strukturelle Äquivalenzen blockweise identifizieren lassen. Diese permutierte Soziomatrix lässt sich auf eine Imagematrix reduzieren, in der nicht mehr einzelne Beziehungen, sondern spezifische Positionen im Netzwerk dargestellt werden [18]. Diese Positionen spiegeln die Struktur des Netzwerks wider. Da es sich für den Anwendungsfall von Verkehrsnetzwerken um komplexe Systeme handelt, soll ein stochastisches Blockmodell Anwendung finden, das die dynamische Entwicklung des Netzwerkes berücksichtigen kann [19].

4 Datengrundlage

Die Datenbasis für die strukturelle Analyse sollen Bewegungsdaten aus dem Raum Mittelsachsen bilden, die im Rahmen des Flowbile-Projekts der Hochschule Mittweida erhoben wurden [17]. Im Rahmen des Projekts wurden über Smartphone-Sensoren ca. 25.000 Wegstrecken freiwilliger Probanden im Raum Mittelsachsen und Leipzig anonymisiert aufgezeichnet [17]. Im Aufbereitungsprozess wurden bereits Informationen über den wahrscheinlichsten Modal Split jeder Wegstrecke berechnet, sodass zwischen fünf unterschiedlichen Bewegungsarten unterschieden werden kann.

In einem folgenden Schritt sollen in Anlehnung an Alexander et al. [4] die wichtigsten Ruhepunkte der Teilnehmer inferenzstatistisch identifiziert und klassifiziert werden. Auf diese Weise lassen sich Wohnorte und Arbeitsplätze der Teilnehmer bestimmen. Mit Hilfe von Daten aus Geoinformationssystemen sollen außerdem weitere wichtige Orte wie Einkaufszentren oder Freizeitstätten identifiziert werden. Sowohl der

Modal Split, als auch die zu identifizierenden Geopunkte sind wichtige Ordnungsparameter, die das Verkehrsnetz als Ganzes stabilisieren.

Mit den so gewonnenen Informationen lassen sich die Bewegungen jedes einzelnen Teilnehmers kleinteilig in Cluster zusammenfassen. Diese Cluster können anschließend die Grundlage der Blockmodellanalyse bilden.

5 Ausblick

Die vorliegende Arbeit beschreibt neue konzeptionelle Ansätze zur Analyse des Nahverkehrs. Es konnte gezeigt werden, dass zwar ein reges wissenschaftliches Interesse an der Analyse von Verkehrsdaten besteht. Das Fehlen von geeigneten theoretischen Annahmen führt allerdings dazu, dass es an aussagekräftigen Ergebnissen, die eine praktische Anwendung folgen lassen können, mangelt.

Mit Hilfe des netzwerktheoretischen Ansatzes Whites wird eine passende Grundlage zur Analyse des Nahverkehrs gegeben. Zusammen mit den Daten des Flowbile-Projektes bietet die Theorie die Möglichkeit, den Nahverkehr in Mittelsachsen zu strukturieren und so die Grundlage für weitere Anwendungen zu schaffen.

Insbesondere aus verkehrsplanerischer Sicht können die so gewonnenen Informationen wertvoll werden. Sie können zum Beispiel als Basis für neue Simulationsverfahren genutzt werden, um das Teilnetzwerk des öffentlichen Personennahverkehrs zu optimieren und Fahrpläne bedarfsgerecht anzupassen. Der entwickelte Ansatz bringt außerdem den Vorteil mit sich, dass er generalisierbar auf andere Verkehrsnetze übertragbar ist, sobald eine passende Datengrundlage geschaffen wurde. Die Aufbereitung und Anreicherung der Daten des Flowbile-Projektes stellen den nächsten Schritt dar um zu zeigen, dass die vorgestellte Analyse ihrem Anspruch gerecht wird.

Danksagung

Diese Arbeit wurde gefördert durch den Europäischen Sozialfonds (ESF) und den Freistaat Sachsen.

