2015 IEEE Symposium Series on Computational Intelligence

Ensembles of Support Vector Data Description for Active Learning based Annotation of Affective Corpora

Patrick Thiam
Institute of Neural Information Processing
Ulm University
Germany
Email: patrick.thiam@uni-ulm.de

Markus Kächele
Institute of Neural Information Processing
Ulm University
Germany
Email: markus.kaechele@uni-ulm.de

Günther Palm
Institute of Neural Information Processing
Ulm University
Germany
Email: guenther.palm@uni-ulm.de

Friedhelm Schwenker
Institute of Neural Information Processing
Ulm University
Germany
Email: friedhelm.schwenker@uni-ulm.de

***Abstract*—The present work aims primary at developing an approach to detect irregular and spontaneous facial gestures in video sequences. The developed approach should help a system distinguish between neutral facial expressions characterized by the absence of facial gestures and facial events characterized by the presence of observable facial gestures in video sequences. For this purpose, an active learning approach is proposed in order to avoid the task of annotating an entire video sequence before proceeding with the classification. It is well known that the annotation task is hard, expensive and error prone. Each video sequence is segmented into smaller segments that are then to be investigated and annotated based on the absence or presence of facial gestures. The approach consists in first selecting a set of samples classified as uncharacteristic through the majority vote of a committee of support vector data description (SVDD) models generated randomly. The base learner then focuses on the selected outliers and not on the whole annotated corpus. Different query strategies are used to select the most informative samples among the selected outliers. Those samples are then annotated by the user and added to a pool of annotated samples. The latter is subsequently used to train the base learner again before the next iteration can take place. Experiments suggest that the proposed active learning approach performs as well as a system trained on a fully annotated corpus, while dramatically reducing the cost of annotation.**

I. Introduction

Semi-supervised learning is a learning paradigm based on the combination of unsupervised and supervised learning mechanisms. This field of research bears its means of existence from the fact that the globally available amount of data increases so fast, that it is almost impossible to properly assess the content of the data, let alone annotate or label it in a meaningful way. The demand for algorithms that are able to process huge amounts of unlabelled data together with a small set of labelled instances has gained, especially in recent years, increased attention with the emergence of fields such as web document classification, mining of social data and applications that generally fall under the term *big data* [1]. With increasing technologization of everyday life, the human factor starts to play a significant role in pattern recognition and machine learning applications. Data analysis in the context of social signals has become an emerging field with many researchers contributing to the subfields of emotion recognition, human-computer interaction, language processing, and recognition of gestures and facial expressions.

Those subfields are highly data driven and advancements are generally strongly correlated with having more or better data collections. Designing and recording such data collections is an interdisciplinary task that demands a large amount of effort. The concept phase of a recording includes the design of the actual experiment and the emotional stimulation, as well as the recruitment of participants. For the actual recording phase the hardware setup has to be prepared and the emotional stimulation has to be carried out.

While those points may use the majority of the scheduled time, an important point that has to be considered is the annotation of the recorded material. In the annotation process groundtruth labels are generated for subsequent classification experiments. As the annotation process causes high cognitive load in the annotators, it usually takes more time to annotate a single video than its duration. If more than one emotional dimension are to be annotated, the video has to be watched several times. Additionally to obtain reliable labels, more than one rater is used as the outcome is highly dependent on the individual. Those issues render annotation a very expensive task. To tackle the annotational overhead, machine learning methods from the fields of semi-supervised and active learning can be used. Active learning [2], [3] is a method in which a classifier selects which instances should be labeled by an expert. The decision should be made based on the information gain that is achieved when selecting specific points. Active learning has for example been successfully applied in combination with crowd sourcing [4]. Bandla et al. [5] used active learning with uncertainty sampling for action detection in video sequences. Tong and Chang [6] applied active learning with a binary Support Vector Machines (SVM) [7] as base learner in the domain of image retrieval using relevance feedback.

Recently, active learning has been used for the emotional annotation of the Interaction Game corpus [8] using support vector methods [9]. The approach proposed in the latter aimed at detecting events in video sequences by using an One Class

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.251

Support Vector Machine (OCSVM) [10] as a base learner in an active learning scenario. The use of an OCSVM is motivated by the assumption that most of the outliers (events) are located in regions of low density in comparison to neutral instances which are located in regions of high density. Therefore using the distance from each sample to the OCSVM decision boundary as a confidence measure would help in detecting and annotating as much outliers as possible. The presented experimental results proved the effectiveness of the approach. Therefore, we build upon this work and investigate the use of ensemble methods for the detection and annotation of outliers. Furthermore, we investigate the combination of the described outliers detection approach with different selection criteria that are used in the active learning decision making process for the development of a system that accurately distinguishes between neutral instances and events (see Figure 1) in video sequences while reducing the cost of annotation.

