USING KERNEL PARTIAL LEAST SQUARES (K-PLS) REGRESSION FOR TIME-SERIES FORECASTING

By

Hui Wang

A Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: INDUSTRIAL AND SYSTEMS ENGINEERING (DECISION SCIENCES AND ENGINEERING SYSTEMS)

Approved by the
Examining Committee:

Mark J.Embrechts, Thesis Adviser

Thomas R. Willemain, Member

Thomas C. Sharkey, Member

Albert S. Paulson, Member

Rensselaer Polytechnic Institute
Troy, New York

(For Graduation August 2011)

ABSTRACT

Time-series forecasting is being increasingly applied in numerous and diverse domains such as transportation and finance. Most conventional forecasting methods are based on linear models. However, linear models are sometimes insufficient to model the behavior of time-series such as electricity loading and financial data. Consequently, there has been a strong market need for an advanced time-series forecasting model that can be applied effectively to diverse types of time-series, including nonlinear time-series. Most of the work in the past decade has focused on extending Support Vector Machines (SVMs) to nonlinear time-series forecasting. As the broader framework behind SVMs, kernel partial least squares (K-PLS) algorithm shows a great potential as a simpler and more computationally efficient time-series forecasting tool.

The K-PLS based method has been widely used in the field of biometrics and chemometrics, but it has not been considered for time-series analysis. This thesis proposes a novel K-PLS based time-series forecasting system $K-PLS_{subseries}$, which can be applied effectively to nonlinear time-series. The focus of this research is to address the practical challenges encountered when using K-PLS for time-series forecasting. This thesis develops a practical and efficient model parameter selection method and a robust forecasting framework. It also introduces one simple evaluation metric to measure prediction leading power. This work uses the M-3, NN3 and NN5 competition data to validate $K-PLS_{subseries}$ performance against other advanced time-series forecasting models. Recommendations on the suitable forecasting model between $K-PLS_{subseries}$ and benchmark models are made for different classes of benchmark time-series.

One of the main issues with K-PLS is parameter selection. It is computationally expensive to obtain the optimal set of K-PLS parameters and generally requires cross validation via a grid search. In order to resolve this issue, this thesis develops

an efficient K-PLS parameter selection method by using a reduced search space for K-PLS parameters and marginal effect of different K-PLS parameters on prediction accuracy.

$K - PLS_{subseries}$ also incorporates statistical techniques to make its forecasts general and robust. This thesis applies $K - PLS_{subseries}$ to the M-3, NN3 and NN5 competitions and shows that $K - PLS_{subseries}$ is competitive with most advanced forecasting models, outperforming most advanced models that participating in M-3 and NN3 competitions for the required forecasting horizon.