Combining Spatial and Temporal Aspects of Prediction Problems to Improve Prediction Performance

William Groves
Department of Computer Science and Engineering
University of Minnesota
groves@cs.umn.edu

1 Introduction

Quantitative prediction problems involving both spatial and temporal components have appeared prominently in several disparate research areas including finance, supply chain management, and civil engineering. Unfortunately, either the spatial or temporal aspect tends to dominate the other in many prediction formulations. We briefly examine the underlying formulations used in spatial and temporal prediction. Then, we outline a method that combines these approaches and improves prediction results in high-dimensional economic domains by integrating multivariate and time series techniques which require minimal tuning but achieve superior performance compared to previous methods. We present preliminary results in the context of the Trading Agent Competition for Supply Chain Management.

2 Completed Work

Much of the work completed so far focuses on economic analysis and prediction techniques relevant to data derived from the Trading Agent Competition for Supply Chain Management (TAC SCM). TAC SCM is a multi-agent supply chain simulation involving an oligopoly of competing, fully autonomous agents who seek to maximize profit. In the simulation, agents purchase component parts from suppliers, construct computers with the component parts, and sell the computers to customers. This is a complex economic simulation competition with an annual tournament since 2003 that has attracted university teams from around the world.[1]

2.1 Understanding Complex Economic Data

Understanding the effect of changes in behavior in a dynamic, noisy economic environment is a challenge. Actions that in off-line analysis should improve agent behavior can have adverse unintended effects due to the inter-related nature of operating in a highly optimized and competitive multi-agent environment. Much of our early work focuses on understanding how low-level tactical decisions in TAC SCM combine into high-level strategic behaviors using visualization [Groves *et al.*, 2010; 2011].

Figure 1 illustrates a technique for visualizing supply and demand pressures for an individual day in the component

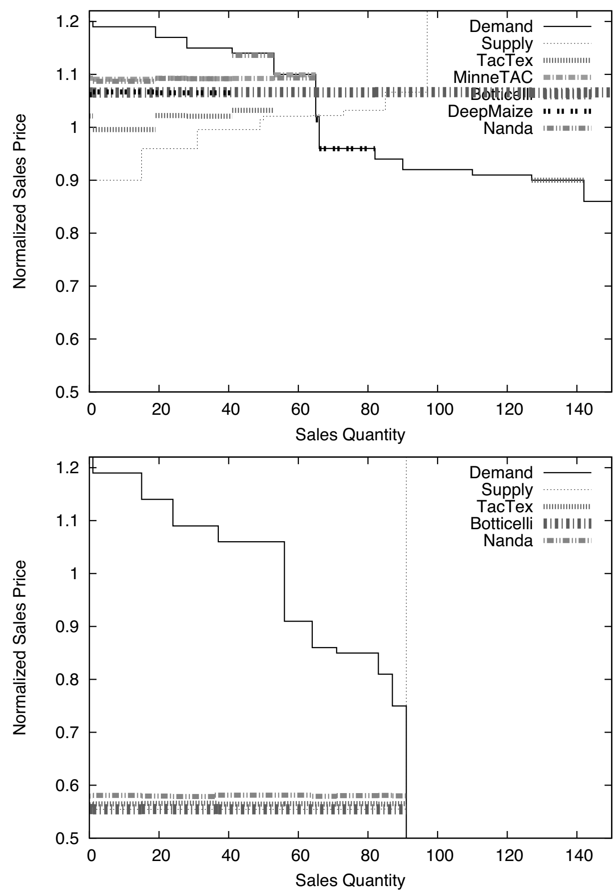

Figure 1: Customer market supply and demand for a product on day 4 (top) and day 89 (bottom) of game tac01-#3454. The demand line is the total demand. The dotted lines correspond to the offers made by each agent and the height of the line indicates the offer price. The agent with the lowest offer price for a given request will receive the order.

purchase market. The "Supply" line represents the ascending price-sorted set of offers made by the agents in response to customer requests. The total demand from customers is shown in the "Demand" line, sorted by decreasing reserve price. Offers from individual agents are visible as short hor-

[1] For more information on TAC SCM, please see http://tac.cs.umn.edu/ .

izontal lines set vertically at the price specified in the offer. The offer with the lowest price becomes the accepted order for each request. Using this graph it is possible to estimate the current market pressure based on how tightly the bids are compressed in terms of price (y-axis). Visualizations such as this are useful for formulating strategic and tactical behaviors in future runs of the simulation.

