Elderly Care

Using Deep Learning for Multi-Domain Activity Classification

Li , Shaoxuan ; Jia, Mu; Le Kernec, Julien; Yang, Shufan; Fioranelli, Francesco; Romain, Olivier

DOI
10.1109/UCET51115.2020.9205464

Publication date
2020

Document Version
Final published version

Published in
2020 International Conference on UK-China Emerging Technologies (UCET)

Elderly Care: Using Deep Learning for Multi-Domain Activity Classification

Shaoxuan Li
James Watt School of Engineering
University of Glasgow
Glasgow, UK
2288945L@student.gla.ac.uk

Mu Jia
James Watt School of Engineering
University of Glasgow
Glasgow, UK
2288929J@student.gla.ac.uk

Julien Le Kernec
James Watt School of Engineering
University of Glasgow
Glasgow, UK
julien.lekernec@glasgow.ac.uk

Shufan Yang
James Watt School of Engineering
University of Glasgow
Glasgow, UK
shufan.yang@glasgow.ac.uk

Francesco Fioranelli
MS3-Microwave Sensing Signals and Systems, TU Delft, Delft, The Netherlands
f.fioranelli@tudelft.nl

Olivier Romain
ETIS – Signal and Information Processing lab, University Cergy-Pontoise, Cergy, France
olivier.romain@cyu.fr

***Abstract*——Nowadays, health monitoring issues are increasing as the worldwide population is aging. In this paper, the radar modality is used to classify with radar signature automatically. The classic approach is to extract features from micro-Doppler signatures for classification. This data representation domain has its limitations for activities presenting similar accelerations like a frontal fall and picking up an object from the floor that lead to wrongly labeled activities. In this work, we propose to combine multiple radar data domains with deep learning. Features are extracted from four domains, namely, Range-Time, Range-Doppler, Doppler-Time, and Cadence Velocity Diagram. The extracted features are set as the input of a Convolutional Neural Network, yielding 91% accuracy with 10-fold cross-validation based on the University of Glasgow "Radar signatures of human activities" open dataset.**

Keywords—Machine Learning, Radar, Assisted Living, Human Activity Recognition, Multi-domain

I. INTRODUCTION

In the 21st century, people aged 60 and over have reached 900 million in 2015, and it is predicted there will be 2 billion in 2050 all over the world [1] increasing the number of chronic diseases dramatically to monitor away from primary care units [2]. Human activity classification is vital for assisted living for fall detection as well as monitoring the activity levels, which are highly correlated with health status. Different sensing modalities exist in the context of assisted living, such as cameras and wearable sensors. Cameras are very efficient thanks to advances in computer vision, but their performances are sensitive to lighting conditions. Furthermore, there are rising concerns about the legal, data ownership, and privacy aspects of end-users. Wearable devices are also common for elderly care. The accuracy is high, but the end-users have to wear them for them to work and may require challenging maintenance for cognitively impaired patients for battery charging or handling [3, 4].

In comparison, the radar sensing modality does not capture optical images of the indoor environment or end-users. Therefore, it protects the intimacy of the users even in case of data being leaked since there is no identifiable data, and it requires expert knowledge to process and interpret the data. Furthermore, radar systems are low-power embedded systems for indoor environments, which means they can be packaged discreetly to avoid stigmatization [5].

Classic approaches in radar-based activity monitoring consist of extracting features directly from micro-Doppler signatures [6]. However, when considering activities presenting similar accelerations such as falling forward and picking up an object from the floor, walking and walking while carrying an object with two hands, the spectrograms might show the similar patterns which is not friendly to image processing [3].

In [7, 8], the authors used deep learning approaches with the help of spectrogram only for activities classification. Algorithms have difficulties finding effective boundaries to separate those classes effectively, especially when it comes to Convolutional Neural Network (CNN) because they did not exploit different radar data domains. In this work, we propose to exploit multiple radar data domains. The radar domains are not only Spectrogram (Doppler-time (DT)), used in [7], but also Range-time (RT), Range-Doppler (RD), and Cadence Velocity Diagram (CVD) to improve classification accuracy. This is compared to the performance of a convolutional neural network applied to the micro-Doppler signatures (DT domain) to evaluate the gain in performance as well as classic statistical learning.

