Hidden Markov Models vs. Maximum Entropy Markov Models

Ivan Oropeza
Dept. of Computer Science
University of Texas at Austin

William Xie
Dept. of Computer Science
University of Texas at Austin

1. INTRODUCTION

A frequent problem in many disciplines is the challenge to do sequence labelling. DNA sequencing, video semantic analysis, and Part-Of-Speech tagging are just some examples where sequence labelling is a crucial tasks that needs to be solved [1, 4, 2]. In Natural Language Processing (NLP), Part-Of-Speech (POS) tagging is an stepping stone into solving more complex problems such as syntactic parsing of sentences. The typical techniques used to solve this problem are Hidden Markov Models (HMM), Maximum Entropy Markov Models (MEMM), and Conditional Random Fields (CRF). While CRFs are considered to be the state-of-the-art in POS tagging we want to compare the performance of the other models HMM and MEMM. This paper is structured in the following way: first we provide a brief description of the datasets used in this investigation. Then, we compare HMMs and MEMMs. Next, we discuss how the training process works for POS tagging. Finally, we compare the performance of HMMs vs MEMMs under similar situations.

2. DATASETS

3. HMM VS. MEMM

A HMM is generative model for the joint distribution of states and observations. It follows a Markovian assumptions. The Markov assumption restricts the transitions between states to be dependent only on the immediate past [3]. On the other hand, MEMM is discriminative since it models the conditional probability of the state given the observation.

An MEMM is an enhancement on the Maximum Entropy model, also known as a multinomial regression model which in turn attempts to do classification by making the fewest number of assumptions. Figure 1 shows the pictorial differences between both graphical models. The added benefit of MEMMs over HMMs is that MEMMs are not limited to only modelling two aspects: $P(S_i|P_{i-1})$ and $P(O_i|S_i)$. Instead, MEMMs consider information derived directly from the observation through the use of features applied on the observation in addition to the previous state knowledge. For example, capitalization often occurs with Nouns, specific suffixes such as "ed" and "ing" tend to be associated with Verbs are some useful features that can not be modelled by an HMM but can be modelled by an MEMM. Moreover, the viterbi algorithm can be trivially modified to solve the decoding problem in an MEMM [3].

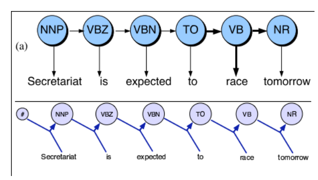

Figure 1: Hidden Markov Model(top) and Maximum Entropy Markov Model(bottom) [3]

4. REFERENCES

[1] W. Ching, E. Fung, and M. Ng. Higher-order hidden markov models with applications to dna sequences. In J. Liu, Y.-m. Cheung, and H. Yin, editors, *Intelligent Data Engineering and Automated Learning*, volume 2690 of *Lecture Notes in Computer Science*, pages 535–539. Springer Berlin Heidelberg, 2003.

[2] F. Hasan, N. UzZaman, and M. Khan. Comparison of different pos tagging techniques (n-gram, hmm and brillâĂŹs tagger) for bangla. In K. Elleithy, editor, *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, pages 121–126. Springer Netherlands, 2007.

[3] D. Jurafsky and J. H. Martin. *Speech and Language Processing (2Nd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2009.

[4] G. Xu, Y.-F. Ma, H.-J. Zhang, and S.-Q. Yang. An hmm-based framework for video semantic analysis. *Circuits and Systems for Video Technology, IEEE Transactions on*, 15(11):1422–1433, Nov 2005.