Learning imprecise hidden Markov models

Arthur Van Camp, Gert de Cooman, Jasper De Bock, Erik Quaeghebeur, Filip Hermans
SYSTeMS research group, Ghent University, Belgium


Precise hidden Markov model

Consider a stationary precise hidden Markov model (HMM) with $2n$ variables: $n$ hidden states $X_i$, taking values in a set $\{1, \ldots, n\}$ and $n$ observations $Y_i$, taking values in $\{1, \ldots, n\}$. Both the marginal model $P_0(Q_1)$ and the transition models $P_{X_k|X_{k-1}}(x_k|x_{k-1})$ and the emission models $P_{Y_k|X_k}(y_k|x_k)$ are known.

Baum–Welch algorithm

Given the observation sequence $(O_1, O_2, \ldots, O_n)$, we can use the Baum–Welch algorithm to obtain a maximum-likelihood estimate of these local models. With this algorithm, the likelihood of the observation sequence converges to a local maximum, but it is not guaranteed that we find the global maximum.

Learning imprecise HMMs

Imprecise hidden Markov model

An imprecise hidden Markov model (iHMM) has the same graphical model but the local models are imprecise.

Using Baum–Welch

With the classical Baum–Welch algorithm, we obtain precise local models. We present a method for learning imprecise transition models in an iHMM. We use the expected number of transitions, obtained by the Baum–Welch algorithm after sufficient iterations, to construct imprecise transition models.

Imprecision

The lower and upper probabilities have the following property: the imprecision increases by increasing number of states $n$.

$$Q(i|j) - Q(i|j) = \frac{s_{i,j}}{s_{i,j}}$$

Here $n_i = \sum n_{i,j}$, the expected number of times that state $i$ occurs in the $n - 1$ variables $X_1, \ldots, X_{n-1}$.

The higher and lower probabilities have the following property: the imprecision increases by increasing number of states $n$.

$$Q(i|j) - Q(i|j) = \frac{s_{i,j}}{s_{i,j}}$$

The lower and upper probabilities have the following property: the imprecision increases by increasing number of states $n$.

$$Q(i|j) - Q(i|j) = \frac{s_{i,j}}{s_{i,j}}$$

We apply our method to the following problem: based on counted number of annual earthquakes in 107 subsequent years (1900 – 2006), we are interested in predicting the earthquake rate in future years.

We assume that

- the earth can be in 3 different seismic states $\lambda_1, \lambda_2,$ and $\lambda_3$;
- in each state, the emission of earthquakes is a Poisson process: $P_0(o|\lambda) = e^{-\lambda} \frac{\lambda^o}{o!}$.

Transition model

The credal sets in the upper eight simplices on the right represent, for different values of the pseudocounts $\lambda$, the transition models. The grey credal set represents $Q(\lambda)$, the blue credal set represents $Q(\lambda)$, and the red credal set represents $Q(\lambda)$.

Prediction

With the transition models learned with our method, we predicted the earthquake rate in the years 2007, 2016, 2026 and 2036. We did this in two cases: the pseudocounts $\lambda = 2$ and $\lambda = 5$. The lower four simplices on the right show conservative approximations (the smallest hexagons with vertices parallel with the vertices of the simplices) for the credal sets representing the global model $\Sigma_{\lambda}(\cdot|\lambda)$, updated to the observation sequence. The grey credal set represents the updated global model with $\lambda = 2$ and the blue credal set represents the updated global model with $\lambda = 5$. As expected, the global model for $\lambda = 5$ include the global model for $\lambda = 2$.

Dilation

Learning imprecise probability models in an iHMM, like our method does, is necessary before being able to make inferences from such a model, e.g., with the MePCTr2 algorithm. Dilation appears here as the increase of the imprecision of the inferences when the target node $X_0$ goes to the first state $X_1$. The interpretation of this phenomenon is not yet clear. We did some experiments to estimate the dilation in an iHMM with $n = 50$. The 30 simplices represent conservative approximations of the updated global model $\Sigma_{\lambda}(\cdot|\lambda)$. The red dots indicate the observations. The local models are linear-vacuum mixtures, of which the precise components are given in the table above.

| $\lambda$ | $p_{15}(\cdot|\lambda)$ | $p_{16}(\cdot|\lambda)$ | $p_{17}(\cdot|\lambda)$ | $p_{18}(\cdot|\lambda)$ | $p_{19}(\cdot|\lambda)$ | $p_{20}(\cdot|\lambda)$ | $p_{21}(\cdot|\lambda)$ | $p_{22}(\cdot|\lambda)$ | $p_{23}(\cdot|\lambda)$ | $p_{24}(\cdot|\lambda)$ | $p_{25}(\cdot|\lambda)$ | $p_{26}(\cdot|\lambda)$ | $p_{27}(\cdot|\lambda)$ | $p_{28}(\cdot|\lambda)$ | $p_{29}(\cdot|\lambda)$ | $p_{30}(\cdot|\lambda)$ |
|-------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 0.0   | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             |
| 0.3   | $\mathbf{0.8}$ | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.8             |
| 0.4   | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.1             | 0.8             |

The authors acknowledge financial support from the Research Foundation – Flanders (FWO) and the Ghent University. The first author has been supported by the Postdoctoral Research Grant (G050314N).