The remainder of this paper is organized as follows. Section II introduces the data collection and the experimental setup. The pre-processing for the four data domains is described in Section II. Section III describes the methodology for feature extraction and classification. Section IV presents the results and conclusions are given in Section V.

II. DATA COLLECTION AND AND PRE-PROCESSING

The data were collected at the University of Glasgow and in two care homes using a Frequency-Modulated Continuous Wave (FMCW) radar. It operates at 5.8 GHz and has an instantaneous bandwidth of 400 MHz with 1 kHz pulse repetition frequency. These six activities are shown with their corresponding labels in Table I. The data comprises 72 participants with over 1000 data files [9, 10].

TABLE I. SIX ACTIVITIES. THESE SIX ACTIVITIES ARE CLASSIFICATION OBJECTIVES IN THIS PROJECT.

Activities	Label	Activities	Label
Walking	1	Picking up an object	4
Sitting down	2	Drinking	5
Standing up	3	Falling forward	6

The receiver will receive the backscattered signals from the target. The transmitted signal is mixed with the received signal giving the beat frequency [11]. The beat frequency is directly proportional to the distance to the target and is also offset by the target Doppler. The distance is recovered by applying a simple Fast Fourier Transform (FFT) on the length

of a sweep [12, 13]. The beat frequency is sampled at 128 kHz, meaning 128 samples/sweep. A more detailed description of the dechirping process can be found in [14].

There are four domains in total, which are RT, RD, DT, and CVD.

The RT domain is extracted from the raw I&Q data by applying a 128-point FFT. This gives a range profile for every pulse. The clutter is suppressed using a single delay-line Moving Target Indicator filter.

$$\Delta R = \frac{c}{2B} = 37.5cm \qquad (1)$$

Where c is the speed of light, B is the instantaneous bandwidth.

Before processing the radar data, an MTI (Moving Target Indicator) filter is applied on the radar data. Then, a Fast Fourier Transform is used to identify the range bins containing the subject's signature by looking at the range-Doppler maps.

The RD domain is obtained by applying a 10000-point FFT for every range bin in the time direction. RD diagram shows the Doppler response of different frequency components at different ranges.

The DT domain is obtained from the RT domain by summing the STFT (Short-Time Fourier Transform)results of each range bins together which does FFT on the time axis of each range bin, using a sliding window moving by Δt. The result is commonly known as micro-Doppler signature or spectrogram representing the evolution of Doppler as a function of time at the cost of resolution because of the smaller window size. However, it allows capturing the motions of moving parts of a structure [15-17]. The overlap factor is 0.95, and the window size is 0.2s, giving a 200-point FFT zero-padded to 800. The Doppler Resolution is 2.5Hz per bin, which is calculated by sweep time and FFT point which are $1ms$ and 800 respectively.

CVD is obtained by further processing the spectrogram (Doppler-time) with a 981-point FFT for every Doppler bin in the time direction. CVD is well-suited to extract periodic motions.

III. Methodology for Feature Extraction

A. Handcrafted-Feature extraction

In this study, 36 features are considered from four domains, such as energy, entropy, skewness, centroid, bandwidth. The full listing can be found in Table II rearranged as a $6x6$ matrix and set as the input of a CNN.

B. Convolutional Neural Network (CNN)

CNN is a kind of deep learning. It utilizes layers with convolution filters that are applied to given features, which is the input [18]. The CNN network is operating in 2D. This CNN was programmed with python using the programming environment details, as shown in Table III.

The CNN (Fig. 1) has 5 layers with two dropout layers. There are 2 convolutional layers with 32 and 64 convolutional 3×3 filters respectively, a max-pooling layer, a fully-connected layer, and a softmax output layer. The first dropout layer is between convolution layer and pooling layer with 0.5 dropout rate. It is used to eliminate redundant features extracted by the convolutional layer. The second one is between pooling layer and fully-connected layer whose function is to prevent overfitting phenomenon. The batch size is 128, and the categorical cross-entropy loss function is applied in this architecture while the learning rate is adaDelta which is a self-adaptive learning rate. The function of categorical cross-entropy loss [19] is shown in Equation 4.

$$C = -\frac{1}{n}\sum_x [y \ln a + (1-y)\ln(1-a)] \qquad (4)$$

Where y is a label or expected output, x is the number of samples and a is the actual output.

