Active Learning for Audio-based Home Monitoring

Mulu Weldegebreal Adhana MULUWELDEGEBREAL.ADHANA@KULEUVEN.BE
ESAT-ETC/AdvISe, KU Leuven TC Geel, Kleinhoefstraat 4, 2440, Geel, Belgium
Bart Vanrumste BART.VANRUMSTE@KULEUVEN.BE
ESAT-ETC/AdvISe, KU Leuven TC Geel, Kleinhoefstraat 4, 2440, Geel, Belgium
ESAT-STADIUS, KU Leuven, Kasteelpark Arenberg, 3001, Leuven, Belgium
iMinds, Medical IT, Kasteelpark Arenberg, 3001, Leuven, Belgium
Peter Karsmakers PETER.KARSMAKERS@KULEUVEN.BE
ESAT-ETC/AdvISe, KU Leuven TC Geel, Kleinhoefstraat 4, 2440, Geel, Belgium
ESAT-STADIUS, KU Leuven, Kasteelpark Arenberg, 3001, Leuven, Belgium

Keywords: active learning, supervised learning, semi-supervised learning, sound recogntion

Abstract

In this work we investigate an active learning scheme in the context of audio-based home monitoring. Our experimental result shows that, using the public dataset NAR, the active learning scheme applied to a fast approximation of kernel logistic regression reduces the manual annotation by more than 84% while attaining similar performance to the supervised learning approach in terms of classification accuracy.

1. Introduction

Most standard supervised learning algorithms assume the presence of labeled data in abundance. However, acquiring sufficient annotated dataset is proven to be costly. Semi-supervised learning is a technique applied in situations where there is a lack of labeled data. This work focuses on a special type of such learning called *active learning*, which allows learning algorithms to assume that only a small portion of a dataset is annotated. The rest of the dataset without label is actively annotated based on defined labeling rules. The work proposed by (Su & Fung, 2012), demonstrates the use of active learning in recognizing the types of emotions in music. Furthermore, (Stikic et al., 2008) applied active learning for sound based activity recognition. Other works in (Nigam & McCallum, 1998; Tong & Koller, 2002; Qi et al., 2008; Yang et al., 2009; Joshi et al., 2009; Li & Guo, 2013) show pool-based SVM active learning applied to solve text and image classifications.

Although generative models are applied as well in audio-based classification problems, we focus on discriminative modeling as a base learning. There are several powerful discriminative classifiers already developed; Support Vector Machine (SVM) is considered as a benchmark in this domain. The experiment carried by (Zhu & Hastie, 2012) shows that Kernel Logistic Regression (KLR) has comparable performance to SVM in terms of classification accuracy. Moreover, KLR has distinct advantages over SVM. The outcome of the KLR model, unlike in SVM, has probabilistic meaning and it also has a natural extension to a multiclass classification problem. To minimize the computational cost, (Karsmakers, 2010) proposed a fast approximate version of KLR, termed Fixed Size KLR (fs-KLR). These are important attributes to implement active learning schemes in the context of audio-based home monitoring.

2. Active Learning Methodology

We implement active learning scheme according to the following procedures. Starting from a small pool of annotated data an initial classifier fs-KLR model is learned. This classifier model then starts processing unannotated examples. When an example is attributed a high uncertainty value a label is asked from the user. In the other cases the newly acquired example is automatically assigned the label from the

Appearing in *Proceedings of Benelearn 2016*. Copyright 2016 by the author(s)/owner(s).

most probable class according to the current available model. Preliminary experiments indicated that the following uncertainty measure gives satisfactory performance. A class prediction is attributed to be uncertain if the difference of the posterior probabilities of the two most likely classes (according to the current model) is below some threshold, T. In such cases the user is required to provide the desired label for the novel data point. For the other cases the novel example is automatically assigned the most likely class label. When an example or set of examples is added to the training set, the classifier model is updated. This iterative process is repeated until the whole training set is annotated.

3. Experiments and Results

In this experiment, we use the dataset of sounds recorded using the humanoid robot NAO that can be found publicly [1] with 21 different classes and 431 annotated examples. 2/3 of the dataset is used to train the model. The remaining samples are used for evaluation. Each sound is represented by a set of 14-dimensional Mel-Frequency Cepstral Coefficients (MFCC) (Han et al., 2006) using a window size of 25 ms with an overlap of 10 ms. This set is further compacted to a single 28-dimensional feature vector by computing the mean and variance of each MFCC dimension.