The remainder of this work is organized as follows. In the next section, the theoretical basis of our system is introduced together with the proposed decision criteria. The data collection that was used is described in Section 3. In Section 4, the experimental results are presented before in Section 5 a conclusion is drawn.

II. Theoretical Background

In this section, the theoretical basis is introduced which includes SVDD, active learning and the proposed decision criteria.

A. Support Vector Data Description (SVDD)

Introduced by Tax and Duin [11], Support Vector Data Description (SVDD) is inspired from the Support Vector classifier and aims to characterize a set of objects in order to distinguish the latter from atypical and uncharacteristic objects. Thus SVDD can be used for outlier detection as well as for imbalanced classification problems [12], [13]. To this end, SVDD generates a closed spherically shaped boundary around the set of target objects with minimal volume. Objects located out of the boundary are therefore considered as outliers while those located inside the boundary are considered as normal or typical objects.

Let $\{x_i\}_{i=1}^{N}$, with $x_i \in \mathbf{R}^n \; \forall i \in \{1, \ldots, N\}$ and $(N, n) \in \mathbf{N}^* \times \mathbf{N}^*$ be a set of objects belonging to a target class. The SVDD generates a closed boundary called hypersphere with a radius $R > 0$ and a center $a \in \mathbf{R}^n$ with minimal volume around the target class. The volume of the hypersphere is minimized by minimizing R^2 and all points x_i of the target class should be within its boundary. Thus, the optimization problem to be solved is defined as follows:

$$\begin{aligned} &\textit{Minimizing} \quad F(R, a) = R^2 \\ &\textit{s.t.} \quad \|x_i - a\|^2 \leq R^2, \forall i \end{aligned} \tag{1}$$

Furthermore, the distance from an object x_i to the center of the hypersphere a should not be strictly smaller than R^2 but larger distances should be penalized. These requirements enhance the flexibility of the boundary. To this end, slack variables $\xi_i \geq 0$ are introduced, as well as a parameter $C \geq 0$, which controls the trade-off between the volume of the hypersphere and the amount of miss classifications. The optimization problem becomes:

$$\begin{aligned} &\textit{Minimizing} \quad F(R, a) = R^2 + C \sum_i \xi_i \\ &\textit{s.t.} \quad \|x_i - a\|^2 \leq R^2 + \xi_i, \; \xi_i \geq 0 \quad \forall i \end{aligned} \tag{2}$$

The optimization problem is solved by introducing the Lagrange multipliers $\alpha_i \geq 0$ and $\gamma_i \geq 0$:

$$\begin{aligned} L(R, a, \alpha_i, \gamma_i, \xi_i) &= R^2 + C \sum_i \xi_i \\ &- \sum_i \alpha_i \{R^2 + \xi_i - (\|x_i\|^2 - 2a.x_i + \|a\|^2)\} \\ &- \sum_i \gamma_i \xi_i \end{aligned} \tag{3}$$

L is minimized with respect to R, a, ξ_i and maximized with respect to α_i and γ_i. By applying the Karush-Kuhn-Tucker conditions, the following equations are derived:

$$\frac{\partial L}{\partial R} = 0 \Rightarrow \sum_i \alpha_i = 1 \tag{4}$$

$$\frac{\partial L}{\partial a} = 0 \Rightarrow a = \sum_i \alpha_i x_i \tag{5}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C - \alpha_i - \gamma_i = 0 \tag{6}$$

Equation 5 shows that the center of the hypersphere is a linear combination of the objects x_i for which the Lagrange multipliers $\alpha_i \neq 0$. Those objects are therefore called the support vectors of the description (sv). Furthermore, based on equation 6 and due to the fact that $\forall i \; \alpha_i \geq 0$ and $\gamma_i \geq 0$, the following constraint can be derived:

$$0 \leq \alpha_i \leq C, \quad \forall i \in 1, \ldots, N \tag{7}$$

By substituting the Equations 4, 5, 6 into Equation 3, the optimization problem can be specified as follows:

$$\begin{aligned} &\textit{Maximizing} \quad L = \sum_i \alpha_i (x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \\ &\textit{s.t.} \quad 0 \leq \alpha_i \leq C \quad \textit{and} \quad \sum_i \alpha_i = 1, \quad \forall i \in 1, \ldots, N \end{aligned} \tag{8}$$