Literatur

1. C. von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov, "Public transport networks: Empirical analysis and modeling," *The European Physical Journal B*, vol. 68, no. 2, pp. 261–275, 2009.
2. A. Haznagy, I. Fi, A. London, and T. Nemeth, "Complex network analysis of public transportation networks: A comprehensive study," in *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 2015, pp. 371–378.
3. T. Shanmukhappa, I. W.-H. Ho, and C. K. Tse, "Spatial analysis of bus transport networks using network theory," *Physica A: Statistical Mechanics and its Applications*, vol. 502, pp. 295–314, 2018.
4. L. Alexander, S. Jiang, M. Murga, and M. C. González, "Origin–destination trips by purpose and time of day inferred from mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 240–250, 2015.
5. S. Jiang, G. A. Fiore, Y. Yang, Ferreira, Joseph Jr., Frazzoli, Emilio, and M. C. González, "A review of urban computing for mobile phone traces: current methods, challenges and opportunities," *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, 2013.
6. C. G. Prato, "Route choice modeling: Past, present and future research directions," *Journal of Choice Modelling*, vol. 2, no. 1, pp. 65–100, 2009.
7. H. Soh, S. Lim, T. Zhang, X. Fu, G. K. K. Lee, T. G. G. Hung, P. Di, S. Prakasham, and L. Wong, "Weighted complex network analysis of travel routes on the singapore public transportation system," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 24, pp. 5852–5863, 2010.
8. X.-l. An, L. Zhang, and J.-g. Zhang, "Research on urban public traffic network with multi-weights based on single bus transfer junction," *Physica A: Statistical Mechanics and its Applications*, vol. 436, pp. 748–755, 2015.
9. M. Zou, M. Li, X. Lin, C. Xiong, C. Mao, C. Wan, K. Zhang, and J. Yu, "An agent-based choice model for travel mode and departure time and its case study in beijing," *Transportation Research Part C: Emerging Technologies*, vol. 64, pp. 133–147, 2016.
10. J. L. Toole, C. Herrera-Yaqüe, C. M. Schneider, and M. C. González, "Coupling human mobility and social ties," *Journal of the Royal Society, Interface*, vol. 12, no. 105, 2015.
11. J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González, "The path most traveled: Travel demand estimation using big data resources," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 162–177, 2015.

12. H. C. White, *Identity and control: A structural theory of social action*. Princeton, NJ: Univ. Press, 1992.
13. —, *Identity and control: How social formations emerge (Second Edition)*. Princeton: Princeton University Press, 2012.
14. F. Lorrain and H. C. White, "Structural equivalence of individuals in social networks," *The Journal of Mathematical Sociology*, vol. 1, no. 1, pp. 49–80, 1971.
15. H. C. White, S. A. Boorman, and R. L. Breiger, "Social structure from multiple networks. i. blockmodels of roles and positions," *American Journal of Sociology*, vol. 81, no. 4, pp. 730–780, 1976.
16. M. Schmitt and J. Fuhse, *Zur Aktualität von Harrison White*. Wiesbaden: Springer Fachmedien Wiesbaden, 2015.
17. M. Czerny, M. Denk, H. Höhne, G. Kind, S. Leuoth, K. Liepelt, and M. Schubert, *Flowbile. Kleinteilige Publikumsforschung im sächsischen Personennahverkehr: Bericht über Verlauf und Ergebnisse eines Drittmittelprojekts der Hochschule Mittweida mit Zweckbestimmung "Relationale Publikumsforschung"*, Mittweida, 2018.
18. R. Heidler, *Die Blockmodellanalyse*, 1st ed. Wiesbaden: DUV Deutscher Universitäts-Verlag, 2006.
19. T. P. Peixoto, "Inferring the mesoscale structure of layered, edge-valued, and time-varying networks," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 92, no. 4, p. 042807, 2015.