Using the whole training set, we perform five-fold cross validation to tune the hyper-parameters. Trivially, having less data to tune the hyper-parameters increases the risk of having improper values assigned to them which will impact the classification performance. Although an important topic for practical use, for now to exclude these effects the hyper-parameters are tuned once using the full annotated data. As a preprocessing step fs-KLR requires the selection of Prototype Vectors (PVs). In these experiments 80 PVs were selected using all available training set without incorporating knowledge about the class labels. We assume a single annotated training example per class to create and initialize model. Then the active learning procedure is executed as described in Section 2. At each active learning round, the test set is used to evaluate the model.

Using the full annotated dataset with a training set size of 284 and test set size of 147, KLR achieves 94.56% of classification accuracy and needed 1.95 s of time to estimate the model parameters. In addition to reducing the training time to 0.43 seconds, a slightly improved classification accuracy of 95.1% is achieved using fs-KLR.

[1] https://team.inria.fr/perception/nard/

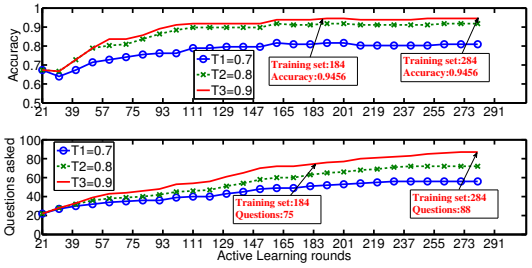


Figure 1. fs-KLR active learning evaluated in terms of accuracy and number of manual annotations averaged over 5 runs

Fig. 1 shows results of fs-KLR active learning both in terms of accuracy (see top plot) and the number of questions asked (see bottom plot) for manual annotation using three threshold values. Generally, the accuracy improves when the threshold value increases. Both the results for accuracy and number of questions were averaged over 5 runs. However, a higher threshold value causes a raise in the number of questions asked. Taking the highest threshold value at 0.9 guides the active learning to lessen the number of training examples to 184 (75 examples annotated manually) which obtains similar accuracy (94.56%) compared to the case where all training data was used. When selecting a random subset of training examples to estimate an fs-KLR model, a subset size 3 times that of the active learning approach was needed on average to achieve similar classification performance as that obtained in the case of active learning.

4. Conclusion and Future Works

We have demonstrated that KLR and fs-KLR can be deployed to perform audio-based home monitoring. We further showed using the active learning principles that while having similar accuracy compared to a fully supervised approach, only 26% of the training data needed manual annotation. We plan to adopt more powerful techniques of uncertainty measures and adaptive means of selecting threshold values to further reduce the number of data points that require annotation without decreasing the classification performance. We also envision to investigate a robust approach to select appropriate values for the hyper-parameters of the active learning scheme.

References

Han, W., Chan, C.-F., Choy, C.-S., & Pun, K.-P. (2006). An efficient mfcc extraction method in speech recognition. *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on* (pp. 145–148).

Joshi, A. J., Porikli, F., & Papanikolopoulos, N. (2009). Multi-class active learning for image classification. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 2372–2379).

Karsmakers, P. (2010). *Sparse kernel-based models for speech recognition*. Doctoral dissertation, PhD Thesis.

Li, X., & Guo, Y. (2013). Adaptive active learning for image classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 859–866).

Nigam, K., & McCallum, A. (1998). Pool-based active learning for text classification. *Conference on Automated Learning and Discovery (CONALD)*.

Qi, G.-J., Hua, X.-S., Rui, Y., Tang, J., & Zhang, H.-J. (2008). Two-dimensional active learning for image classification. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1–8).

Stikic, M., Van Laerhoven, K., & Schiele, B. (2008). Exploring semi-supervised and active learning for activity recognition. *Wearable computers, 2008. ISWC 2008. 12th IEEE international symposium on* (pp. 81–88).

Su, D., & Fung, P. (2012). Personalized music emotion classification via active learning. *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies* (pp. 57–62).

Tong, S., & Koller, D. (2002). Support vector machine active learning with applications to text classification. *The Journal of Machine Learning Research*, *2*, 45–66.

Yang, B., Sun, J.-T., Wang, T., & Chen, Z. (2009). Effective multi-label active learning for text classification. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 917–926).

Zhu, J., & Hastie, T. (2012). Kernel logistic regression and the import vector machine. *Journal of Computational and Graphical Statistics*.