EXPECTATION MAXIMIZATION FOR ELECTION PREDICTION IN A HMM-LIKE PROBABILISTIC GRAPHICAL MODEL.

PATRICK VARIN AND ARJUN ALETTY

This paper is dedicated to our adviser, Paul Ruvolo.

ABSTRACT. In this paper we present a probabilistic graphical model (PGM) similar to the Hidden Markov Model in which state transitions are also affected by previous observations. We also develop the Expectation Maximization algorithm specific to this PGM, an analogue to the Baum-Welch algorithm for the HMM. Finally we present this new PGM in the context of of election prediction and we show the performance of the EM algorithm in a series of experiments.

1. INTRODUCTION

The Markov condition is often a good descriptor of a variety of dynamical models. More likely than not, however, we cannot gather complete information from the system in question, in these cases a Hidden Markov Model is often used to analyze the system. Often times these systems are parametrized by a series of unknowns. These unknown parameters can be inferred, however, using the Expectation Maximization algorithm, which chooses the set of parameters that maximizes the likelihood of the observed dataset a famous application of this is in speech recognition, in which a speaker makes a series of utterances, the observed dataset, with the intent of constructing a word, the unknown parameter. The specific implementation of Expectation Maximization for HMMs is called the Baum-Welch algorithm.

There is a set of problems, however, in which the observations can influence the transition to the next state. In an election, for example, popular opinion can be influenced by the outcome of a poll. If an election cycle is considered a series of events in discrete time, the true public opinion can be considered the latent variable that influences the outcome of polls, the observed variables. There are a variety of parameters involved in this model, for example certain polls may have more influence on public opinion and certain polls may have some inherent bias towards or against different candidates.

In this paper we will discuss the computational methods necessary to analyze the graphical model discussed above. We will also revisit this example of using Expectation Maximization to predict the outcome of an election based on poll data.

2. The Model

The hidden Markov Model is given by a series of hidden states $\mathbf{Z} = \{z_t\}$, each of which emits an observable variable $\mathbf{X} = \{x_t\}$. According to the Markov condition state transitions rely only on the previous state, and observations rely only on the immediate state. As a result the model can be described by three probability distributions, the initial state probability $P(z_0)$, the transition probability $P(z_t|z_{t-1})$ and the emission probability $P(x_t|z_t)$.

For problems in which there is no prior knowledge of the form of these probability distributions it is common to tabulate each of the probabilities in three matrices: $\pi = \{\pi_i\}$, $A = \{\alpha_{ij}\}$, $B = \{\beta_{ij}\}$. The initial state probability is given by $P(z_0 = i) = \pi_i$, the transition probability is given by $P(z_t = j|z_{t-1} = i) = \alpha_{ij}$ and the emission probability is given by $P(x_t = j|z_t = i) = \beta_{ij}$ The Expectation-Maximization algorithm, then attempts to compute each element in these tables. This specific implementation of Expectation-Maximization is called the Baum-Welch algorithm.

The model that we introduce here introduces a new dependency in the state transitions. In this new model these transitions rely not only on the previous state, but also the previous observation. This changes the form of the transition probability to $P(z_t|z_{t-1}, t_{x-1})$.

3. Expectation-Maximization

In some applications, in which the form of these probability distributions is unknown, it may make sense to parametrize the probability distributions in a similar manner to the HMM described above. In general, however, the structure of Expectation-Maximization is independent of the choice of parametrization and we will use the variable θ to describe a general set of parameters, e.g., the parametrization used in the Baum-Welch algorithm is $\theta = (\pi, A, B)$.

The Expectation-Maximization procedure is generally separated into two steps, the E-step and the M-step

The E-Step. The goal of the E-Step is to formulate a fuction, $Q(\theta, \theta^t)$, the computes the expected value of the log-likelihood of parameters with respect to the probabilities of the latent variables given the last best guess of the parameters.

$$Q(\theta, \theta^t) = \mathbf{E}_{\mathbf{Z}|\mathbf{X}, \theta^t} \left[\log P(\mathbf{X}, \mathbf{Z}|\theta)\right]$$

The M-Step. The purpose of the M-Step is to optimize the log-likelihood function formulated in the E-Step.

The following is an example of a proof.

Proof. Set $j(\nu) = \max(I \backslash a(\nu)) - 1$. Then we have

$$\sum_{i \notin a(\nu)} t_i \sim t_{j(\nu)+1} = \prod_{j=0}^{j(\nu)} (t_{j+1}/t_j).$$

Hence we have

$$
(3.1) \quad \prod_{\nu} \left(\sum_{i \notin a(\nu)} t_i \right)^{|a(\nu-1)|-|a(\nu)|} \sim \prod_{\nu} \prod_{j=0}^{j(\nu)} (t_{j+1}/t_j)^{|a(\nu-1)|-|a(\nu)|}
$$
$$
= \prod_{j \geq 0} (t_{j+1}/t_j)^{\sum_{j(\nu) \geq j}(|a(\nu-1)|-|a(\nu)|)}.
$$

By definition, we have $a(\nu(j)) \supset c(j)$. Hence, $|c(j)| = n - j$ implies (5.4). If $c(j) \notin a$, $a(\nu(j))c(j)$ and hence we have (5.5). $\qquad \square$

> This is an example of an 'extract'. The magnetization M_0 of the Ising model is related to the local state probability $P(a) : M_0 = P(1) - P(-1)$. The equivalences are shown in Table 1.

