TimeFork: Mixed-Initiative Time-Series Prediction

Sriram Karthik Badam[*]
University of Maryland
Student Member, IEEE

Jieqiong Zhao[†]
Purdue University

Niklas Elmqvist[‡]
University of Maryland
Senior Member, IEEE

David S. Ebert[§]
Purdue University
Fellow, IEEE

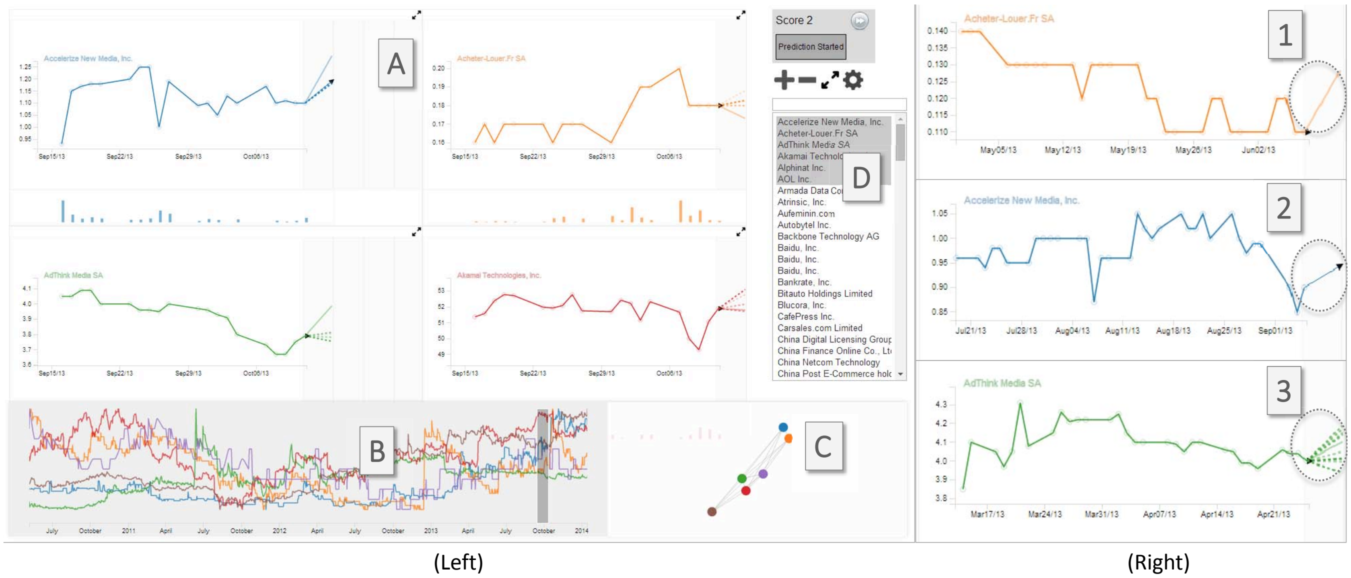

Figure 1: **(Left)** TimeFork interface showing (A) line chart and bar chart for each stock, (B) overview chart, (C) correlation viewer, and (D) the list of companies. **(Right)** Three types of predictions identified in the technique: (1) Temporal, (2) user-guided, and (3) spatial predictions.

Abstract

We present TimeFork, an analytics technique for predicting the behavior of multivariate time-series data originating from modern disciplines such as economics (stock market) and meteorology (climate), with human-in-the-loop. We identify two types of machine-generated predictions for such datasets: temporal prediction that predicts the future of an attribute; and spatial prediction that predicts an attribute based on the other attributes in the dataset. Visual exploration of this prediction space, constituting of these predictions of different confidences, by chunking and chaining predictions over time promises accurate user-guided predictions. In order to utilize TimeFork technique, we created a visual analytics application for user-guided prediction over different time periods, thus allowing for visual exploration of time-series data.