TABLE II. The Matrix Formed by 36 Features from 4 Domains: Doppler-time (Green), RT (Blue), RD (Yellow), and CVD (Gray).

Matrix Formed by Features					
Entropy	Skewness	Centroid (mean)	Centroid (var)	Bandwidth (mean)	Bandwidth (var)
SVD mean of 1st col of U	SVD mean of 1st col of V	SVD mean of 2nd col of U	SVD mean of 2nd col of V	SVD mean of 3rd col of U	SVD mean of 3rd col of V
SVD var of 1st col of U	SVD var of 1st col of V	SVD var of 2nd col of U	SVD var of 2nd col of V	SVD var of 3rd col of U	SVD var of 3rd col of V
Energy Curve-mean	Energy Curve-variance	Energy Curve-integral	SVD var of 1st col of U	SVD var of 2nd col of U	SVD var of 3rd col of U
SVD var of 2nd col of V	SVD var of 3rd col of V	SVD var of 1st col of U	SVD var of 2nd col of U	SVD var of 3rd col of U	SVD var of 2nd col of V
SVD var of 3rd col of V	Entropy	Energy max	Energy min	Energy difference	Max Velocity Frequency

*col: column, var: variance

TABLE III. Version of Some Essential Packages or Libraries.

Package	Version
Anaconda	3
Tensorflow	1.6.0
CUDA	v9.0
cuDNN	v7
Keras	2.1.5

*CUDA: Computer Unified Device Architecture

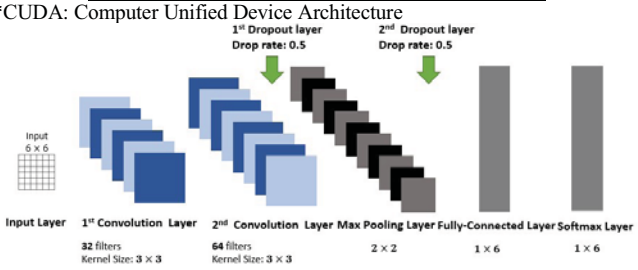

Fig. 1. The structure of CNN to process the feature matrix

IV. Classification Methods

In this paper, four methods are introduced. Method 1 (M1) is to input the matrix formed by features which is a regular method. From 36 features in Table II, a matrix is formed and set as the input of classifier, CNN. In method 2 (M2), features are extracted from cropped radar domains from the regions of interest as described in Fig. 2. This is to dismiss either redundant information or regions containing only noise, which are detrimental to the accuracy. The original and cropped domains are shown in Fig 2. Method 3 and 4 (M3 & M4) are two control groups for comparison. M3 uses images as the input of the classifier resized to 45×45. M4 uses a statistical machine learning approach, namely, Support Vector Machine (quadratic SVM) instead of CNN. The normalized features from cropped radar data domains are set to be the input of SVM. To make the results more accurate, the SBS (Sequential Backward Selection) is used in this paper to remove redundant features. All methods use 10-fold cross-validation, and the accuracy is the average of the 10 runs.

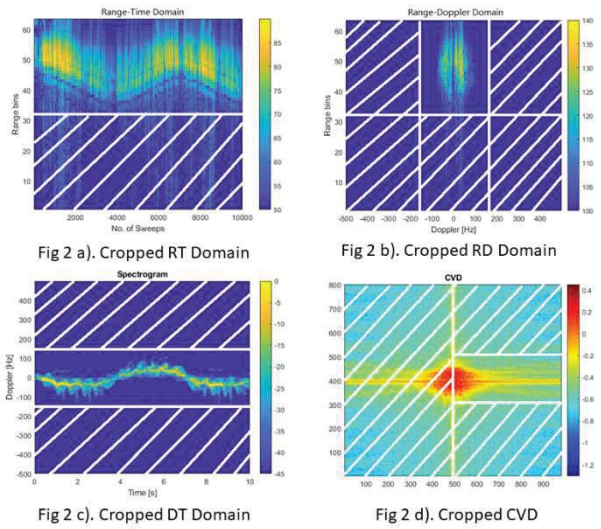

Fig. 2. a). - d). Original and Cropped RT & RD Domains. These figures represent the comparison of cropped radar data and original radar data.

