Discriminative Spoken Language Understanding Using Word Confusion Networks

Matthew Henderson, Milica Gašić, Blaise Thomson, Pirros Tsiakoulis, Kai Yu and Steve Young

Confusion Network Decoder

Problem: Semantic decoding while maintaining multiple paths in the speech recognition

Confusion Network Features are the expected n-gram frequencies e.g. IN THE = prob. first path + prob. second path

Dialogue Context Features are also extracted from the previous system act.

The Confusion Network Decoder has a chance of getting the correct semantics:

Dialogue act type, d-type = inform slots, S = {area=south}

Semantic Tuple Classifier Method

(Adopt approach of Mairesse et al. 2009)

• Discriminative approach so may use arbitrary features.
• Is trained from unaligned pairs of utterances and semantic labels
• One multiclass Support Vector Machine (SVM) is trained for the dialogue act type, and one binary SVM is trained for each of the slot-value pairs.

\[
P(d\text{-type}, S \mid u) = P(d\text{-type} \mid u) \prod_{s \in S} P(s\mid u) \prod_{s \in S} \left(1 - P(s\mid u)\right)
\]

Baseline Decoder

• A Phoenix grammar which has been used in publications in this domain for several years is used for comparison.
• Designed to be robust to recognition errors
• Spots keywords. Searches for longest spanning phrases

Conclusions

• Confusion network features and dialogue context features both significantly improve the accuracy of SLU, particularly in high noise environments
• The online evaluation shows dialogues were improved by the more robust SLU