Proceedings of the Post-Graduate Conference on Robotics and Development of Cognition
10-12 September 2012, Lausanne, Switzerland

Active Learning in a Computational Model of Word Learning

Maarten Versteegh[1,2], *Christina Bergmann*[1,2], *Louis ten Bosch*[2], *Lou Boves*[2]

[1]International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands
[2]Centre for Language Studies, Radboud University Nijmegen, The Netherlands
m.versteegh@let.ru.nl

1. Abstract

Infants learning the meaning of their first words are faced with a difficult task involving multiple sources of uncertainty. Both the speech they hear from their caretakers and their visual environment can be viewed as stochastic sources of information. This paper investigates a computational model of word learning that takes a closer look at these sources of uncertainty. We present an algorithm, based on exploiting the statistical regularities in the input, and show that this is sufficient to reliably learn words from speech and vision under highly adverse circumstances. The learning strategies the algorithm implements supports conjectures about infant learning, most notably the crucial roles of statistical learning and of taking an active approach to in processing the environmental input.

We identify two types of uncertainty. The first relates to the fact that the infant must discover recurrent acoustic stretches from a highly variable speech stream and eventually form meaningful combinations of these building blocks. Research has shown that the ability of young learners to detect patterns in spoken language is at least partly based on the use of the statistical properties of the speech signal.

The second type of uncertainty is due to possible inconsistencies between patterns in the speech stream and objects in the scene surrounding the learner. Any pairing of a spoken word and a communicative scene presents the learner with a large number of possible referents. When a caretaker utters a sentence pertaining to an object in the visual environment, it may thus be highly ambiguous from the viewpoint of the learner. Infants use statistical evidence across many individually ambiguous word-scene pairings to discover cooccurrences of auditory and visual information, thus resolving the ambiguity. They are also able to detect when form-referent pairings do not match their previous experience and will actively attempt to resolve these perceived mismatches by aligning a different visual referent with a given auditory form.

We present a computational model of word learning that incorporates both cross-situational statistical learning as well as the corrective process after detection of a mismatch, which has an analogy in active learning procedures in machine learning techniques. Learning in our model consists of multi-modal pattern finding, combining information extracted from speech utterances and images with the goal of both identifying the basic building blocks of language and forming word-referent pairings from situations of referential ambiguity.

Results from the model show the advantages of combining cross-situational statistics and active learning. We show that cross-situational statistics by itself provides the learner with a high level of robustness against referential ambiguity. Incorporating active learning, we then show that this improves the model's robustness even further, where the gain is proportional to the level of ambiguity in the input. Our model shows the importance of the active learning procedure attested in infants, in reliably learning words under highly uncertain conditions, such as the conditions an infant faces when learning his or her first words.