V. Results

Fig. 4 to 6 show the test accuracy and loss function for methods 1 to 3, respectively whereas the accuracy per activity is presented for all 4 methods using confusion matrices in Fig. 6 to 9. The objective is to classify 6 activities, and the overall performances are summarised in Table IV, and the result is the accuracy of the classifier. As shown in Fig. 3 and 6 and Table IV, the test accuracy of M1 is 89.72%, and test loss is 0.331. This model took 1 min 52s to finish 1500 epochs training with 10-fold cross-validation. For M2, Fig. 4 and 7 show that the test accuracy increases by ~2%, reaching 91.59%, which proves that this operation (cropping) is helpful for the final result. The errors occur for the classification of activity 4 (picking up an object) with an accuracy of 76%. The test accuracies of other activities are much higher than it of activity 4, at 90% (drinking) to 100% (falling).

In control groups, Fig. 5 and 8. shows the results of M3. This mode achieved 97.20% accuracy within 1min 25s using 100 epochs. M4 took 2.734s to achieve 92.7% test accuracy, and the confusion matrix is shown in Fig. 9. Fig. 10 shows the results of SBS. The accuracy is 92.70% without SBS, and the accuracy increased by 2% with the help of SBS, which used 10 features.

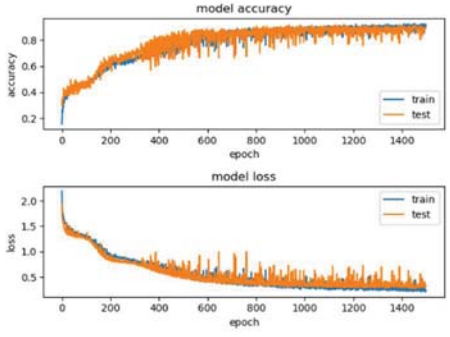

Fig. 3. The accuracy top) and loss bottom) of the CNN for method 1. The testing accuracy is ended at 89.72%, while the test loss is 0.331 at epoch 1500.

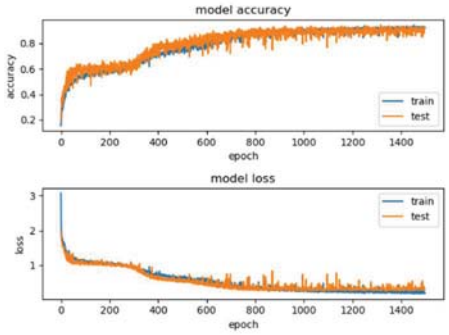

Fig. 4. The accuracy top) and loss bottom) of the method 2 whose features are extracted from the cropped area. The test accuracy is 91.59%, and test loss is 0.308 at epoch 1500.

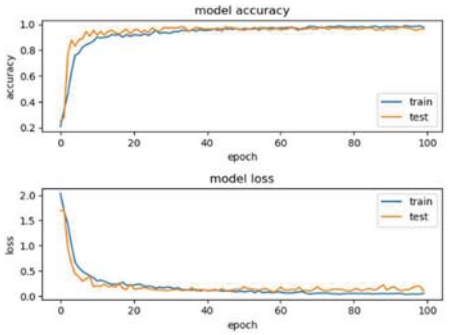

Fig. 5. The accuracy top) and loss bottom) of the CNN approach when the input is Spectrogram Image. The test accuracy is 91.59%, and test loss is 0.308 at epoch 1500.

Overall 89.72%		Predicted label					
		1	2	3	4	5	6
True label	1	94%	0	0	6%	0	0
	2	0	100%	0	0	0	0
	3	0	0	88%	0	12%	0
	4	0	12%	0	71%	18%	0
	5	0	0	0	10%	90%	0
	6	6%	0	0	0	0	94%

Fig. 6. Confusion Matrix for method 1 (original). The accuracy of activity 4 is lowest at 71% while the accuracy of activity 2 is highest at 100%.

Overall 91.59%		Predicted Label					
		1	2	3	4	5	6
True tabel	1	94%	0	0	6%	0	0
	2	0	94%	0	0	6%	0
	3	0	0	94%	0	6%	0
	4	0	6%	0	76%	18%	0
	5	0	0	0	10%	90%	0
	6	0	0	0	0	0	100%

Fig. 7. Confusion Matrix for method 2 (cropped). The result of activity 4 is the lowest at 76% while the accuracy for other activities is >90% and 100% for activity 6.

