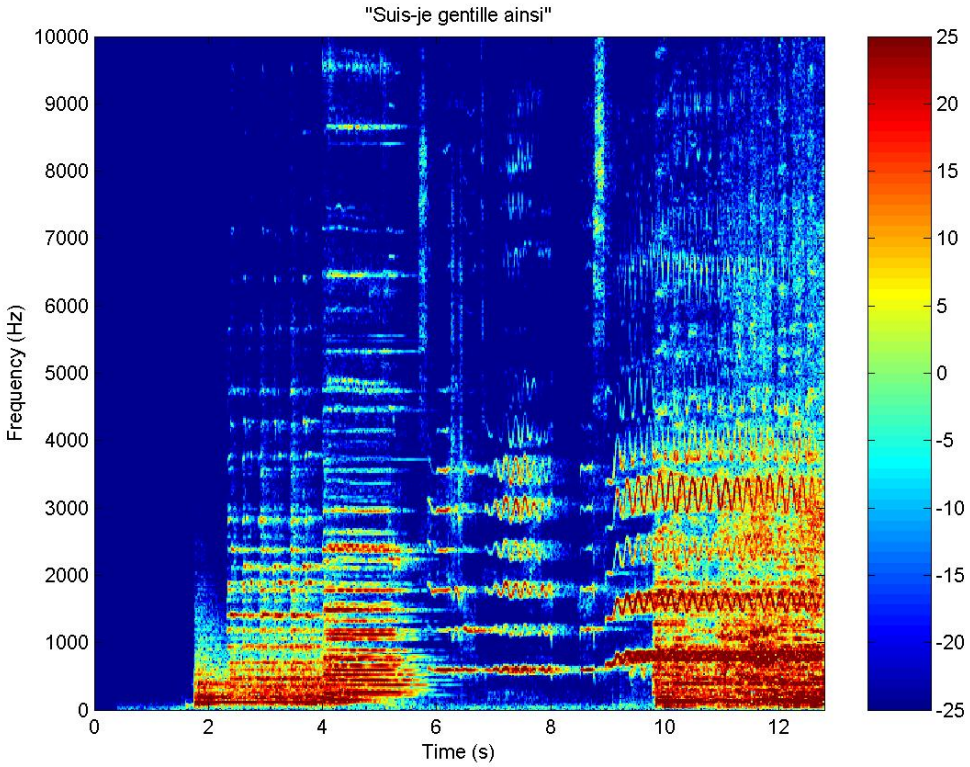
Solo voice detection using the SVM-HMM

Christine Smit

December 11, 2007

The problem: sequence labeling



Structured Learning

- Learn the entire structure (i.e. sequence, string, tree…) of the problem

What is the SVM-HMM?

- A specific case of

Set of input feature sequences

Set of output label sequences

Discriminant function →

$$F : X \times Y \rightarrow \Re$$

$$f(x;\omega) = argmax_{y \in Y} \; F(x,y;\omega)$$

Combined feature representation

Formulate for "HMM" here →

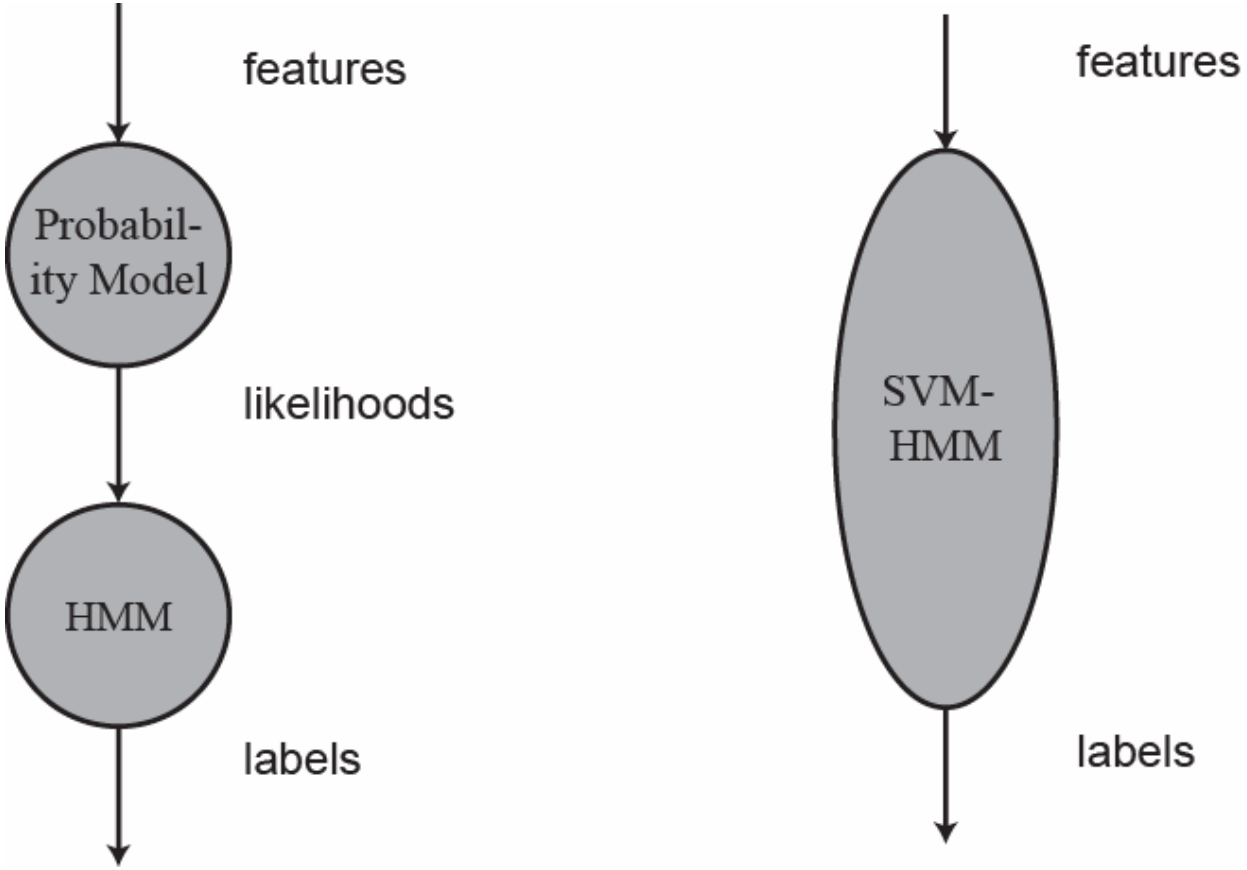
$$F(x,y;\omega) = \langle \omega, \Psi(x,y) \rangle$$