Solving Equation 8 gives the set of α_i. Based on this, the radius of the hypersphere, which is the distance from its center to any support vector on its boundary, can be computed:

$$R^2 = \|x_k - a\|^2 = (x_k \cdot x_k) - 2 \sum_i \alpha_i (x_i \cdot x_k) + \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \tag{9}$$

Subsequently, an object z is classified as belonging to the target class if the distance from z to the center of the hypersphere a is smaller or equal to the radius R of the hypersphere.

$$\|z - a\|^2 = (z \cdot z) - 2 \sum_i \alpha_i (x_i \cdot z) + \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \leq R^2 \tag{10}$$

Similarly to the Support Vector classifiers, a more flexible data description can be obtained by using the kernel trick. Let $\phi : \mathbf{R} \rightarrow \mathbf{H}$ be a function which maps the target set into the Hilbert space $\mathbf{H}$. This can be achieved by replacing the inner product in the previously derived equations by an appropriate kernel function $K(x_i, x_j) = (\phi(x_i), \phi(x_j))$.

B. Active Learning

Supervised and Semi-Supervised Learning rely on labeled samples in order to successfully perform classification or regression tasks ([14]). The robustness and the performance of a model trained with such techniques highly depend on the amount of annotated samples available as well as on the quality of the annotation process. Thus the creation of a large and reliable annotated corpus is known to be the bottleneck in the domain of Supervised and Semi-Supervised Learning. In most real world applications, the annotation process is hard, expensive both in time and costs, cumbersome and error prone, in particular when the amount of data needed is very large ([15]). Furthermore, a large corpus of annotated samples may contain redundant and irrelevant samples for the classification, respectively the regression task.

Active learning is a widely used technique to address these issues ([16]). The main characteristic of active learning is the ability of the learner to select the samples from which it learns. More specifically, the learner selects the most informative unannotated samples to be annotated by an oracle (e.g. a human annotator). In this way the learner is able to improve its performance and accelerate its learning process, while reducing the costs of annotating an entire corpus.

Several approaches have been proposed for the selection of the samples, the most common approaches being:

1) Membership Query Synthesis [17] : In this approach, the learning model generates itself new samples by synthesizing some selected unannotated instances, before requesting an annotation of the synthetic samples from the oracle.
2) Stream-based Selective Sampling [18]: In this approach samples are presented in a stream and the learning model decides whether or not to query its label.
3) Pool-Based Sampling [19]: In Pool-Based Sampling, it is assumed that a small set of annotated samples L is available, additionally to a large pool of unannotated instances U; the learning model then selectively chooses one or more informative instances from the pool U to be annotated and added to the pool L. A new model is then learned using the pool L and the process is iterated.

The informativeness of each unannotated instance is evaluated by the learner. Depending on the evaluation strategy, the most informative instances are selected and annotated. Several evaluation strategies have been proposed through the years, amongst others:

1) Query by committee [20]: This strategy involves maintaining a committee of prediction models, which are trained on the current annotated pool L; each committee member votes on the annotation of each queried sample; the informativeness of the queried sample can be expressed in two different ways, depending on the nature of the models in the committee: the most informative sample is considered to be the one about which most of the committee members either disagree or agree.
2) Uncertainty sampling [21]: This strategy consists in querying those unannotated samples for which the current prediction model is least certain about the corresponding labels;
3) Expected error reduction [22]: This strategy consists in querying those unannotated samples that would most reduce the generalization error of the prediction model.

III. SVDD BASED ACTIVE LEARNING

In this section we present a pool-based active learning approach based on outlier detection combined with binary classification. To this end, both the SVDD and the binary SVM are used.

Since the approach is pool-based, the algorithm begins with a small set of annotated samples L and a large set of unannotated samples U. At each iteration the most uncharacteristic samples of the unannotated set are detected using a committee of randomly generated SVDD models. Each model is trained and tested on the entire pool U. The samples are then selected by majority vote, thus are those for which most of the committee members agree. Subsequently there are two variants of the algorithm:

1) The first variant consists in annotating the whole selection of outliers, adding the annotated samples to the pool L, training a binary SVM on L and applying the model to U. This process is iterated until a termination criteria is satisfied. It should be noted at this point that the committee, which is trained at each iteration on the unannotated set, decides solely on the samples to be annotated. The binary SVM is used in this variant just to measure the performance of the active learning method and does not get involved in the selection process of the samples to be annotated.
2) The second variant consists in involving the binary SVM in the selection process of the samples to be annotated. At each iteration k, the binary SVM model trained at the previous iteration $k-1$ on the annotated set L_{k-1} is tested on the set of samples selected by the committee of SVDD models. Instead of annotating the whole selection as in the previous variant of the algorithm, a subset of the selected samples is annotated based on the binary SVM model. Those samples that are annotated are added to L_k, while the other samples are left in U. This process is iterated until a termination criteria is satisfied.