Wissenstransfer in der Bürgerbeteiligung auf kommunaler Ebene

Christopher M. Brinkmann

Hochschule Mittweida, Fakultät Medien, D-09648 Mittweida,
brinkmann@hs-mittweida.de

Zusammenfassung Das Wissen der Akteure in Beteiligungsprozessen (z.B. Stadträte, Vereine, lokale Unternehmen, Bürgerinnen und Bürger) stellt eine wichtige Grundlage für kommunale Entwicklungen dar. Bisherige Partizipationsmodelle benennen zwar, dass im Rahmen von Bürgerbeteiligungsprozessen das Fach- und Expertenwissen politischer Entscheider durch das Erfahrungswissen der Bürgerinnen und Bürger ergänzt wird. Eine weitere Vertiefung, wie dieser Wissenstransfer gestaltet und für zukünftige Entwicklungen tradiert werden kann, erfolgt allerdings nicht. Auf Grundlage theoretischer Konzepte (Wissensmanagement, lernende Organisation, Bürgerkommune) und ergänzender empirischer Daten wird in einem kooperativen Promotionsprojekt zwischen der Hochschule Mittweida und der Universität Bremen der Ansatz einer wissensorientierten Beteiligungskommune entwickelt. Das Paper gibt einen Einblick in das Themenfeld.

Schlagwörter Wissensmanagement, Bürgerbeteiligung, Wissenstransfer, Beteiligungskommune

1 Einleitung: Bürgerbeteiligung im Wandel

Im deutschen repräsentativen Demokratiemodell werden politische Sachentscheidungen sowie Entwicklungsplanungen durch Abgeordnete in Gremien, Parlamenten oder Räten getroffen und abgestimmt. Bürgerinnen und Bürger geben mit der Teilnahme an Wahlen den Repräsentanten die Legitimität, als Vertreter des Volkes entsprechende Entscheidungen eigenverantwortlich zu treffen. Die Bürgerbeteiligung

an diesem System beruht, neben der Möglichkeit mit den Abgeordneten zum Beispiel bei Bürgersprechstunden in Kontakt zu kommen, auf dem aktiven und passiven Wahlrecht. In den vergangenen Jahren sind allerdings auch immer wieder Bemühungen zu sehen, die repräsentative Demokratie durch direktdemokratische (z.B. Bürgerentscheide) oder deliberative Bürgerbeteiligungsformen (dialogorientierte Partizipationsprozesse) zu ergänzen und weiterzuentwickeln [1, S.18] [2, S.94]. Dabei stehen vor allem drei große Trends als Ausgangspunkte dieser Reformbestrebungen. Zum einen werden die Politikverdrossenheit, Abstinenz bei Wahlen und die daraus resultierende fehlende Legitimierung der Abgeordneten als Ausgangspunkt der Forderung nach mehr Bürgerbeteiligung genannt [3, S. 243]. Bürgerinnen und Bürger sollen durch Möglichkeiten der Einflussnahme aus ihrem politischen Desinteresse geholt und zur demokratischen Teilnahme aktiviert werden [4, S. 388]. Dem gegenüber stehen die Bürgerproteste, für die die Demonstrationen gegen Stuttgart 21 häufig als Beispiel genannt werden [5, S. 18] [1, S. 18]. Mit Bürgerbeteiligungsprozessen sollen Meinungen und Vorbehalte in der Bevölkerung in einem Planungsprozess frühzeitig aufgedeckt sowie aufgegriffen und damit Protesten sowie Einwänden im weiteren Projektverlauf proaktiv begegnet werden [6]. Als dritter größerer Trend der Reformbewegung für mehr Bürgerbeteiligung gelten die Digitalisierungsprozesse und der Wandel der Kommunikationsstrukturen. Durch Open Government, Open Data oder E-Government sollen das Handeln von Regierungen und Verwaltungen für die Bevölkerung ubiquitär abrufbar in Onlineforen transparent dargestellt, das Vertrauen der Bürgerinnen und Bürger in die Institutionen gestärkt und die gesellschaftliche Teilnahme an Entwicklungsplanungen gefördert werden [7, S. 25] [8, S.46]. Durch die intensive Beschäftigung mit verschiedenen Formen der Bürgerbeteiligung in der Praxis und die wissenschaftliche Begleitung sowie Auswertung beispielhafter Partizipationsprojekte in der Forschung entwickelte sich eine ausgeprägte Methodenliteratur bestehend aus Leitfäden, Handlungsrichtlinien und Tool-Baukästen für die bürgerschaftliche Partizipation [9] [10]. Die kommunale Ebene wird dabei durch die räumliche Nähe der Akteure und das vornehmliche Interesse der Bürgerinnen und Bürger an Entwicklungen in ihrem direkten Umfeld als Ausgangspunkt und Ort für Bürgerbeteiligung benannt. Im Rahmen von formellen, gesetzlich geregelten (z.B. Bauleitplanungen) und informellen, für eine Kommune freiwilligen (z.B. Bürgerpanel,