Overall 97.20%		Predicted Label					
		1	2	3	4	5	6
True tabel	1	100%	0	0	0	0	0
	2	0	94%	0	0	6%	0
	3	0	0	100%	0	0	0
	4	0	0	0	100%	0	0
	5	0	0	0	14%	86%	0
	6	0	0	0	0	0	100%

Fig. 8. Confusion Matrix When the Input is Spectrogram Image. The test accuracy of activity 5 is lowest, at 86% while the test accuracy of activity 2 is 94%. The accuracy of other activities are 100%.

Overall 92.70%		Predicted Label					
		1	2	3	4	5	6
True tabel	1	100%	0	0	0	0	0
	2	0.5%	96%	0	0.5%	2%	0
	3	0	0	92%	4%	4%	0
	4	0	1%	5%	82%	12%	0
	5	0	0.5%	3%	9.5%	87%	0
	6	0	0	1%	0.5%	0.5%	98%

Fig. 9. Confusion Matrix of SVM. The lowest accuracy occurs at the classification of activity 4, which is 82% and 12% of activity 4 samples are classified as activity 5.

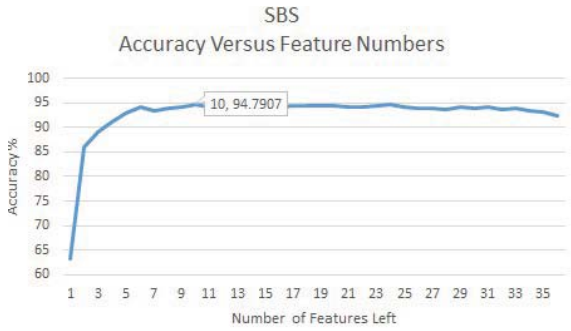

Fig. 10. SBS (Sequential Backward Selection) feature selection result. When there are 36 features which is original condition, the accuracy is 92.70%. When SBS is applied, the accuracy achieved 94.79% with 10 features.

TABLE IV. TRAINING DETAILS AND RESULTS OF FOUR METHODS.

	Deep Learning			Statistical Learning
	original	cropped	image	SVM-SBS
Time	1 min 52s	1min 49s	1min 25s	0.75s
Test Accuracy	89.72%	91.59%	97.20%	94.79%
Cross-validation	10	10	10	10
Epoch	1,500	1,500	100	-
Test Loss	0.331	0.308	0.136	-

VI. CONCLUSIONS

In this paper, 36 features are successfully extracted from multiple radar data domains, rearranged into a matrix and then classified with CNN. The features are extracted from cropped domains in order to improve accuracy by focusing on regions of interest and discard redundant information or noise. Although the cropping improved accuracy by ~2% compared to uncropped, the accuracy of SVM and simple image processing from CNN has finished the classification of the 6 activities using the control methods 3 and 4 still outperform method 2 by 5.61% and 3.2%. The two activities 4 and 5 are the most difficult to distinguish from each other. In our case, we see that the classic image processing (M3) and statistical learning (M4) outperform our designed method. This means that CNN is not adequate to work on finding a pattern in features directly as it does on signals or images. Furthermore, more salient features will need to be designed to discriminate activities 4 and 5 effectively. Furthermore, even though deep learning outperforms SVM by 2.31% in accuracy training a CNN takes much longer (~1mn) than finding statistical learning boundaries (<1s).

REFERENCES

[1] WHO. "Ageing and health." https://www.who.int/news-room/fact-sheets/detail/ageing-and-health#:~:text=Key%20facts,%2D%20and%20middle%2Dincome%20countries. (accessed 2020).

[2] WHO, "Noncommunicable Diseases (NCD) Country Profiles, 2018 - China," 2018.

[3] J. Le Kernec *et al.*, "Radar Signal Processing for Sensing in Assisted Living: The Challenges Associated With Real-Time Implementation of Emerging Algorithms," *IEEE Signal Processing Magazine,* vol. 36, no. 4, pp. 29-41, 2019, doi: 10.1109/msp.2019.2903715.

[4] J. Le Kernec, F. Fioranelli, S. Yang, J. Lorandel, and O. Romain, "Radar for Assisted Living in the Context of Internet of Things for Health and Beyond," in *26th IFIP/IEEE Int. Conf. Very Large Scale Integr. VLSI-SOC*, Verona, 8-10 Oct 2018, pp. 1-5.