Four Query strategies with three of them based on the output of the binary SVM model are used to select the samples to be annotated:

- *Random Sampling*: the samples to be annotated are queried randomly among those selected by the committee of SVDD models.
- *Shannon Entropy*: this is a widely used uncertainty measure [23]; the uncertainty measurement function based on the entropy estimation of a model's posterior distribution can be expressed as:

$$H(x) = -\sum_{y \in Y} P(y|x) log P(y|x), \tag{11}$$

Fig. 1: Frame sequence depicting a typical event (laughter in this case) characterized by the presence of observable facial gestures. The approach should enable a system to distinguish such events from sequences where no facial gesture is observable.

where y is the label and x is the feature vector of the sample; the selection of the most informative sample s can be formulated as:

$$s = \arg\max_{x \in U} H(x) \tag{12}$$

- *Nearest To The Border*: this strategy involves annotating the samples that are closest to the decision boundary of the binary SVM model.
- *Farthest from The Border*: this strategy involves annotating the samples that are farthest from the decision boundary of the binary SVM model.

The second variant of the algorithm helps the annotator by enabling him to chose the number of samples to be annotated at each iteration. This is not possible while using the first variant of the algorithm, since the SVDD models are generated randomly. Based on the work of Chang et al [24], for each SVDD model the cost parameter C is chosen in the interval $[\frac{1}{N}, 1]$ where N is the number of samples in the unannotated set U. In order to have a more flexible boundary the radial basis function (RBF) kernel is used for each SVDD model. The RBF kernel is given by

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{13}$$

For each SVDD model the parameter γ is randomly selected in the interval $[\frac{1}{f}, 1]$ where f is the number of feature vectors. This randomness in the parameter selection for each member of the committee leads to a fluctuation of the number of samples selected as outliers at each iteration. Nevertheless, instead of searching through the whole set of unannotated samples at each iteration, the binary SVM focuses uniquely on a small set of uncharacteristic samples. By annotating such instances the algorithm should be able to quickly compute an efficient model for the classification of the unseen instances.

IV. Data Description

In this work a dataset is used that was recorded as part of an interaction experiment [8]. The data consists of a set of 30 video sequences with a length of about 30 minutes, each depicting an experiment involving a participant who interacts by speech and touch input with a multi-modal system. The experiment involves resolving a series of subsequent puzzles with different levels of difficulty. Two cameras capture the interaction between the participant and the system. One of them is placed in front of the participant and records his/her face while the other one is used to monitor the progress of the experiment. Audio is recorded as well. The experiment is designed such that the participant has to solve a series of timed puzzles with increasing difficulty. As an incentive, the participant can earn money based on his/her performance. The longer it takes to solve a puzzle, the less money is rewarded.

While the dataset has not been primarily designed to elicit emotions in the participants, spontaneous reactions occur (more or less) frequently in the interaction process. By the design of the experiment the content of the recordings mainly includes two different reaction types. First, a so called neutral class that spans the search phases of each puzzle and the event class which comprises events such as exclamation of happiness or frustration (if the puzzle was solved correctly/incorrectly) (see Figure 1). The distribution between neutral and event is skewed towards neutral (compare Table I). The two classes have been annotated manually for a set of 6 participants drawn such that expressive and non expressive persons are included. For more details about the annotation or the dataset, the reader is referred to [9].

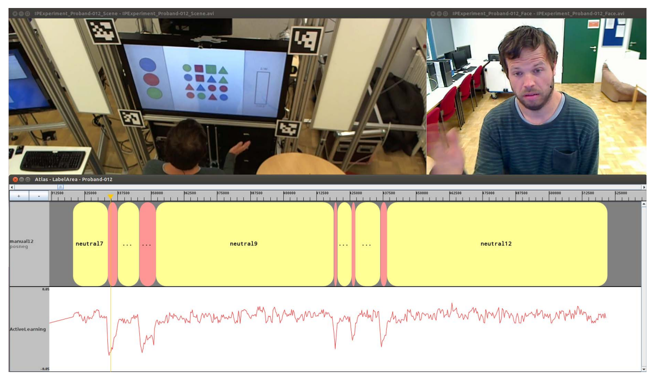

Fig. 2: Interaction Game dataset. **Top left:** Interface with an instance of the puzzle. **Top right:** Typical recording of a participant. **Bottom:** Annotated events such as search phases and outcome of the task.

