

Preface

The application of machine learning techniques to time-series prediction is a challenging research problem, particularly for the case of real-world problems which are often nonlinear and non-stationary. An important preprocessing step is to extract salient features from the raw data. Better features produce better predictions. This brief proposes a systematic method to discover such features, utilising context-free grammars to generate a large candidate set from which a subset is selected. Expert knowledge can be incorporated in the grammar, and the selection process avoids redundant and irrelevant features. The readers can use the techniques in this brief to a wide range of machine learning applications to represent complex feature dependencies explicitly when machine learning cannot achieve this by itself.

The utility of grammar-based feature selection is demonstrated using examples involving stock market indices, peak electricity load and foreign exchange client trade volume. The application of machine learning techniques to predict stock market indices and electricity load is not a recent development. Yet it continues to attract considerable attention due to the difficulty of the problem, which is compounded by the nonlinear and non-stationary nature of the time-series. In addition to these two popular applications, we have also demonstrated the application of the technique to foreign exchange client trade volume time-series (irregular time-series). This unique application particularly focuses on hedging the risk exposure of a bank by predicting client trade volumes and directions.

This brief was produced as a result of the research carried out at the Computer Engineering Laboratory (CEL), University of Sydney, Australia with the collaboration of Westpac Banking Corporation. Anthony Mihirana De Silva was funded by the University of Sydney International Scholarships (USydIS) scheme. We also gratefully acknowledge the support of the Australian Research Council's Linkage Projects funding scheme (Project number LP110200413) and Westpac Banking Corporation.

We are indebted to Dr. Barry Flower who has been a long-time companion of our research and a constant source of insights. Dr. Ahmed Pasha and Dr. Richard Davis provided valuable ideas and discussions regarding feature selection, and Farzad Noorian's help with crucial ideas and implementation has greatly improved the quality of this work.

The University of Sydney, December 2014

Anthony Mihirana De Silva
Philip H.W. Leong

Grammar-Based Feature Generation for Time-Series
Prediction

De Silva, A.M.; Leong, P.H.W.

2015, XI, 99 p. 28 illus., Softcover

ISBN: 978-981-287-410-8