1 Introduction

Forecasting is widely used in many fields from economics and finance to meteorology. The data generated in these fields are multi-faceted [2] in nature, often times spatiotemporal and multivariate with many attributes. In the presence of such complexities, visualization has been proven to successfully provide insights chiefly through interactive visual analysis. However, forecasting/prediction sometimes exceeds the capabilities of a typical visualization application especially in presence of big data. There have been many attempts at forecasting multiple time-dependent variables in a dataset through automated data analysis algorithms [4] that typically use predictive analytics. These algorithms are capable of efficient prediction but they are offline, computationally expensive, have limited tolerance to data artifacts, and lack the ability to aid a user in real-time decision-making. In this context, we propose a mixed-initiative technique for integrating visual analytics (VA) applications with automated data analysis algorithms, thus allowing for direct interaction with both the front-end visualization and the back-end predictive analytics model.

The TimeFork technique connects the data analysis and visual representation steps of the VA pipeline thus allowing automated analysis models to provide predictions (with varying confidences) based on historical patterns and allow user to filter this prediction space to choose a prediction for decision making. For example, stock traders utilize typical patterns from historical data to buy/sell a stock based on a prediction. For this, using a machine learning model trained to offer multiple predictions and allowing the user to choose one or more prediction for each stock (attribute), based on other sources of information such as news and social network, can improve the interoperability between human and machine for a decision-making process. For prototyping the TimeFork technique, we developed a visual analytics tool for predicting stock market.

Machine learning models for stock prediction have centered around **temporal prediction**: prediction of a stock value based on previous values; where various indicators are used to effectively determine when to trade. However, stocks affect each other in many

[*]email: sriram.karthik6@gmail.com
[†]email: zhao413@purdue.edu
[‡]email: elm@umd.edu
[§]email: ebertd@purdue.edu

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

sectors and this can be utilized for **spatial prediction** to find the behavior of one stock based on the behavior of others. Therefore, TimeFork utilizes **user-guided prediction**, a process in which the user choosing one or more of the machine-generated predictions through direct interaction with the visualization. TimeFork can also extend the capability of the user to change the parameters of the predictive analytics model directly to provide context-dependent predictions for subsets of big data, as in TimeSearcher3 [1].

2 TimeFork Design

The goal of TimeFork is to allow close coupling of the visual representations and predictive analysis algorithms to facilitate user-guided predictions. This provides an opportunity for the user to select a prediction at every time instance and chain them to predict the behavior over a period of time. For example, predicting a behavior of a stock over the next one month is a hard time consuming task for a data analytics algorithm as it involves finding the path in the prediction space with maximum overall confidence for the time period or training multiple machine learning algorithms for predicting various time periods. For utilizing the TimeFork technique, we identify the following requirements for any time-series dataset:

- **R1** Provide options to visualize the attributes in the dataset for any width of the time scale (e.g.: week) and connect them to an overview visualization through brush-and-link interaction.
- **R2** Identify the typical patterns in the dataset and embed them within the visualizations. For example, static and dynamic patterns [3] help decide on the future of a stock and these need to be focus of some visual representations.
- **R3** For the predictions, type (spatial or temporal) and confidence can to be mapped to visual variables. There are many design choices for this process. (e.g.: shape and opacity).
- **R4** Allow fluid movement through the prediction space in order to chain predictions into paths for a time period. For example, it should be possible to predict ‘t+1’, ‘t+2’ and so on.
- **R5** Manage clutter while presenting the temporal and spatial patterns. This occurs because the prediction space can be huge based on the generated spatial and temporal predictions.
- **R6** Capture and retrieve prediction paths over a time period.

3 TimeFork Technique Implementation

Our implementation (Figure 1) of the TimeFork technique is targeted for stock market data. It has attribute views, overviews, and selection options, and is web-based, written in HTML, JavaScript, and CSS. Once the user selects some stocks, the line charts (R1) of the stock value and bar charts of trade volume for each stock are added to the layout. Every line chart is attached with a visual space on the right side to represent the predictions (R3, R4) that are generated by machine learning models, trained on a Java server on the historical data of each stock. The overview visualization contains the entire historical data for the selected stocks and can be used to select specific time periods through brush-and-link interaction.