Runde Tische) Partizipationsformen soll sich die Bürgerschaft auch außerhalb von Wahlen Gehör verschaffen können und mit ihrem Engagement Einfluss auf kommunale Projekte und lokale Entwicklungen nehmen [11, S. 104]. Dabei ist zu sehen, dass in der bisherigen Forschung zwar benannt wird, dass im Rahmen von Bürgerbeteiligungsprozessen das Fach- und Expertenwissen politischer Entscheider durch das Erfahrungswissen der Bürgerinnen und Bürger ergänzt, bestätigt oder widerlegt wird (was zu einer besseren Qualität und breiteren Akzeptanz in der Entscheidungsfindung führen soll) [2, S. 79] [3, S.249]. Eine weitere Betrachtung der Prozesse hinter dem Wissenstransfer zwischen den Akteuren in der Bürgerbeteiligung erfolgt allerdings nicht. Es fehlt daher bisher ein Erklärungsansatz oder Modell zur Strukturierung des Wissens auf kommunaler Ebene und dem Tradieren des Wissenstransfers in Partizipationsprozessen für zukünftige, regionale Entwicklungen. Kommunale Bürgerbeteiligungen sind damit heute häufig einzelne, abgeschlossene und themenspezifische Teilhabeprojekte. Die Einrichtung dauerhaft mit der Zivilgesellschaft vernetzter und über die anlass- und projektbezogene Mitwirkung hinausgehender kommunaler Strukturen und Beteiligungen besteht damit zumeist nicht [3, S. 248]. Außerdem zeigen Erfahrungen aus der Praxis, dass bei den Akteuren einer Beteiligungsplanung ein hohes Frustrationspotential besteht, wenn entsprechende Prozesse nicht strukturiert und transparent gestaltet sowie Erwartungen an die Partizipation nicht erfüllt werden [12]. Ein kommunales Wissensmanagement könnte hier für eine wissensorientierte Beteiligungskommune strukturbildend wirken, indem es durch die transparente Darstellung beteiligungsrelevanten Wissens Akteure in Partizipationsprozesse integriert, bei beteiligungswilligen Akteuren entsprechendes Handlungswissen zur Teilhabe ausgebildet, Rahmenbedingungen für einen Wissenstransfer in Bürgerbeteiligungsprozessen setzt sowie das Wissen der lokalen Akteure in den gegebenen Strukturen vernetzt und schließlich auch für regionale Entwicklungen tradiert. Die Kommunikation in einer solchen Beteiligungskommune erfolgt über cross-medial vernetzte Plattformen und Medienkanäle, um möglichst unterschiedliche, aber im Inhalt sowie Darstellung abgestimmte Einstiegs- und Kontaktpunkte für beteiligungswillige Bürgerinnen und Bürger zu bieten. Darüber hinaus ermöglicht die Crossmedialität unterschiedliche Akteure auf verschiedenen Kanälen über beteiligungsrelevante Thematiken zu informieren.

