in the PROSITE motif database. Of the sequences, we use the 32 which are actually in the RCI sequences as positive examples and the others as negative examples. We evaluated our method and the Baum-Welch by the problem of distinguishing the positive examples from the negative ones. Note that this problem is difficult enough, since we must classify the sequences having a consensus pattern into two classes, i.e. positive and negative.

In a series of experiments, both positive and negative examples are used for learning our method while the Baum-Welch uses positive examples alone. The HMMs used are fully-connected, in each of which a state connects to any state, and have six to sixteen states. The discrimination ability of our method and the Baum-Welch are measured by the average number of discrimination errors of ten trials with random initial values. In each trial, the stopping condition of the two iterative learning methods are determined so as to mininize the number of discrimination errors. Figure 1 shows the average number of errors varies depending on the number of states.

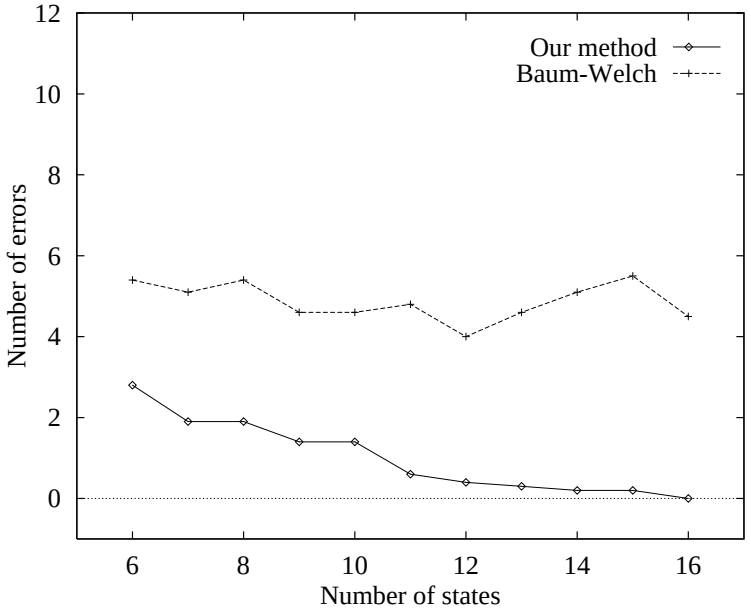

Figure 1: Average number of discrimination errors

As shown in the figure, our method greatly reduces the average number of errors made by the Baum-Welch in any fully-connected HMMs having six to sixteen states. This result indicates that negative examples are effectively used by our method for improving HMMs to have higher discrimination ability.

References

[1] A. Krogh, M. Brown, I. S. Mian, K. Sjölander and D. Haussler "Hidden Markov Models in Computations Biology Application to Protein Modeling," *Journal of Molecular Biology*, Vol. 235, pp. 1501-1531, 1994.

Learning Hidden Markov Models Using Back-Propagation through Time

Hiroshi Mamitsuka
mami@sbl.cl.nec.co.jp

C&C Research Labs., NEC Corporation
1-1, Miyazaki 4-chome, Miyamae-ku, Kawasaki, Kangawa 216, Japan

1 Introduction and Methods

Hidden Markov models (HMMs for short) are a type of stochastic models that have been used for a number of applications including speech recognition. Recently, HMMs have been applied to biological sequences[1] as tools of sequence alignments, validating motif patterns and even predicting protein structures.

The main reason why HMMs are widely used is that they have an efficient learning algorithm which is called the Baum-Welch or the Forward-Backward and is categorized into a class of the EM (Expectation Maximization) algorithm. However, the biggest disadvantage of this algorithm lies in its low discrimination ability when only a small dataset is given.

On this problem, we present a learning method of HMMs using back-propagation through time, which is a typical learning method of neural-network learning, by regarding HMMs as a type of real-time recurrent neural networks having a linear input-output function. This smooth learning method of HMMs allows us to deal with both positive and negative examples for a given sequence class, and appears to solve the problem of low discrimination ability of HMMs for given small datasets. Through the experiments using actual protein sequences, we show how our method improves the discrimination ability of HMMs trained by the Baum-Welch algorithm.

2 Experimental Results and Conclusion

We use *the Receptor Cytokinase I* (RCI for short) motif for experiments. ¿From the SwissProt database release 29.0, we obtained 47 sequences of 15 to 16 residues long, all of which have the consensus pattern of RCI, i.e. "C - [LVFYR] - x(7,8) - [STIVDN] - C - x - W", which is noted

馬見塚 拓： NEC C&C 研究所，〒 216 川崎市宮前区宮崎 4-1-1