V. Experimental Settings & Results

The proposed approach was tested on 6 participants drawn from the Interaction Game dataset described earlier. As it can be seen in Table I, each participant's set exhibits a severe imbalance ratio between both event and neutral classes. Appropriate features were extracted by first localizing and extracting the facial region in each frame of a video sequence. Subsequently, the facial region was divided into a fixed number of overlapping blocks. The local binary pattern on three orthogonal plane (LBP-TOP) [25] was then applied on each cuboid consisting of each block of facial region for the entire video sequence to generate the corresponding histogram of description. Finally the histograms generated from each single cuboid were concatenated into a final feature vector describing the video sequence.

In order to measure the performance of the proposed approach a stratified 3-fold cross validation was applied on each

TABLE I: Dataset description and class membership distributions.

Participant ID	Neutral Instances	Events
12	3387	108
15	3365	135
17	4149	66
23	3406	208
26	3409	429
30	4489	665

single set. The active learning algorithm was then successively applied to two of the folds, while the performance of the generated model was tested on the left out fold. A binary RBF SVM was also trained on the annotated two folds and tested on the left out fold. The result of this classification is used as baseline in order to assess the performance of the active learning approach. Furthermore, since the sets are particularly imbalanced, the geometric mean ($g - mean$) was chosen as performance metric in order to assess the performance of the generated model on the left out fold. The geometric mean is defined as follows:

$$g - mean = \sqrt{sensitivity \cdot specificity}, \tag{14}$$

where $sensitivity$ is the accuracy on the neutral class and $specificity$ is the accuracy on the event class.

During the experiments the size of the committee is fixed to 10. For the second variant of the approach a maximum of 20 samples are annotated after each iteration. The annotation of the selected samples is followed by a classification with a binary RBF SVM. This classification is preceded by a grid search in order to select the appropriate parameters for the SVM classification. At the beginning of each experiments a small set of 59 instances chosen randomly are annotated to form the initial annotated pool U.

Figure 3 shows the ratio of annotated samples for the dataset specific to the participant 12. The results of the annotation process are plotted for each of the variants:

- OD: Consists in annotating the whole set of samples depicted as outliers by the committee.
- ODCB: Detection of the outliers closest to the border of the model trained in the previous step.
- ODFB: Detecting and annotating those outliers that are farthest from the border of the model.
- ODE: Detection based on the entropy of the distribution.
- ODRS: Detection of the outliers followed by the random sampling approach.

The figure shows clearly that the first variant of the approach leads to a quicker annotation of the dataset, while the second variant is much slower. More precisely, the first variant OD needs around 60 to 65 learning rounds to annotate around 80% of the entire dataset. Meanwhile, the second variant (ODCB, ODFB, ODE, ODRS) needs around 100 learning rounds to reach the same amount. This is due to the randomness of the generation of the committee members which occurs at each learning round and causes the fluctuation in the number of samples that are classified as being outliers.

Meanwhile figure 4 shows the results of the approach on each participant's set. These results depict clearly that the algorithm is able to generate a model that performs at the very least as good as a supervised model trained on the whole set of annotated samples, and this by annotating between 50% and 60% of the available samples. In some cases the algorithm generates a model that performs even better than a fully supervised trained model. It can also be seen that limiting the number of annotated samples at each iteration instead of annotating the whole set of detected outliers does not hinder the performance of the approach. The main difference of both variants is the number of iteration that is needed in each case to annotate a precise amount of samples (the reader is referred to figure 3). The different query strategies perform also well, but in order to find out which query strategy performs better, further experiments are to be undertaken.

In figure 5, the result of the annotation process can be seen. For a small time window, the continuous distances to the hyperplane are visualized together with the manually annotated ground truth. As can be clearly seen, the classifier detects deviations from the neutral class in each case (here because of wrongly solved puzzles) as negative distance to the decision border.

VI. Conclusion

In this work, we proposed an active learning approach for the annotation and detection of unusual events in video sequences. The first step of the approach consists in localizing and selecting uncharacteristic samples using a committee of SVDD models. The search for informative samples does not cover the whole set of unannotated samples, but instead focuses on those samples tagged as uncharacteristic by the committee. Subsequently, common active learning query strategies are applied to select the most informative samples to be annotated and used for the training of the base learner. The results of the undertaken experiments on a set of video sequences drawn from an interactive game dataset clearly depict the effectiveness of the proposed approach for both annotation and classification tasks. As a matter of fact, the proposed approach performed in some cases better than a binary SVM trained on the fully annotated set of samples. Further experiments involving more datasets are to be undertaken for a better assessment of the performance of the approach. These experiments should also include the comparison of the approach with other common and successful active learning algorithms. Moreover the applicability of the approach to multimodal datasets should be investigated.

