LEARNING IN MARKOV MODELS USING THE IMPRECISE DIRICHLET MODEL

Erik Quaeghebeur, SYSTeMS, GHENT UNIVERSITY

Objectives

1. develop a method for learning the transition probabilities in (hidden) Markov models using imprecise probabilities
2. application of this method to real-life problems

Markov models (MM)

Considered systems Stochastic processes in discrete time with a finite number of states.

Transitions The (unknown) transition probabilities θ_{ij} between states i and j have the Markov property, i.e. they only depend on i and j. They are grouped in the transition probability matrix Θ where the rows correspond to initial states and the columns to destination states.

Observations A state sequence $x_1 x_2 \cdots x_n$ is observed. The observation matrix N consists of components n_{ij}, the number of observed transitions $i \to j$.

Hidden Markov models (HMM) A MM with unobservable states that generate a random output. The probability for a state i to produce y is ξ_{iy} and Ξ is the corresponding output matrix.

The imprecise Dirichlet model (IDM)

Precise Dirichlet model A model for statistical inference, that uses

- a Dirichlet probability density function $f_\alpha(\Theta)$ to express a prior assessment about Θ (α is a parameter),
- a multinomial likelihood function $L(N|\Theta)$ that expresses the likelihood of N (each row is considered to be an independent multinomial sample) given Θ,
- Bayes' rule, to calculate a posterior density $f_\alpha(\Theta|N) = f_\alpha(\Theta)L(N|\Theta)$.

The posterior density can be used to obtain an estimation $\hat{\Theta}$ of Θ.

Imprecise Dirichlet model An extension of the precise model that uses a class $\{f_\alpha|\alpha \in \mathcal{A}\}$ of density functions. Using imprecise probability theory, we can obtain lower - and upper transition probability matrices $\underline{\Theta}$ and $\overline{\Theta}$ as an estimation for Θ.

RESULTS AND QUESTIONS

Initial results Theory and simulation have shown that

- $\underline{\Theta} < \hat{\Theta} < \overline{\Theta}$,
- the imprecision $\overline{\Theta} - \underline{\Theta}$ is a row property: it remains large for transient states and either remains 1 or converges to 0 for states in an absorbing class.

Immediate question Can the rows of N be considered to be independent multinomial samples (there is a relation between the row-sum and the column-sum for a state)?

HMM questions To use the IDM we need to know

- how to represent the observations of outputs,
- how to incorporate conditional probabilities (transition probabilities must be inferred from output observations using Ξ).

MM APPLICATION:
PRE-FETCHING OF WEB PAGES

Goal

Given the previous history of web pages visited by a user, the server must decide, after each transition to a page, what new pages to pre-fetch.

Formulation in our context

The previous history consists of

1. the previous sessions of the user, i.e. sequences of pages, which are considered realizations N of a MM with transition probability matrix Θ,
2. the web pages visited in the current session, which is a partial realization of the same MM.

Additional modeling

Pre-fetching is a decision problem that uses the class of posterior densities in concurrence with a utility function which describes the cost of pre-fetching a page and the loss associated when it is eventually not requested by the user.

HMM Application:
Aligning of Gene Sequences

Goal

Given two gene sequences, we want to know the most likely alignments and the probabilities associated with these alignments.

Formulation in our context

We consider the aligned pair of sequences to be the realization N of a HMM (with Θ and Ξ) which generates sequences of $\{\mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{T}, -\}$ pairs, where $-$ corresponds to a gap.

Additional modeling

To be useful, we will have to take into account

- begin - and end states,
- the quality of our learning set.