Learning and Adaptivity - Project Proposal
Time Series Prediction using Hidden Markov Models

Saugata Biswas, Ashok Meenakshi Sundaram

May 12, 2014

Introduction

The primary objective of this project is to learn to predict non linear time series. From the perspective of service or medical robotics which need to help people, the real life data like speech, activity, bio-informatics all require time series analysis to predict the future based on the past so that a control decision could be made. Also, in the other domains like meteorology, finance, marketing, web analysis etc. time series analysis is important. However there is no deterministic relation between the past and future. The variation of data are non linear. An effective way to handle this non-deterministic data could be to formulate in terms of probability. This project will focus on one such stochastic machine learning approach called Hidden Markov Model. The use case considered for this project is prediction of stock market based on its performance in the past.

Methods

Hidden Markov Models (HMM) are to be used for training time series prediction model. HMMs are markov process with hidden states i.e. the states are not directly visible but only the output dependent on this hidden state is visible. HMMs are chosen since they are stochastic process and more meaningful to handle non-deterministic data. Other methods to predict time series include multi layer perceptron, auto regressive model, Box-Jenkins approach.

Hidden Markov Model tools available in scikit-learn [1] and weka [2] libraries will be used to complete this project. Future extensions of this project could be done in the directions of adapting HMM to new data as they come and not restricting to the old training set. Also, issues on balancing prediction based on long term series trend and short term series trend could be investigated further for better performance.

Data Set and Features

Data set for the stock market prediction will be taken from Istanbul stock exchange which is available at `http://archive.ics.uci.edu/ml/datasets/ISTANBUL+STOCK+EXCHANGE`. This data set contains indexes of Istanbul stock exchange in comparison with other international indexes e.g.

[1] http://scikit-learn.org/

[2] http://www.cs.waikato.ac.nz/ml/weka/

SP, DAX, FTSE, NIKKEI etc. for a period of two years (2009-2011). This is a labeled data set with 538 entries. 80 percent of the data will be taken as training set and the rest will be used as test set.

Features i.e. other international indexes are already directly available in the data set but pre-processing of the data will be necessary. New additional features can be added by taking different statistical measures from the available measures. Python has implementation of such statistical tools required to collect those features.

Milestones

First week

- Pre-processing of data
- Selection of features and generating new features based on those already available
- Separation of data into training, validation and test set
- Learn theory behind HMM

Second and Third week

- Training HMM using the training set
- Validation of the model using cross validation set

Fourth week

- Finally testing the learned model using the test set
- Compare those result with those already found in research paper

References

- Akbilgic, Oguz, Hamparsum Bozdogan, and M. Erdal Balaban. "A novel Hybrid RBF Neural Networks model as a forecaster." Statistics and Computing (2013): 1-11.
- Zhang, Yingjian. Prediction of financial time series with Hidden Markov Models. Diss. Simon Fraser University, 2004.
- Kavitha, G., A. Udhayakumar, and D. Nagarajan. "Stock Market Trend Analysis Using Hidden Markov Models." arXiv preprint arXiv:1311.4771 (2013).
- Hassan, Md Rafiul, Baikunth Nath, and Michael Kirley. "A fusion model of HMM, ANN and GA for stock market forecasting." Expert Systems with Applications 33.1 (2007): 171-180.
- Nobakht, Behrooz, Carl-Edward Joseph, and Babak Loni. "Stock market analysis and prediction using hidden markov models." Student Conference on Engg and Systems (SCES). 2012.
- Dietterich, Thomas G. "Machine learning for sequential data: A review." Structural, syntactic, and statistical pattern recognition. Springer Berlin Heidelberg, 2002. 15-30.