2 Konzeption: Wissensmanagement als Innovationsförderer

Entsprechende Erfahrungen zur Gestaltung eines Wissenstransfers in Organisationen können aus der Praxis und Theorie der Wirtschaft herangezogen werden. Das Ziel des Wissensmanagements ist hier im Rahmen einer wissensorientierten Unternehmensführung „(...) das Gestalten, Lenken, und Entwickeln der organisationalen Wissensbasis zur Erreichung der Unternehmensziele“ [13, S. 161]. Die Ansätze entsprechender Prozesse sind dabei unterschiedlich. Wissensmanagement im engeren Sinne kann mit Informationsmanagement gleichgesetzt werden. Wissensbeauftragte sammeln für Arbeitsprozesse relevante Informationen, stellen diese zusammen und leiten sie an die passiven Nutzer (z.B. Mitarbeiter oder Führungskräfte) weiter. Dieses Verständnis von Wissensmanagement kommt ohne die Betrachtung von Kompetenzen oder Wertgrundlagen in einem Wissenstransfer aus. Wissensmanagement im erweiterten Sinne ist hingegen kompetenzorientiert. Nicht die alleinige Wissensspeicherung, sondern der Wissensfluss steht im Fokus dieses Wissensmanagementsystems. Durch gezielten Kompetenzaufbau bei den Akteuren in einer Organisation soll die Fähigkeit gefördert werden, Problemstellungen selbstorganisiert und kreativ zu lösen [14, S. 7 f.]. Somit ist Wissensmanagement nicht „(...) die Anhäufung von Wissen, sondern die richtige Interpretation vorhandener Fakten, die Entwicklung kompetenter Suchstrategien für relevantes Wissen und die effiziente Umsetzung von Wissen in Nutzen“ [15, S. 34]. Dafür ist auch relevant anhand von Wissenslandkarten zu erfassen, wer in der Organisation was weiß und wer was wissen sollte, um einen Innovationsprozess zu fördern. Hierarchien in Organisationen werden dadurch zu strukturellen Wissens Ebenen, die für einen nachhaltigen Wissenstransfer möglichst transparent und kooperativ gestaltet werden sollten.

Wissensmanagement lässt sich auf dieser Basis in verschiedenen Gebieten anwenden. Besonders drei Phänomenbereiche stehen dabei im Vordergrund [16, S. 46]:

- Individuelles Wissensmanagement
- Organisationales Wissensmanagement
- Regionales Wissensmanagement

Beim individuellen Wissensmanagement geht es um das persönliche Wissen eines Akteurs und wie dieses entwickelt werden kann [17]. Organisationales Wissensmanagement nimmt die Wissensbasis einer Organisation als Grundlage und beschäftigt sich mit Wissensprozessen in Gruppen [18]. Für das regionale Wissensmanagement ist die Nutzbarmachung lokaler Wissensprozesse für eine vorteilhafte Entwicklung einer Region im Fokus. Dabei geht es auch um die Schaffung einer „lernenden Region“ [16, S. 46]. Dies stellt einen Rückgriff auf das Konzept lernender Organisationen dar. Unter solchen Einrichtungen werden „(...) anpassungsfähige, auf äußere und innere Reize reagierende Organisationen verstanden“ [15, S. 115]. Lernende Organisationen sind so durch individuelle und organisationale Lernprozesse geprägt, was eine Erweiterung des Wissensmanagements darstellt. Organisationales Lernen „(...) kann (damit) als Einstiegsprozess in das (...). Wissensmanagement gesehen werden“ [15, S. 116]. Trotz der unterschiedlichen Schwerpunkte, sind die zwei Prozessbereiche nicht voneinander zu trennen. Zielstellung des Wissensmanagements ist die Bewahrung und Erweiterung respektive Strukturierung des Wissens einer Organisation. Die lernende Organisation hinterfragt und entwickelt sich zudem weiter.

Einen Ansatz die beiden Bereiche zusammenzuführen bildet das Münchner Modell von Reinmann-Rothmeier und Mandl. Zielsetzung dieses Wissensmanagement-Ansatzes ist die Entwicklung einer lernenden Organisation [13, S. 173]. Mit der Fokussierung auf das Lernen sollen Organisationen „(...) [verstehen], mit ihren Wissensressourcen umzugehen und ihre Wissensträger im Aufbau wissensrelevanter Kompetenzen und Einstellungen zu unterstützen (...)" [19, S. 9]. Je besser einer Organisation dies gelingt: „(...) umso leichter kann sie auf Veränderungen in ihrem gesellschaftlichen (...) Umfeld reagieren und innovative Prozesse anstoßen“ [19, ebd.]. Damit werden individuelles und organisationales Lernen miteinander verbunden. Um auf beide Lernzyklen zugreifen zu können, schlägt das Münchner Modell vier Wissensprozess-Kernbereiche vor, für deren Darstellung zunächst das grundlegende Verständnis von Wissen relevant ist. Wissen gilt im Münchner Modell zum einen als Objekt (z.B. in Enzyklopädien festgehaltene Erkenntnisse) und zum anderen als Prozess (auf Erfahrung basierend und im menschlichen Kontakt eingebracht). Es lässt sich damit in Informationswissen und Handlungswissen unterteilen. Zwi-

