Final Project Proposal: Implementation of Hidden Markov Model in Weka

Alexandre Savard

February 17, 2006

Machine learning algorithms are definitly very powerfull tools designed to acheive many different task over large amount of datas. Weka is an open source project that aims to make those techniques generally available as well as to develop new machine learning algorithms. The area of concern of these tools is wide and regroups a large amount of different discipline.

Weka include a large number of well known learning algorithms as we think about neural network, logistic model trees, document profiling, bayesian network. However, it seems still to lack the Hidden Markov Model, which is one of the most popular machine learning algorithm. A reason for this absence is mainly a difficulty to pre-process uniformly the datas that are going to be analized by the model.

This project aims to implement Hidden Markov Model in weka. There already exists many different implementation of it in different programming languages. A first task will be to find an appropriate and efficient one to be used or to be translated in JAVA programming language. A second step will be to program a basic implementation according to the weka's uniform interface for a specefic type of data and test the results over an existing database. Further steps should be how the datas can be generally processeed in order to handle different type of databases. It is sure that we will keep in mind the specificity of musical information troughout this project.

Bibliography

Frank, E., Hall, M. A., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H. and Trigg, L. "In Data Mining and Knowledge Discovery Handbook", Chapter 8, Springer-Verlag, 2005.

Dugad R. and Desai U. "A tutorial on Hidden Markov Models", Signal Processing and Artificial Neural networks Laboratory, technical report no.: SPANN-96.1, Bombay, 1996.