$$\forall i \in \{1,\ldots,n\}, \; \forall y \in Y \backslash y_i : \; \langle \omega, \Psi(x_i,y_i) - \Psi(x_i,y) \rangle \geq 0$$

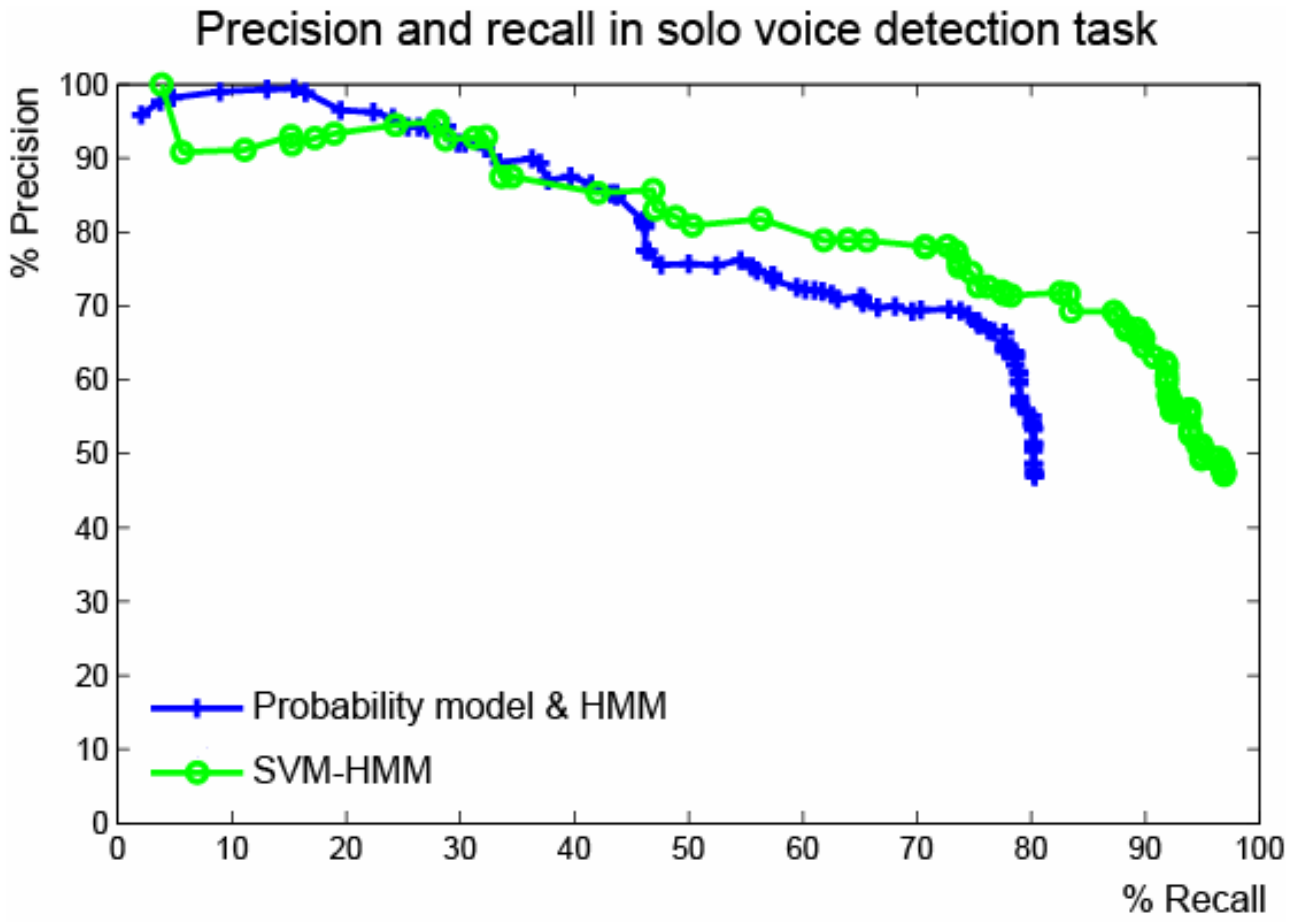
Features

- Ratio of pitch cancelled residual to original power
- normalized original power

Different strategies



Results


Precision and recall in solo voice detection task
% Precision
% Recall
0
10
20
30
40
50
60
70
80
90
100
Probability model & HMM
SVM-HMM

References

- I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun, Large Margin Methods for Structured and Interdependent Output Variables, Journal of Machine Learning Research (JMLR), 6(Sep):1453-1484, 2005. http://jmlr.csail.mit.edu/papers/volume6/tsochantaridis05a/tsochantaridis05a.pdf

- Y. Altun, I. Tsochantaridis, T. Hofmann, *Hidden Markov Support Vector Machines*. International Conference on Machine Learning (ICML), 2003.