Acknowledgment

This paper is based on work done within the Transregional Collaborative Research Centre SFB/TRR 62 Companion-Technology for Cognitive Technical Systems funded by the German Research Foundation (DFG). Markus Kächele is supported by a scholarship of the Landesgraduiertenfoderung

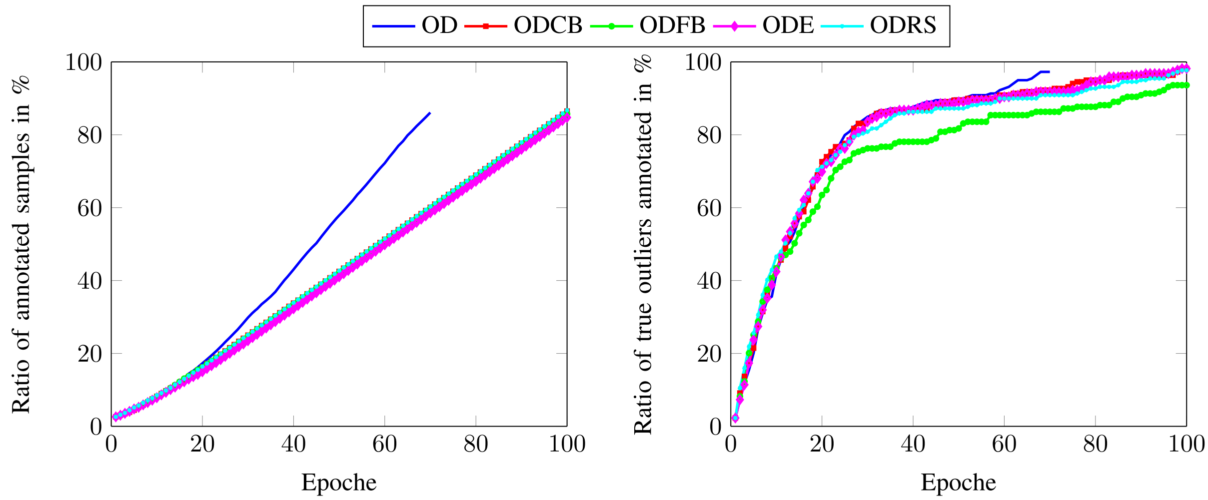

Fig. 3: Active learning results specific to the participant 12. **Left:** Using the first approach, nearly 80% of the samples are annotated after a total of 70 iterations. Around 100 iterations are needed for the other variants in order to annotate the same amount of samples. **Right:** Almost all true outliers are detected using each variant, whereby the first variant needs a total of 70 iterations and the other variants nearly 100 to reach the same amount.

