State sequence prediction in imprecise hidden Markov models

Jasper De Bock

Department of Electrical Energy, Systems and Automation, Ghent University

Supervisor(s): Gert De Cooman

Consider the following problem: you have a sequence of data you would like to know, but cannot directly observe. For example: you are interested in the weather for each day of some week in the past and you do not have access to this data. However, you do have access to a different sequence of data that is related to the first one. You might know how many people visited the Belgian shore each day. The objective is to use the second sequence to try and find the first one.

Of course there is no one-to-one correspondence between these two sequences of data, but there's definitely a connection, which can be expressed by means of probabilities. If the weather is nice, chances are many people will visit the shore. If today is a very warm day, it's unlikely that tomorrow will be extremely cold.

All these probabilities can be bundled into a hidden Markov model (HMM). Using this HMM and the information that we have about the visits to the shore, one can calculate which weather has the highest probability of having occurred during that week and use that as an estimation of the actual weather. An efficient way to do this calculation is by using the Viterbi algorithm, which is used in various fields of application to solve the kind of problems described above. Some examples are speech recognition, bio-informatics, optical character recognition and crypto-analysis.

However, the Viterbi algorithm and HMMs fail to incorporate robustness into their results. If you would alter the probabilities used in the calculations by just a slight amount and get totally different results, this indicates that the obtained results are not reliable and it would be advisable to gather more data. Unfortunately, the Viterbi algorithm does not detect such situations.

In order to avoid this, one can use the theory of imprecise probabilities. This theory essentially allows you to consider not just one probability, but a range of probabilities and provides ways to do calculations with them. If you have access to huge amounts of data, you might be confident enough to say that, for example, the probability that it is sunny today is exactly 40% if it was sunny yesterday. However, if this is based on only a couple of observations, it might be wise to be more cautious and say that the probability lies between 30% and 50%.

If we now describe our knowledge about the weather and the shore visits of people in terms of imprecise probabilities and bundle all this information, we obtain an imprecise HMM and we can again try and solve the problem mentioned above. Unlike the precise case, the mathematics gets complicated and the algorithmic complexity of solving the problem tends to explode. Up till now, no efficient solution existed. The work I present here at the symposium is a solution to this problem. I present an algorithm that solves the problem in a manner that is both exact and efficient, thereby generalizing the existing Viterbi algorithm to allow for imprecision.

Jasper De Bock is a research fellow of the Research Foundation - Flanders (FWO - Vlaanderen)