

# Active Learning by Selecting New Training Samples from Unlabelled Data

Ho Gyeong Kim, Cheong-An Lee, and Soo-Young Lee

Department of Electrical Engineering and Brain Science Research Center  
Korea Advanced Institute of Science and Technology  
373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea

**Abstract.** Human utilizes active learning to develop their knowledge efficiently, and a new active learning model is presented and tested for automatic speech recognition systems. Human can self-evaluate their knowledge systems to identify the weak or uncertain topics, and seek for the answers by asking proper questions to experts (or teachers) or searching books (or webs). Then, the new knowledge is incorporated into the existing knowledge system.

Recently this active learning becomes very important to many practical applications such as speech recognition and text classification. On the internet abundant unlabelled data are available, but it is still difficult and time consuming to get accurate labels for the data. With the active learning paradigm, based on uncertainty analysis, a few data will be identified to be included in the training database and the corresponding labels will be asked to users. The answers will be incorporated into the current knowledge base by an incremental learning algorithm. This process will be repeated to result in a high- accuracy classification system with minimum number of labelled training data.

The active learning algorithm had been applied to both a simple toy problem and a real-world speech recognition task. We introduced a uncertainty measure for each unlabelled data, which is calculated from the current classifier. The developed algorithm shows better recognition performance with less number of labelled data for the classifier training. In the future we will also incorporate a smooth transition on the selection strategy based on the exploitation-exploration trade-off. At the early stage of learning human utilizes exploitation while exploration is applied at the later stage.