

Dipl.-Ing. Oliver Walter - 24.6.2014

Unsupervised Language Acquisition - The Unsupervised Learning of the Components of a Language

Abstract:

Unsupervised language acquisition is the task of acquiring the building blocks of a language without any supervision. Techniques to discover these components directly from audio recordings of continuous speech are known under the term zero resource speech technologies and are currently an area of active research. Components of a language are particularly words and their building blocks, the phonemes. A language model, describing word probabilities in the context of other words, is a further component of a language.

Conventional speech recognition systems rely on supervised learning, where the patterns to be recognized (the words) and their acoustic building blocks (the phonemes) are known a priori and where labeled training data are used to learn the models. In unsupervised vocabulary discovery, however, the word inventory is not known in advance and the training data come without labels. There are several applications where such a problem arises, such as the development of an automatic speech recognition (ASR) system for languages for which low or zero resources are available. Another application is teaching a robot an environment-specific vocabulary by speech interaction. Further, the techniques can serve as a computational model for early child language acquisition. Unsupervised word discovery is essentially based on searching for recurring sequential patterns in the audio data.

Several algorithms have been proposed for this task. Segmental dynamic time warping has been used for discovering repeated acoustic patterns in the input speech. Clustering techniques are then applied to group these patterns into word-like units. Other algorithms rely on a two-stage approach, where first the basic acoustic building blocks, i.e. the phonemes or similar entities of a language, are discovered and models are trained for them. Once this has been achieved, the real-valued speech input can be transcribed into sequences of symbols, the phoneme labels. The second task is then the discovery of lexical units, i.e., the words, as recurring sequences of these symbols. In the process of discovery words a language model can also be learned.

In this talk I will present algorithms for unsupervised language acquisition and particularly give a more detailed description for an hierarchical algorithm as one example on how to deal with the problem at hand.