Appears in the Proceedings of the Fourteenth National Conference on Artificial Intelligence, July 1997, Providence, RI

Naive Mixes for Word Sense Disambiguation*

Ted Pedersen
Department of Computer Science & Engineering
Southern Methodist University
Dallas, TX 75275–0122
pedersen@seas.smu.edu

The Naive Mix is a new supervised learning algorithm based on sequential model selection. The usual objective of model selection is to find a single probabilistic model that adequately characterizes, i.e. fits, the data in a training sample. The Naive Mix combines models discarded during the selection process with the best–fitting model to form an averaged probabilistic model. This is shown to improve classification accuracy when applied to the problem of determining the meaning of an ambiguous word in a sentence.

A probabilistic model consists of a parametric form and parameter estimates. The form of a model describes the interactions between the features of a sentence with an ambiguous word while the parameter estimates give the probability of observing each possible combination of feature values in a sentence.

The class of models in a Naive Mix is restricted to decomposable log–linear models to reduce the model search space and simplify parameter estimation. The form of a decomposable model can be represented by an undirected graph whose nodes represent features and whose edges represent the interactions between features. The parameter estimates are the product of the marginal distributions, i.e. the maximal cliques in the graph of the model.

Model selection integrates a search strategy with an evaluation criterion. The search strategy determines which decomposable models are evaluated during the selection process. The evaluation criterion measures the fit of each model to the training sample. (Pedersen, Bruce, & Wiebe 1997) report that the strategy of forward sequential search (FSS) and evaluation by Akaike's information criteria (AIC) selects models that serve as accurate classifiers for word–sense disambiguation. Here, this combination is shown to result in Naive Mixes that improve the accuracy of disambiguation over single selected models.

Model selection guided by FSS evaluates the fit of decomposable models at increasing levels of complexity, where complexity is defined as the number of edges in the graph of the model. The best–fitting model of complexity level i is designated the current model, m_i. The models evaluated at complexity level $i+1$ are generated by adding one edge to m_i and checking that the resultant model is decomposable. The evaluation begins with the model of independence where there are no interactions between features, $i = 0$, and ends when none of the generated models of complexity level $i+1$ sufficiently improves on the fit of m_i.

The result is a sequence of models $(m_0, m_1, \ldots, m_r)$, where the subscript represents the number of edges in the model and m_r is the best–fitting model. Parameter estimates are made for every model in the sequence. These estimates are averaged across all of the models to formulate a Naive Mix. This is an averaged joint probability distribution and serves as the basis of a probabilistic classifier. An ambiguous word is assigned the sense that maximizes the averaged parameter estimates given the observed features in the sentence.

The Naive Mix is a general–purpose method that performs feature selection and has a compact and intuitive graphical representation. Experimental results disambiguating four nouns, four verbs, and four adjectives show that it is competitive with a variety of machine learning algorithms. The average accuracy over twelve words for the Naive Mix is 85% while a decision tree (C4.5) achieves 86%, rule induction (CN2) 84%, nearest–neighbor classification (PEBLS) 84%, and Naive Bayes classification 85%. These results are discussed in detail in (Pedersen & Bruce 1997).

Acknowledgments

This research was supported by the Office of Naval Research under grant number N00014-95-1-0776.

References

Pedersen, T., and Bruce, R. 1997. A new supervised learning algorithm for word sense disambiguation. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence.*

Pedersen, T.; Bruce, R.; and Wiebe, J. 1997. Sequential model selection for word sense disambiguation. In *Proceedings of the Fifth Conference on Applied Natural Language Processing.*