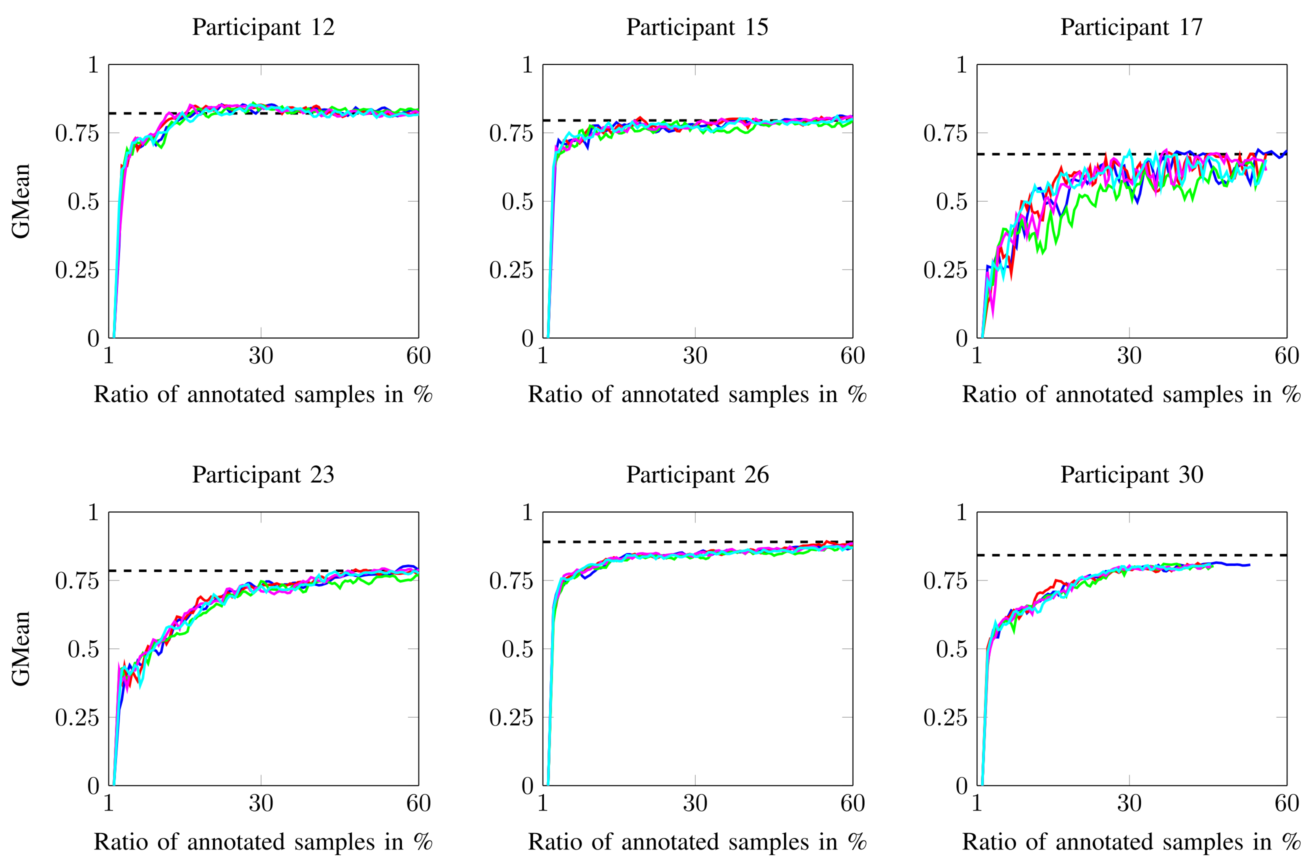

Fig. 4: 3-fold cross validation active learning results for each of the 6 participants; (a) Participant 12, (b) Participant 15, (c) Participant 17, (d) Participant 23, (e) Participant 26, (f) Participant 30

Baden-Wüttemberg at Ulm University. Further we would like to thank Felix Schüssel and Frank Honold for providing the

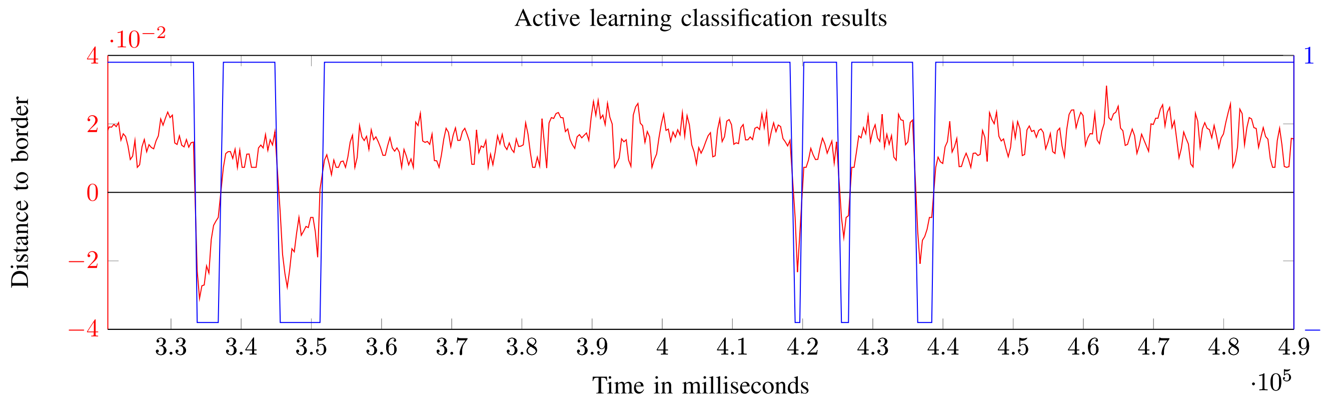

Fig. 5: Active learning classification of a video segment specific to the participant 12. During the manual annotation process, neutral instances were labeled 1 and events were labeled -1. The manual annotation results are plotted in blue. A binary SVM trained with the proposed approach on a portion of the entire video sequence was applied on this specific video segment. The results of the classification, which in this case are represented by the distances of the samples to the hyperplane, are depicted in red. It can be clearly seen that the classifier performs rather well and detects each of the manually annotated events.

interactive game dataset.

