A New Algorithm for Automatic Configuration of Hidden Markov Models

Makoto Iwayama¹ and Nitin Indurkhya² and Hiroshi Motoda¹

¹ Hitachi Advanced Research Laboratory, Hatoyama, Saitama 350-03, Japan
² Telecom Research Laboratories, 770 Blackburn Road, Clayton, Vic 3168, Australia

Abstract. Hidden Markov Models (HMM) (i.e. doubly stochastic probabilistic networks) have been widely used in analyzing time-series data such as those obtained from speech and molecular biology. A crucial issue in modeling time-series data using HMM, is the problem of determining the appropriate model architecture: the number of states and the links between the states. While current HMM training procedures iteratively optimize model parameters, they usually require the model configuration to be fixed. The task of model configuration is done manually by trained experts. In this paper we present a procedure that addresses the problem of automatically configuring HMM's. It starts with a large, possibly over-fitted HMM, and attempts to prune it down to the appropriate complexity fit. The procedure can be seen as a generalization of the well-known iterative Baum-Welch Algorithm. The parameter re-estimates in our procedure can be formally derived and its local convergence can be formally proved. Compared to existing methods, our procedure offers the following advantages: (1) better convergence characteristics than the standard Baum-Welch algorithm, (2) automatic reduction of model size to the right complexity fit, (3) better generalization, and (4) relative insensitivity to the initial model size. We demonstrate these features by presenting empirical results on the problem of recognizing DNA promoter sequences.

1 Introduction

Hidden Markov Models (HMM) (i.e., doubly stochastic probabilistic networks) have been widely used in analyzing time-series data such as those obtained from speech[15, 10] and sequence data in domains such as molecular biology[8, 4, 3]. A crucial issue in modeling time-series data using HMM, is the problem of determining the appropriate model architecture: the number of states and the links between the states. Model complexity and predictive performance are highly related. While increased complexity gives better fits on the training data, it usually results in poorer predictive performance on new cases. Determining the optimal model complexity is crucial for good performance on new cases. For HMM's, this task is done manually by trained experts. Once the configuration is decided, several HMM training procedures are available that optimize model parameters while keeping the model configuration fixed [11, 15]. In domains where human expertise for determining model configuration is unavailable, efficient use of HMM's is difficult. This knowledge acquisition bottleneck hinders
the application of HMM technology. Automatic methods for determining model configuration would be of considerable assistance in applying HMM's in new domains.

In this paper we present a new procedure that addresses the problem of automatically configuring an HMM to the right complexity fit. It starts with a large, possibly over-fitted, HMM and attempts to iteratively prune it down to the appropriate complexity fit. The parameter re-estimates in our procedure can be formally derived and it can be shown that the procedure optimizes the maximum likelihood. Other features of our procedure include: (1) better convergence characteristics than the standard Baum-Welch algorithm, (2) automatic reduction of model size to the right complexity fit, (3) better generalization, and (4) relative insensitivity to the initial model size.

We consider only sequences that consist of discrete symbols. Our pruning strategy focuses on single link deletions. We always start with an ergodic (i.e. fully-connected network) model, although our procedure is not limited by this condition.

2 Hidden Markov Models

Hidden Markov Models (HMM) are a special class of Markov models that exhibit two levels of stochastic behavior, such that one level is not directly observable (hence the word "hidden") but can only be inferred through another stochastic process that produces a sequence of observable symbols. Several well-written introductions to HMM's, such as [10, 15], are available in existing literature, which emphasize their applications to speech recognition, a domain where HMM's have proven quite successful.

A HMM consists of a set of states which are inter-connected by links. In each state, it is possible to make a probabilistic transition to another state via a link between the two states. As a result of the transition, an output symbol is produced according to a probability distribution associated with the link. The sequence of output symbols is directly observed and represents the observable stochastic process. The underlying sequence of state transitions is not directly observed, but can only be inferred through the sequence of output symbols. Given a HMM and a sequence of symbols, a standard procedure based on dynamic programming concepts is available to determine the probability that the sequence was produced by the HMM. For details, the reader may refer to [15]. Our interest is in methods for training HMM's and this shall be the focus of this paper.

The best known training procedure for HMM's is the Baum-Welch Algorithm, an iterative algorithm that re-estimates model parameters so as to maximize the probability of generating the time-series [11]. The algorithm has a solid theoretical foundation and many analytical results as to its optimality are available. A major restriction is that it requires the model configuration to be fixed throughout the training process. This means that model configuration must be determined beforehand – usually based on human expertise. We use the procedure as a starting point of our method described in Sect. 3. Unlike the standard Baum-Welch, we allow the number of model parameters to be reduced in each iteration.