



Profile Hidden Markov Models

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics (Oxford, England)*, 14(9), 755-763.

Presenter Alim Gokkaya



Outline

1. Introduction
2. HMM Overview
3. What is a “profile”
4. Building Profile HMMs
5. Aligning Sequences with Profile HMMs
6. Conclusions
7. Discussion



Introduction

In this paper;

- A very influential review of the profile HMM algorithms
- Substantial background on the methods used

has been described.

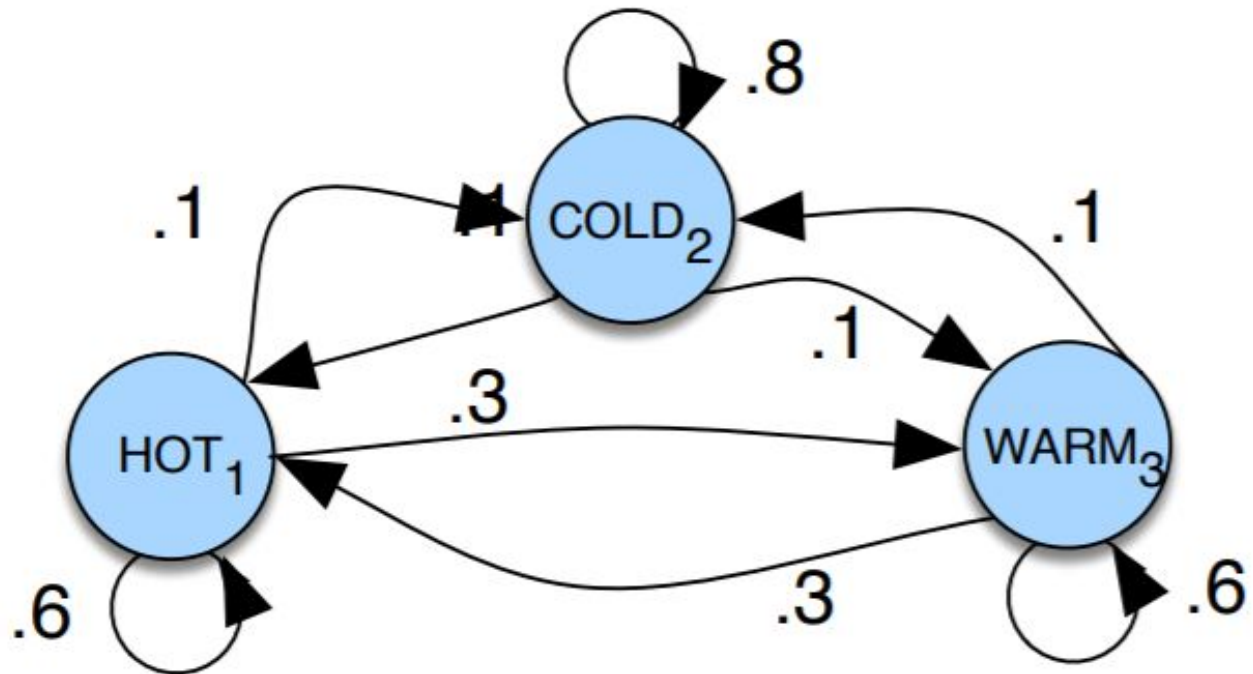


Introduction

Profile Hidden Markov Model

- Is sequence learning method from bioinformatics domain
- Is often preferred over pairwise alignment methods
- Has been very influential in the last decades
 - Approximate pattern recognition, e.g., finding genes
 - Error correction

Introduction





HMM Overview

The hidden states represent the emission probabilities for different events.