Apart from these, we also provide a correlation viewer (R2), a force-directed graph with each stock of interest as a node. This visualization delivers the bigger picture or the correlations between stocks to aid better user-guided prediction. Stocks with positive correlations are placed closer to each other in this graph while the stocks with negative correlations are further away. This visualization aids the target user in exploring the prediction space. The current implementation manages clutter (R5) by opacity blending and allows saving the prediction path (R6).

For implementation purposes, we have chosen a stock market dataset containing two years of historical data from the internet information providers sector. We used neural networks for the various types of predictions: a feed forward neural network (input layer - 15 neurons, two hidden layers - 41 neurons each) for the temporal prediction of a stock, and a self-organizing map (625 neurons) modified to provide multiple spatial predictions at any time step.

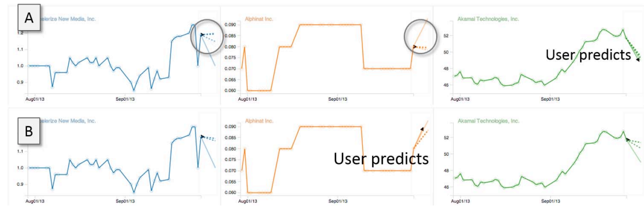

Figure 2: Two possible prediction paths are shown (one in each row). In Path A, the user selects the temporal prediction for stock 3 (green) and finds that the temporal predictions for stock 1 and 2 do not match the spatial predictions. In Path B, the user selects the temporal prediction for stock 2 (orange), and spatial pattern match with the temporal patterns better. Therefore, path B is preferred by a target user.

4 Use-case: Novice Amy

Amy is a novice stock trader interested in just three companies (A, B, and C) and she is trying to figure out when to buy stocks (Fig. 2). She looks at the past two days of data, and identifies that the temporal predictions (with highest confidence) suggest a 10% increase of stock A and 5% decrease of stocks B and C each at the next time instance. She interacts with the prediction space by accepting the major prediction (prediction with most confidence) for stock A and finds that the most of the spatial predictions are suggesting around 6% decrease for stock B, and a 10% increase for stock C (a conflicting suggestion compared to the temporal prediction). She saves this ‘What if’ scenario (Path 1) and decides to further explore the possibilities. She then goes back to the previous step and selects the major temporal prediction for stock C. This provides spatial predictions of 7% increase (high confidence) for stock A and a 5% decrease for stock B (Path 2). She decides that path 2 is most probable as it satisfies the temporal prediction and spatial prediction model with minimal conflict, and therefore buys stock A. She chooses to continue this process to predict the behavior over a time period.

5 Conclusion and Future work

We have presented TimeFork, a technique to explore multi-faceted time-series datasets to make user-guided predictions by combining visual analysis with predictive analytics models. While the use of TimeFork is well motivated, the success of the technique both in terms of the usability and accuracy is yet to be established. Future work includes (1) applying the technique to other datasets, (2) developing formal design requirements for using this technique, and (3) extending the span of the technique to work more closely with the internals of the predictive analytics models.

References

[1] P. Buono, C. Plaisant, A. Simeone, A. Aris, B. Shneiderman, G. Shmueli, and W. Jank. Similarity-based forecasting with simultaneous previews: A river plot interface for time series forecasting. In *Proceedings of the International Conference on Information Visualization*, pages 191–196, 2007.

[2] J. Kehrer and H. Hauser. Visualization and visual analysis of multi-faceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):495–513, 2013.

[3] A. V. Moere. Time-varying data visualization using information flocking boids. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 97–104, 2004.

[4] K. Venugopal, K. Srinivasa, and L. Patnaik. Fuzzy based neuro - genetic algorithm for stock market prediction. In *Soft Computing for Data Mining Applications*, pages 139–166. 2009.