schen diesen beiden Polen ist Wissen stetig in Bewegung und führt von rezipierter Information zum Handeln [19, S. 14 ff.]. Diese Wissensbewegung wird in vier Phänomenbereichen erfasst. In der Wissensrepräsentation wird versucht, Wissen in eine greifbare und bearbeitbare Form zu bringen. Dafür müssen die beteiligten Akteure bereit sein, ihr Wissen nach außen und in eine Organisation zu geben. Bei der Wissensnutzung soll Wissen in beobachtbare Handlungen eingebacht werden. Dafür müssen die beteiligten Akteure Gewohnheiten überwinden und Handlungsfreiräume nutzen. Bei der Wissenskommunikation wird Wissen ausgetauscht und verteilt. Dieser Prozess beruht auf einem Geben und Nehmen. Beteiligte Akteure befinden sich in einer Win-Win-Situation, kommunizieren ihr Wissen und nehmen neues Wissen für sich heraus. In der Wissensgenerierung werden Informationen zu handlungsorientiertem Wissen. Dies ist auf Grundlage eines stetigen Lernens möglich [19, S.22ff.]. Für die Gestaltung der Wissensbewegung haben Communities eine zentrale Bedeutung. Sie bilden innovative Organisationsstrukturen und die Keimzelle des Wissensmanagements. Unter Communities sind selbst organisierende Interessengemeinschaften oder Personennetzwerke zu verstehen, die aus einem einflussreichen Kern und einer lose gekoppelten Peripherie bestehen. Zentrale Prozesse der Communities sind Kommunikation, Kooperation, Erfahrungsaustausch, Wissensschaffung und das wechselseitige Lernen. Als Keimzelle des Wissensmanagements bilden Communities Knotenpunkte für Kommunikation, an denen Informationen und Wissen weitergegeben werden. Wissen und Erfahrungen können so innerhalb einer Organisation oder zwischen Einzelpersonen leichter tradiert werden. Außerdem sind Communities durch ihren informellen und strukturell unabhängigen Charakter Innovationstreiber, indem bereichs- und fachübergreifendes Denken und Problemlösungen gefördert werden [19, S. 28 ff.].

3 Wissenstransfer in der Bürgerbeteiligung gestalten

In der Forschung zu Partizipations- und Bürgerbeteiligungsprozessen entwickelte sich in den vergangenen Jahren das Leitbild der Bürgerkommune, in deren Rahmen sich Bürgerbeteiligung und die Aktivierung zur Partizipation „(...) zu einer übergreifenden Moder-