A dynamic algorithm (Viterbi) can compute the best sequence of hidden states for a given query string in:

$O(NM^2)$ time,

Where N = length of query,

M = # of hidden states



HMM Overview

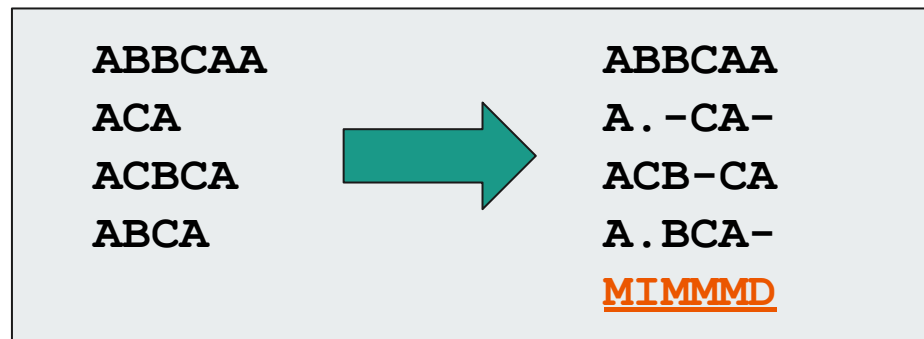
Similar methods, are already being used in morphological disambiguation:

- However these methods uses semi-supervised features for each word rather than letter-by-letter classification

What is a “profile”?

The consensus obtained from a family of related sequences via multiple-alignment.

Objective is to minimize the edit distance towards the consensus:





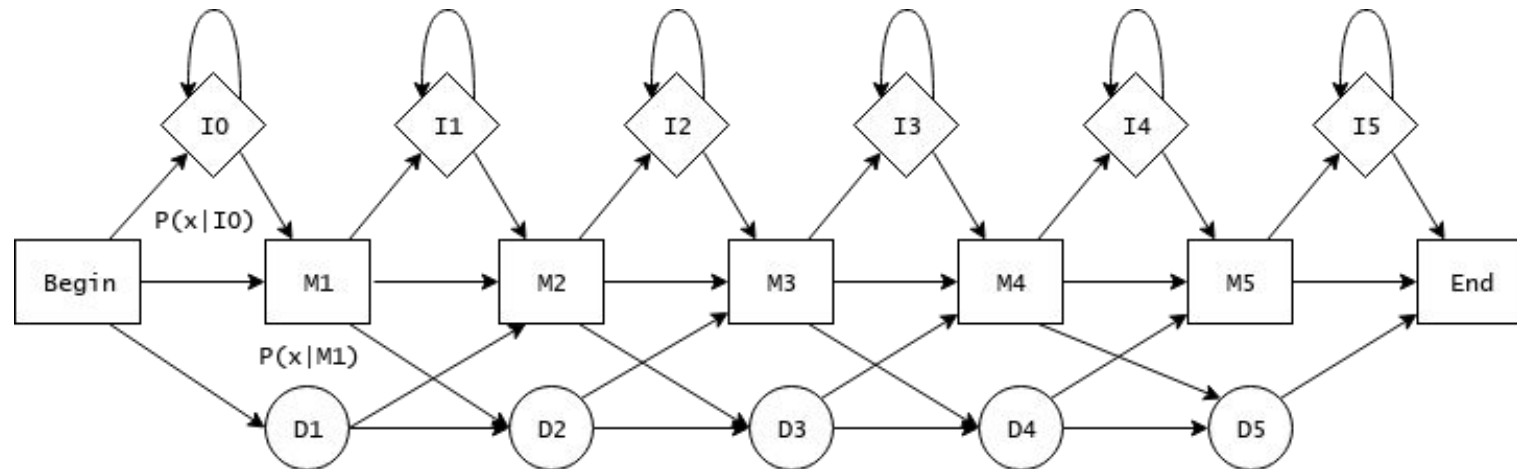
Profile HMM

The probabilities in a profile HMM is learned by:

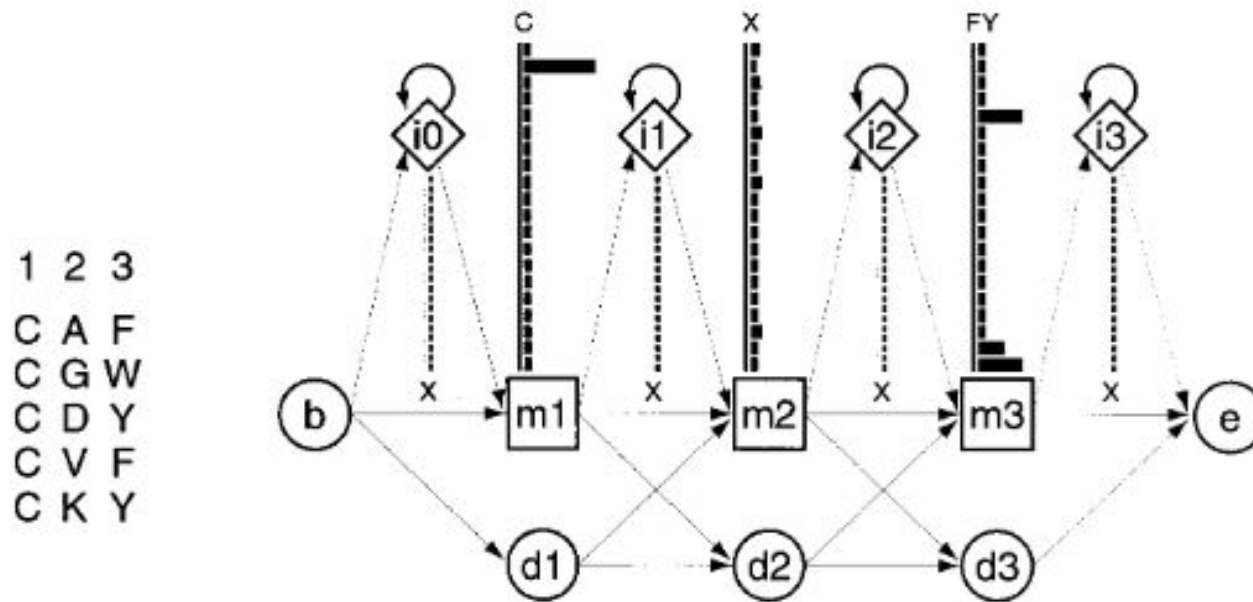
1. Aligning the sequences via an external algorithm
2. For each alignment column following hidden states are added
 - a. Insertion
 - b. Deletion
 - c. Match (or mismatch)

Profile HMM

Each insertion and match state holds emission probabilities for each character of the alphabet