[5] J. L. Kernec *et al.*, "Radar Sensing in Assisted Living: an Overview," in *2019 IEEE MTT-S International Microwave Biomedical Conference (IMBioC)*, 6-8 May 2019 2019, vol. 1, pp. 1-4, doi: 10.1109/IMBIOC.2019.8777748.

[6] *Radar for indoor monitoring : detection, classification, and assessment*. Boca Raton: CRC Press, Taylor & Francis Group, 2018, pp. xiii, 390 pages.

[7] Y. Lin, J. Le Kernec, S. Yang, F. Fioranelli, O. Romain, and Z. Zhao, "Human Activity Classification With Radar: Optimization and Noise Robustness With Iterative Convolutional Neural Networks Followed With Random Forests," *IEEE Sensors Journal,* vol. 18, no. 23, pp. 9669-9681, 2018, doi: 10.1109/jsen.2018.2872849.

[8] Y. Kim and B. Toomajian, "Hand Gesture Recognition Using Micro-Doppler Signatures With Convolutional Neural Network," *IEEE Access,* vol. 4, pp. 7125-7130, 2016, doi: 10.1109/access.2016.2617282.

[9] D. F. Fioranelli, D. S. A. Shah, H. Li, A. Shrestha, D. S. Yang, and D. J. L. Kernec, "Radar sensing for healthcare," *Electronics Letters,* vol. 55, no. 19, pp. 1022-1024, 2019, doi: 10.1049/el.2019.2378.

[10] F. Fioranelli, S. A. Shah, H. Li, A. Shrestha, S. Yang, and J. Le Kernec. *Radar signatures of human activities*, doi: 10.5525/gla.researchdata.848.

[11] A. Melnikov, J. L. Kernec, and D. Gray, "A case implementation of a spotlight range migration algorithm on FPGA platform," in *2014 International Symposium on Antennas and Propagation Conference Proceedings*, 2-5 Dec. 2014 2014, pp. 177-178, doi: 10.1109/ISANP.2014.7026588.

[12] A. Melnikov, J. L. Kernec, and D. Gray, "Porting spotlight range migration algorithm processor from Matlab to Virtex 6," in *2015 IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC)*, 7-11 Sept. 2015 2015, pp. 1429-1432, doi: 10.1109/APWC.2015.7300214.

[13] A. Melnikov, J. L. Kernec, and D. Gray, "FMCW rail-mounted SAR: Porting spotlight SAR imaging from MATLAB to FPGA," in *2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 5-8 Aug. 2014 2014, pp. 780-785, doi: 10.1109/ICSPCC.2014.6986303.

[14] G. W. Stimson, *Introduction to Airborne Radar*. Institution of Engineering and Technology, 1998.

[15] J. Le Kernec and O. Romain, "Performances of multitones for ultra-wideband software-defined radar," *IEEE Access,* vol. 5, pp. 1-1, 2017, doi: 10.1109/access.2017.2693300.

[16] J. L. Kernec and O. Romain, "Multitones' Performance for Ultra Wideband Software Defined Radar," in *Applications of Digital Signal Processing through Practical Approach*, S. Radhakrishnan Ed.: Intech open, 2015, ch. 3.

[17] J. L. Kernec, D. Gray, and O. Romain, "Empirical analysis of chirp and multitones performances with a UWB software defined radar: Range, distance and Doppler," in *Proceedings of 2014 3rd Asia-Pacific Conference on Antennas and Propagation*, 26-29 July 2014 2014, pp. 1061-1064, doi: 10.1109/APCAP.2014.6992691.

[18] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, "Personnel Recognition and Gait Classification Based on Multistatic Micro-Doppler Signatures Using Deep Convolutional Neural Networks," *IEEE Geoscience and Remote Sensing Letters,* vol. 15, no. 5, pp. 669-673, 2018, doi: 10.1109/lgrs.2018.2806940.

[19] S. Mannor, D. Peleg, and R. Rubinstein, "The cross entropy method for classification," presented at the Proceedings of the 22nd international conference on Machine learning, Bonn, Germany, 2005. [Online]. Available: https://doi.org/10.1145/1102351.1102422.