References

[1] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 26, no. 1, pp. 97–107, Jan 2014.

[2] D. Angluin, "Queries and concept learning," *Machine Learning*, vol. 2, no. 4, pp. 319–342, 1988. [Online]. Available: http://dx.doi.org/10.1007/BF00116828

[3] S. Ayache and G. Quenot, "Evaluation of active learning strategies for video indexing," in *Content-Based Multimedia Indexing, 2007. CBMI '07. International Workshop on*, June 2007, pp. 259–266.

[4] R. Morris and D. McDuff, *Crowdsourcing Techniques for Affective Computing*. Oxford University Press, 2014.

[5] S. Bandla and K. Grauman, "Active learning of an action detector from untrimmed videos," in *IEEE International Conference on Computer Vision*. IEEE, 2013, pp. 1833–1840.

[6] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proceedings of the Ninth ACM International Conference on Multimedia*, ser. MULTIMEDIA '01. New Yoork, NY, USA: ACM, 2001, pp. 107–118.

[7] S. Abe, *Support Vector Machines for Pattern Classification*. London, England: Springer, 2005.

[8] F. Schüssel, F. Honold, M. Schmidt, N. Bubalo, A. Huckauf, and M. Weber, "Multimodal interaction history and its use in error detection and recovery," in *Proceedings of the 16th International Conference on Multimodal Interaction*, ser. ICMI '14. New York, NY, USA: ACM, 2014, pp. 164–171. [Online]. Available: http://doi.acm.org/10.1145/2663204.2663255

[9] P. Thiam, S. Meudt, M. Kächele, G. Palm, and F. Schwenker, "Detection of emotional events utilizing support vector methods in an active learning hci scenario," in *Proceedings of the 2014 Workshop on Emotion Representations and Modelling for HCI Systems*, ser. ERM4HCI '14. New York, NY, USA: ACM, 2014, pp. 31–36. [Online]. Available: http://doi.acm.org/10.1145/2668056.2668062

[10] B. Schlkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," in *Advances in Neural Information Processing Systems 12*, S. S.A., L. T.K., and M. K., Eds. MIT Press, 2000, pp. 582–588.

[11] M. T. David and P. D. Robert, "Support vector data description," *Machine Learning*, 2004.

[12] S. Chaki, A. Verma, A. Routray, W. Mohanty, and M. Jenamani, "A one-class classification framework using svdd: Application to an imbalanced geological dataset," in *Students' Technology Symposium (TechSym), 2014 IEEE*.

[13] M. Kächele, P. Thiam, G. Palm, and F. Schwenker, "Majority-class aware support vector domain oversampling for imbalanced classification problems," in *Artificial Neural Networks in Pattern Recognition*, ser. Lecture Notes in Computer Science, N. El Gayar, F. Schwenker, and C. Suen, Eds. Springer International Publishing, 2014, vol. 8774, pp. 83–92. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-11656-3_8

[14] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, USA: Springer, 2006.

[15] N. V. Chawla, *Data Mining For Imbalanced Datasets: An Overview*. New York, USA: Springer, 2005, ch. 40, pp. 853–867.

[16] B. Settles, "Active learning literature survey," University of Wisconsin–Madison, Computer Sciences Technical Report, 2009.

[17] D. Angluin, "Queries revisited," in *Proceedings of the 12th International Conference on Algorithmic Learning Theory*. London, UK: Springer, 2001, pp. 12–31.

[18] L. E. Atlas, D. A. Cohn, and R. E. Ladner, "Training connectionist networks with queries and selective sampling," in *Advances in Neural Information Processing Systems 2*. Morgan-Kaufmann, 1990, pp. 566–573.

[19] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," in *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New-York, USA: Springer, 1994, pp. 3–12.

[20] *Analysis of rare categories*. Springer, 2012, ch. Survey and Overview.

[21] S. H. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. New York, NY, USA: ACM, 1992, pp. 287–294.

[22] N. Roy and A. McCallum, "Toward optimal active learning through sampling estimation of error reduction," in *Proceedings of the Eighteenth International Conference on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 441–448.

[23] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10 – 21, 1949.

[24] W. cheng Chang, C. pei Lee, and C. jen Lin, "A revisit to support vector data description (svdd)."

[25] Z. Guoying and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.