TABLE 1

	$-\infty$	$+\infty$
$f_+(x,k)$	$e^{\sqrt{-1}kx} + s_{12}(k)e^{-\sqrt{-1}kx}$	$s_{11}(k)e^{\sqrt{-1}kx}$
$f_-(x,k)$	$s_{22}(k)e^{-\sqrt{-1}kx}$	$e^{-\sqrt{-1}kx} + s_{21}(k)e^{\sqrt{-1}kx}$

Definition 3.1. This is an example of a 'definition' element. For $f \in A(X)$, we define

$$
(3.2) \quad \mathcal{Z}(f) = \{E \in Z[X] : f \text{ is } E^c\text{-regular}\}.
$$

Remark 3.2. This is an example of a 'remark' element. For $f \in A(X)$, we define

$$
(3.3) \quad \mathcal{Z}(f) = \{E \in Z[X] : f \text{ is } E^c\text{-regular}\}.
$$

Example 3.3. This is an example of an 'example' element. For $f \in A(X)$, we define

$$
(3.4) \quad \mathcal{Z}(f) = \{E \in Z[X] : f \text{ is } E^c\text{-regular}\}.
$$

Exercise 3.4. This is an example of the `xca` environment. This environment is used for exercises which occur within a section.

The following is an example of a numbered list.

(1) First item. In the case where in G there is a sequence of subgroups

$$
G = G_0, G_1, G_2, \ldots, G_k = e
$$

such that each is an invariant subgroup of G_i.

(2) Second item. Its action on an arbitrary element $X = \lambda^\alpha X_\alpha$ has the form

$$
(3.5) \quad [e^\alpha X_\alpha, X] = e^\alpha \lambda^\beta [X_\alpha X_\beta] = e^\alpha c_{\alpha\beta}^\gamma \lambda^\beta X_\gamma,
$$

(a) First subitem.

$$
-2\psi_2(e) = c_{\alpha\gamma}^\delta c_{\beta\delta}^\gamma e^\alpha e^\beta.
$$

(b) Second subitem.

FIGURE 1. This is an example of a figure caption with text.

FIGURE 2

(i) First subsubitem. In the case where in G there is a sequence of subgroups

$$G = G_0, G_1, G_2, \ldots, G_k = e$$

such that each subgroup G_{i+1} is an invariant subgroup of G_i and each quotient group G_{i+1}/G_i is abelian, the group G is called *solvable.*

(ii) Second subsubitem.
(c) Third subitem.
(3) Third item.

Here is an example of a cite. See [1].

Theorem 3.5. *This is an example of a theorem.*

Theorem 3.6 (Marcus Theorem)**.** *This is an example of a theorem with a parenthetical note in the heading.*

4. SOME MORE LIST TYPES

This is an example of a bulleted list.

- $\mathcal{J}_g$ of dimension $3g - 3$;
- $\mathcal{E}_g^2 = \{$Pryms of double covers of $C = \square$ with normalization of C hyperelliptic of genus $g - 1\}$ of dimension $2g$;
- $\mathcal{E}_{1,g-1}^2 = \{$Pryms of double covers of $C = \square_{P^1}^H$ with H hyperelliptic of genus $g - 2\}$ of dimension $2g - 1$;
- $\mathcal{P}_{t,g-t}^2$ for $2 \leq t \leq g/2 = \{$Pryms of double covers of $C = \square_{C''}^{C'}$ with $g(C') = t - 1$ and $g(C'') = g - t - 1\}$ of dimension $3g - 4$.

This is an example of a 'description' list.

Zero case: $\rho(\Phi) = \{0\}$.

Rational case: $\rho(\Phi) \neq \{0\}$ and $\rho(\Phi)$ is contained in a line through 0 with rational slope.

Irrational case: $\rho(\Phi) \neq \{0\}$ and $\rho(\Phi)$ is contained in a line through 0 with irrational slope.

References

1. T. Aoki, *Calcul exponentiel des opérateurs microdifferentiels d'ordre infini.* I, Ann. Inst. Fourier (Grenoble) **33** (1983), 227–250.
2. R. Brown, *On a conjecture of Dirichlet*, Amer. Math. Soc., Providence, RI, 1993.
3. R. A. DeVore, *Approximation of functions*, Proc. Sympos. Appl. Math., vol. 36, Amer. Math. Soc., Providence, RI, 1986, pp. 34–56.

OLIN WAY, OLIN COLLEGE, NEEDHAM, MASSACHUSETTS 02492
E-mail address: `patrick.varin@students.olin.edu`

OLIN WAY, OLIN COLLEGE, NEEDHAM, MASSACHUSETTS 02492
E-mail address: `arjun.aletty@students.olin.edu`