nisierungsstrategie, die auf die Ausbildung einer neuer Partizipationskultur in der Bürgerkommune zur Optimierung des Gemeinwohls zielt“ verbinden [5, S. 18]. Bürgerkommunen fördern dafür „(...) bürgerschaftliches Engagement (...) zur Erschließung von Informationen und Ressourcen von privaten Akteuren sowie von Akteuren aus dem Dritten Sektor (Vereine, Verbände, Kirchen etc.) für die Bereitstellung öffentlicher Leistungen“ [5, ebd.]. Das Wissen der Bürgerinnen und Bürger einer Stadt wird somit, neben dem offiziellen Budget, zur wichtigen städtischen Ressource [3, S. 250]. Zwischen den Akteuren in Bürgerbeteiligungsprozessen auf kommunaler Ebene besteht somit ein Wissenstransfer, für den es strukturelle und organisatorische Rahmenbedingungen zu definieren gilt. Das Münchner Modell stellt hier einen theoretischen Rahmen für eine entsprechende Modellentwicklung hin zu einer gezielten Wissensnutzung in kommunalen Bürgerbeteiligungsprozessen. Wird das Konzept der Bürgerkommune um entsprechende Prozesse des Wissensmanagements in lernenden Organisationen (respektive lernenden Regionen) ergänzt, so entsteht die Perspektive von einer wissensorientierten Beteiligungskommune. In dieser koordiniert ein Partizipationsbeauftragter das Wissensmanagement in den Beteiligungsprozessen, findet eine Analyse der in einer Region bestehenden Wissensbasis statt und werden regionale Wissensziele (z.B. Zukunftsleitfäden) definiert. Außerdem werden Informationen transparent dargestellt und neues Wissen auf geeigneten Kanälen (z.B. Dokumentation in RegionalWikis, Protokolle / Abschlussberichte von Arbeitskreisen) crossmedial vernetzt für zukünftige Entwicklungen tradiert. Weiterhin bilden zivilgesellschaftliche Communities wie Vereine, Bürgerinitiativen oder soziale Verbände Knotenpunkte der Kommunikation und eine Plattform für Bürgerbeteiligung. Zum einen kommunizieren die Communities Bedarfe und Erfahrungen aus ihrem Engagement nach außen. Zum anderen greifen sie Informationen auf und leiten sie an ihre Mitglieder weiter. Communities sind dabei Orte individueller und organisationaler Lernprozesse in einer Beteiligungskommune. Neben dem Tradieren des kommunalen Wissens aus den verschiedenen Akteursfeldern wird hier auch Handlungswissen bei den lokalen Akteuren ausgebildet, indem Kompetenzen für die Teilnahme an Bürgerbeteiligungsprozessen geschaffen werden (unter anderem Demokratie- und Bürgerkompetenzen [20]). Bürgerbeteiligung löst sich damit von der Konnotation eine Teilnahme an politischen Entschei-

dungsprozessen zu sein und wird zu einer dauerhaften gesellschaftlichen Aufgabe der Akteure einer Region.

4 Ausblick: Empirische Untersuchungen in lokalen Netzwerken

Die vorangegangenen Ausführungen stellen zunächst eine theoriegeleitete Annäherung an den Wissenstransfer in Bürgerbeteiligungsprozessen dar. Für die Entwicklung des Konzeptes der wissensorientierten Beteiligungskommune, in der das Wissen der Akteure in Partizipationsprozessen für die kommunale Entwicklung nutzbar gemacht wird, ein Wissensmanagement das Tradieren dieses Wissens begleitet und das Lernen der Akteure durch das Ausbilden beteiligungsrelevanter Fähigkeiten im Fokus steht, bedarf es einer weiteren empirischen Untersuchung. Eine solche steht im Zentrum eines kooperativen Promotionsprojektes an der Hochschule Mittweida und der Universität Bremen. Als Untersuchungsraum für die qualitative Forschung gilt die durch mehrere Kleinstädte und eine nahegelegene Großstadt geprägte Region Mittweida. In dem Testfeld werden Leitfadenterviews mit lokalen Akteuren aus Politik, Wirtschaft, Wissenschaft, Lokalmedien, Vereinsarbeit und Bürgerschaft geführt (derzeit 20 Interviews/ ø 48 Minuten Länge). Nach Ende der Befragung und Transkription wird das qualitativ erhobene Datenmaterial im Rahmen einer inhaltlich-strukturierenden Inhaltsanalyse ausgewertet. Im Fokus der nachgelagerten Forschungsleistung steht die deduktive und induktive Kategorienbildung am Datenmaterial zu den Themenfeldern Bürgerbeteiligung, lokale Netzwerke, kommunale Kommunikationsstrukturen und Gestaltung eines Wissensmanagements. Damit sollen Aussagen darüber getroffen werden, welche Akteure am Wissenstransfer in Beteiligungsverfahren teilnehmen, an welchen Orten (Medien und Einrichtungen) der entsprechende Austausch stattfindet sowie welche Rahmenbedingungen die Wissenstradierung begünstigen oder auch konterkarieren.

Danksagung

Das Promotionsprojekt wird aus Mitteln des ESF und des Freistaates Sachsen gefördert. Der Autor bedankt sich für die Betreuung bei Prof. Dr. Christian Pentzold vom ZeMKI an der Universität Bremen und Prof. Dr. Tamara Huhle von der Fakultät Medien der Hochschule Mittweida.



Europäische Union

Europa fördert Sachsen.