2.2 Improving prediction by leveraging temporal aspects of data

Economic data generally features a large number of simultaneously observable variables that contain potentially relevant information for prediction. Many types of multivariate algorithms can capture relationships between sets of variables and be used for prediction. [Martens and Næs, 1992] provides an excellent overview of multivariate techniques. There are also many types of time series prediction methods that perform prediction for temporal data.

We introduce a method that augments the feature vector used in multivariate techniques with time features and show that it improves prediction in high-dimensionality domains having a significant temporal aspect. Specifically, we augment the features fed to a partial least squares (PLS) regression model as follows: (1) we include time-delayed observations (that we call *lagged features*) as additional elements in the feature vector, in addition to the most recent observed value, and (2) we introduce a hierarchical segmentation of the features set. In our domain, the 101 raw data streams are divided into 5 classes based on the relationship of each to the target variable. The intuition we use is that earlier observations of the variable to be predicted are most likely to contain predictive information. Data streams on other similar products also provide some information. Finally, data streams from the remaining products are expected to contain the least useful information. The extent of lagged features is set uniformly within each feature class and is determined by a search process. Partitioning the raw features into classes in this way makes searching for a higher-performance combination of time-delayed features tractable and efficient.

	Future Product (p-value)	Future Component (p-value)
PLS regression with Lagged Features	**0.08553** (0.2660)	**0.09910** (0.0012)
PLS regression (no lagged features)	0.09039	0.10947
2008 First Place	0.08726	0.09964
2008 Third Place	0.09934	0.10281

Table 1: RMSE (Root Mean Square Error) of the aggregate prediction scores (lower is more accurate) of an agent implementing our method for each of the 20 day ahead ("future") predictions and the scores of the top performing agents in the 2008 TAC SCM Prediction Challenge.

Table 1 shows the results of PLS regression used both with and without the augmented feature vector. The table also compares results produced by our approach with the scores of other prediction methods used the TAC SCM domain. A p-value less than 0.1 denotes statistical significance at the 90% level using Student's paired t-test. Our method, which is domain-agnostic, achieves competitive and often superior performance compared to the state-of-the-art domain-specific prediction methods used by top agents in the 2008 TAC SCM Prediction Challenge competition. The Prediction Challenge is concerned only with price prediction aspects in TAC SCM.

3 Conclusions and Future Work

We have shown that augmenting multivariate techniques with temporal features facilitates prediction and is easy to generalize to many domains. This methodology can replace elaborate domain specific models in many applications.

For future work, we will seek to apply these prediction techniques in other domains. Our next steps are (1) to add spatial information and (2) to use the computed models to determine mathematical relationships between the variables. Many real-world prediction domains contain both temporal and spatial aspects, and this motivates the need to incorporate spatial relationships into the prediction model. In particular, we believe spatial relationships present in airline ticket price data (see [Etzioni *et al.*, 2003]) are amenable to our approach.

Multivariate techniques such as PLS regression implicitly compute vectors (PLS factors) from the training data to perform dimensionality reduction. These PLS factors, analogous to principal components in Principal Component Analysis, contain information about correlations within the augmented feature set. Analysis of the PLS factors will allow estimates of spatial and temporal relationships. These could serve to validate prediction models computed on data with known relationships. For instance, if the spatial relationships between locations in the air transportation network are estimated from the data, these computed relationships could be verified using known physical relationships. This could serve as a mechanism for prediction model validation.

References

[Etzioni *et al.*, 2003] Oren Etzioni, Rattapoom Tuchinda, Craig A. Knoblock, and Alexander Yates. To buy or not to buy: mining airfare data to minimize ticket purchase price. In *KDD*, pages 119–128, 2003.

[Groves *et al.*, 2010] W. Groves, W. Ketter, J. Collins, and M. Gini. Analyzing market interactions in a multi-agent supply chain environment. In R. Sharman, H. R. Rao, and T. S. Raghu, editors, *Exploring the Grand Challenges for Next Generation E-Business. 8th Workshop on E-Business, WEB 2009, Revised Selected Papers*, volume 52 of *Lecture Notes in Business Information Processing*, pages 44–58. Springer, 2010.

[Groves *et al.*, 2011] W. Groves, W. Ketter, J. Collins, and M. Gini. Analysis of Market Interactions and Decision Support in a Multiagent Supply Chain Network. *IEEE Trans on Automation Science and Engineering*, in review, 2011.

[Martens and Næs, 1992] Harald Martens and Tormod Næs. *Multivariate Calibration*. John Wiley & Sons, July 1992.