Second-Order Belief Hidden Markov Models

Jungyeul Park\textsuperscript{1}, Mouna Chebbah\textsuperscript{1,2}, Siwar Jendoubi\textsuperscript{1,2}, and Arnaud Martin\textsuperscript{1}

\textsuperscript{1} UMR 6074 IRISA, Université de Rennes1, Lannion, France
\textsuperscript{2} LARODEC Laboratory, University of Tunis, ISG Tunis, Tunisia

\{jungyeul.park,mouna.chebbah,arnaud.martin\}@univ-rennes1.fr, jendoubi.siwar@etudiant.univ-rennes1.fr

Abstract. Hidden Markov Models (HMMs) are learning methods for pattern recognition. The probabilistic HMMs have been one of the most used techniques based on the Bayesian model. First-order probabilistic HMMs were adapted to the theory of belief functions such that Bayesian probabilities were replaced with mass functions. In this paper, we present a second-order Hidden Markov Model using belief functions. Previous works in belief HMMs have been focused on the first-order HMMs. We extend them to the second-order model.

Keywords: Belief functions, Dempster-Shafer theory, first-order belief HMM, second-order belief HMM, probabilistic HMM.

1 Introduction

A Hidden Markov Model (HMM) is one of the most important statistical models in machine learning [15]. A HMM is a classifier or labeler that can assign label or class to each unit in a sequence [10]. It has been successfully utilized over several decades in many applications for processing text and speech such as Part-of-Speech (POS) tagging [11], named entity recognition [29] and speech recognition [7]. However, such works in the early part of the period are mainly based on first-order HMMs. As a matter of fact, the assumption in the first-order HMM, where the state transition and output observation depend only on one previous state, does not exactly match with the real applications [13]. Therefore, they require a number of sophistications. For example, even though the first-order HMM for POS tagging in early 1990s performs reasonably well, it captures a more limited amount of the contextual information than is available [27]. As consequence, most modern statistical POS taggers use a second-order model [3].

Uncertainty theories can be integrated in statistical models such as HMMs: The probability theory has been used to classify units in a sequence with the Bayesian model. Then, the theory of belief functions is employed to this statistical model because the fusion proposed in this theory simplifies computations of \textit{a posteriori} distributions of hidden data in Markov models. This theory can provide rules to combine evidences from different sources to reach a certain level of belief [21,28,24,4,23]. Belief HMMs introduced in [6,12,14,16,25,19,28,18], use
combination rules proposed in the framework of the theory of belief functions. This paper is an extension of previous ideas for second-order belief HMMs. For the current work, we focus on explaining a second-order model. However, the proposed method can be easily extended to higher-order models.

This paper is organized as follows: In Sections 2 and 3 we detail probabilistic HMMs for the problem of POS tagging where HMMs have been widely used. Then, we describe the first-order belief HMM in Section 4. Finally, before concluding, we propose the second-order belief HMM.

2 First-Order Probabilistic HMMs

POS tagging is a task of finding the most probable estimated sequence of \( n \) tags given the observation sequence of \( v \) words. According to [15], a first-order probabilistic HMM can be characterized as follows:

- \( N \): The number of states in a model \( S_t = \{ s_{t1}, s_{t2}, \ldots, s_{tN} \} \).
- \( M \): The number of distinct observation symbols. \( V = \{ v_1, v_2, \ldots, v_M \} \).
- \( A = \{ a_{ij} \} \): The set of \( N \) transition probability distributions.
- \( B = \{ b_j(o_t) \} \): The observation probability distributions in state \( j \).
- \( \pi = \{ \pi_i \} \): The initial probability distribution.

Figure 1 illustrates the first-order probabilistic HMM allowing to estimate the probability of the sequence \( s_{t-1} \) and \( s_t \) where \( a_{ij} \) is the transition probability from \( s_{t-1} \) to \( s_t \) and \( b_j(o_t) \) is the observation probability on the state \( s_t \). Regarding POS tagging, the number of possible POS tags that are hidden states \( S_t \) of the HMM is \( N \). The number of words in the lexicons \( V \) is \( M \). The transition probability \( a_{ij} \) is the probability that the model moves from one tag \( s_{t-1} \) to another tag \( s_t \). This probability can be estimated using a training data set in supervised learning for the HMM. The probability of a current POS tag appearing in the first-order HMM depends only on the previous tag. In general, first-order probabilistic HMMs should be characterized by three fundamental problems as follows [15]:

- Likelihood: Given a set of transition probability distributions \( A \), an observation sequence \( O = o_1, o_2, \ldots, o_T \) and its observation probability distribution \( B \), how do we determine the likelihood \( P(O|A,B) \)? The first-order model relies on only one observation where \( b_j(o_t) = P(o_t|s^t_j) \) and the transition probability based on one previous tag where \( a_{ij} = P(s^t_j|s^t_{i-1}) \). Using the forward path probability, the likelihood \( \alpha_t(j) \) of a given state \( s^t_j \) can be computed by using the likelihood \( \alpha_{t-1}(i) \) of the previous state \( s^t_{i-1} \) as described below:

\[
\alpha_t(j) = \sum_i \alpha_{t-1}(i)a_{ij}b_j(o_t)
\]