Europäischer Sozialfonds



Diese Maßnahme wird mitfinanziert durch Steuermittel auf Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Literatur

1. R. Hebestreit, Ed., *Partizipation in der Wissensgesellschaft. Funktion und Bedeutung diskursiver Beteiligungsverfahren*, Wiesbaden, 2013.
2. O. Renn, "Partizipation bei öffentlichen Planungen. Möglichkeiten, Grenzen, Reformbedarf," in *Zivile Bürgergesellschaft und Demokratie. Aktuelle Ergebnisse der empirischen Politikforschung*, Wiesbaden, 2013.
3. J. Franzke, "Engagierte Bürger und lokale Macht. Verändert Partizipation städtische Herrschaft?" in *Bürgerschaft und demokratische Regierbarkeit in Deutschland und Europa*, Wroclawski, 2014.
4. J. Bogumil and L. Holtkamp, "Die kommunale Ebene," in *Engagementpolitik. Die Entwicklung der Zivilgesellschaft als politische Aufgabe*, Wiesbaden, 2010.
5. H. Bauer and L. Hajasch, "Vom passiven Untertan über den Wutbürger zum aktiven Citoyen in der Bürgerkommune," in *Partizipation in der Bürgerkommune*, Potsdam, 2017.
6. K. Selle, "Stadtentwicklung und Bürgerbeteiligung – auf dem Weg zu einer kommunikativen Planungskultur?" *Information zur Raumentwicklung*, vol. Heft 1.2007, pp. 63–71, 2007.
7. B. W. H. Kubicek, Herbert/ Lippa, Ed., *Medienmix in der Bürgerbeteiligung. Die Integration von Online-Elementen in Beteiligungsverfahren auf lokaler Ebene*, Berlin, 2009.
8. Staatsministerium für Wirtschaft, Arbeit und Verkehr, Ed., *Sachsen digital 2017 – Digitalisierungsstrategie des Freistaates Sachsen*. Dresden: <https://publikationen.sachsen.de/bdb/artikel/28672>, 2017.
9. J. Sommer, Ed., *Kursbuch Bürgerbeteiligung*, Potsdam, 2015.

10. C. Beringhaus, G. Wachinger, and O. Renn, Eds., *Bürgerbeteiligung. Konzepte und Lösungswege für die Praxis*, Frankfurt am Main, 2016.
11. S. Bock, "Erfolgsbedingungen kommunaler Bürgerbeteiligung: Perspektiven, Chancen und Fallstricke in der Praxis," in *Partizipation in der Bürgerkommune*, Potsdam, 2017.
12. J. Ermert, "Das Ebenenproblem zwischen Bürgern und Planern in partizipativen Stadtentwicklungsprozessen. Der Vermittlungsversuch eines Praktikers," *pnd Online*, vol. Online I, no. 2018, 2018.
13. K. North, Ed., *Wissensorientierte Unternehmensführung. Wissensmanagement gestalten*, Wiesbaden, 2016.
14. W. Sauter and C. Scholz, Eds., *Kompetenzorientiertes Wissensmanagement. Ge-steigerte Performance mit dem Erfahrungswissen aller Mitarbeiter*, Wiesbaden, 2015.
15. S. Porschen, Ed., *Austausch impliziten Erfahrungswissen. Neue Perspektiven für das Wissensmanagement*, Wiesbaden, 2008.
16. R. Gutounig, Ed., *Wissen in digitalen Netzwerken. Potentiale neuer Medien für Wissensprozesse*, Wiesbaden, 2015.
17. G. Reinmann, Ed., *Individuelles Wissensmanagement – ein Rahmenkonzept für den Umgang mit personalem und öffentlichem Wissen (Arbeitsbericht Nr. 5)*, Augsburg, 5. März 2005.
18. K. Zboralski, Ed., *Wissensmanagement durch Communities of Practice. Eine empirische Untersuchung von Wissensnetzwerken*, Wiesbaden, 2007.
19. G. Reinmann-Rothmeier, Ed., *Wissen managen: Das Münchener Modell (Forschungsbericht Nr. 131)*, München, 2001.
20. M. May, Ed., *Demokratiefähigkeit und Bürgerkompetenzen. Kompetenztheoretische und normative Grundlagen der politischen Bildung*, Wiesbaden, 2007.