An Example Profile HMM





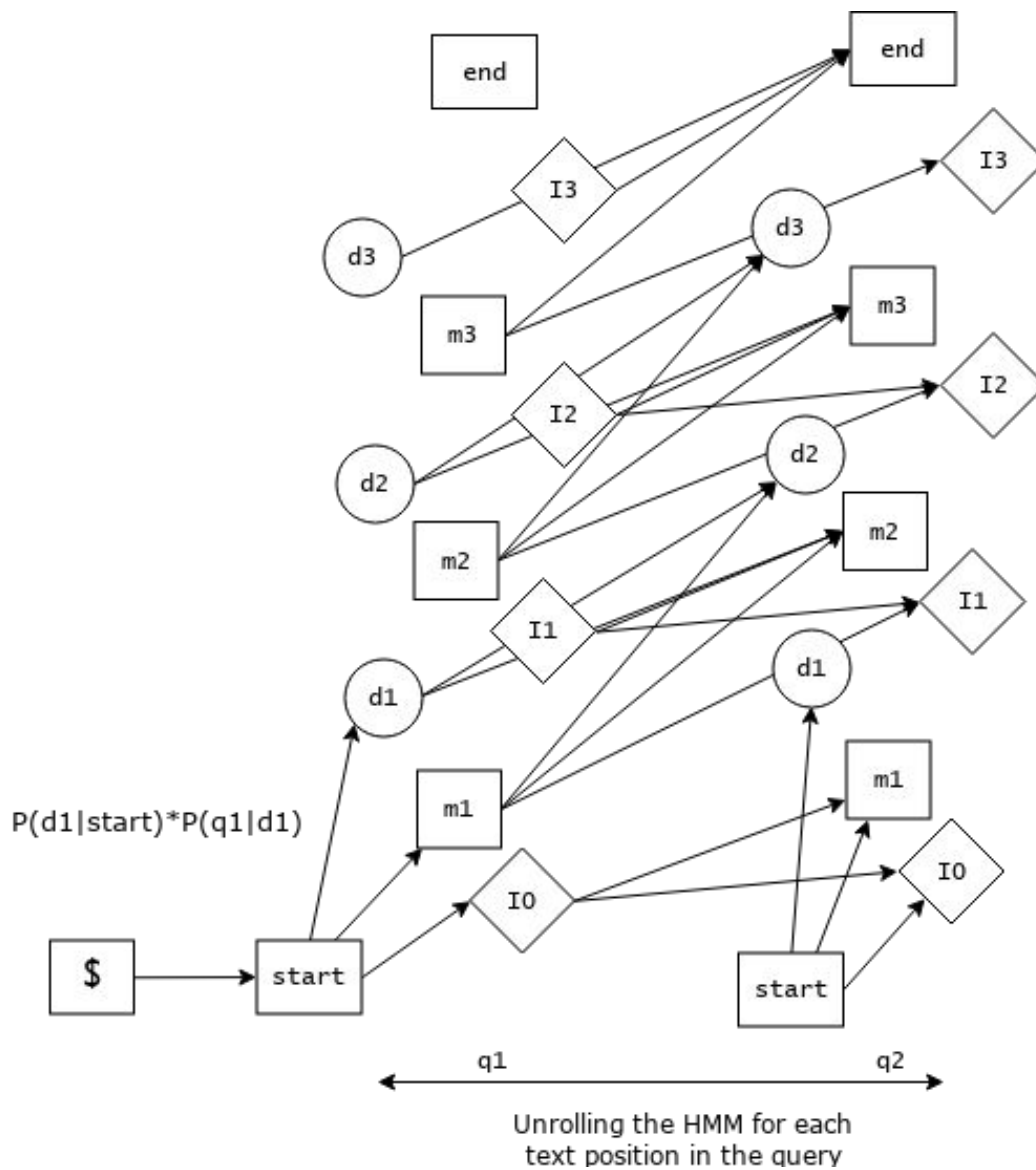
Training a Profile HMM

- Calculating emission probabilities is trivial if the multiple-sequence alignment is already given
- Otherwise model building is as challenging as computing a multiple-alignment
 - The optimal solution is known to be NP, i.e., $O(M^K)$
 - Commonly used heuristic algorithms to train profile HMMs include; expectation-maximization and gradient descent algorithms



Aligning Sequences Profile HMMs

- The Viterbi algorithm is used for alignment
 - Can be adapted for global vs. local alignment
 - Similar to Needleman/Wunsch
 - Affine gap penalties can be to the transitions from insertion->match nodes



Aligning Sequences Profile HMMs



Conclusions

- The profile HMM models gave a way to
 - Recognize the textual patterns
 - In a statistically meaningful way
- There are neural network models to train and implement profile HMMs that can provide better results
 - HMMs have smaller attention span w.r.t. state of the art DL methods
 - Advanced training methods available to avoid local optima



Discussions

We investigate whether;

- The methods used in the Profile HMM algorithms can be adapted to
 - Emit different event signalling the type of each morpheme
 - Based on the given context of the sentence (e.g. stem of the previous word)



References

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics (Oxford, England)*, 14(9), 755-763.

Jurafsky D., Martin J. H. (2018). Chapter A. Hidden Markov Models. *Speech and Language Processing (3rd ed. draft)*, (pp. 4)