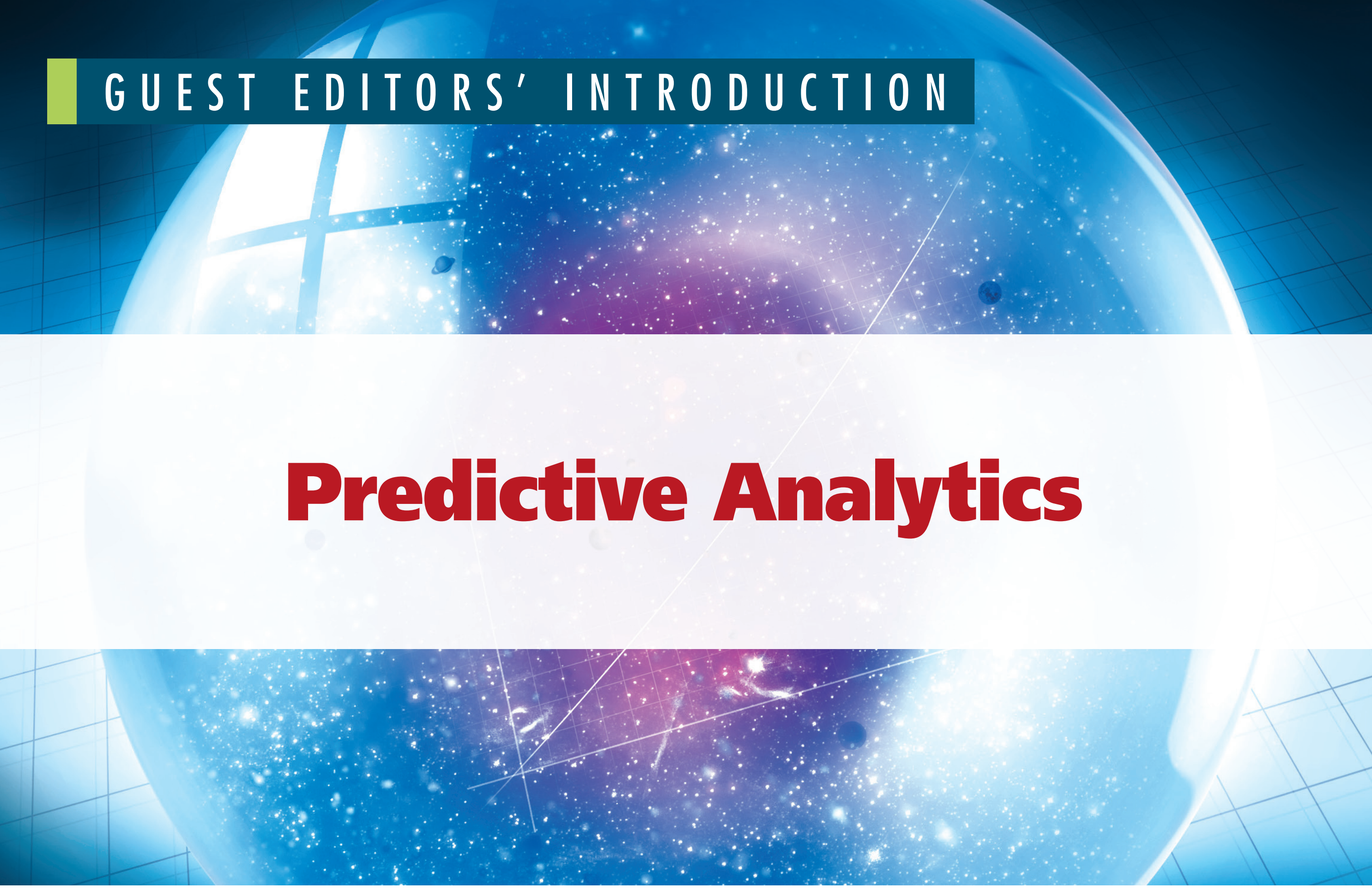

GUEST EDITORS' INTRODUCTION

Predictive Analytics

Donald E. Brown and Ahmed Abbasi, *University of Virginia*

Raymond Y.K. Lau, *City University of Hong Kong*

Predictive analytics is the use of statistical or machine learning methods to make predictions about future or unknown outcomes. Although predictive modeling techniques have been researched by the data mining community for several decades, they've become increasingly pervasive in real-world settings in recent years, impacting every facet of our lives. Novel methods are being applied in areas such as homeland security,[1] crime prevention,[2] infrastructure management, cybersecurity,[3] intelligent transportation, healthcare and bioinformatics,[4,5] text mining,[6,7] fraud detection,[8] social media,[9] and decision support for complex tasks such as mergers and acquisitions.[10] In organizational settings, predictive analytics has gained widespread adoption over the past 10 years as firms look to "compete on analytics."

In the era of big data, the volume, velocity, variety, and veracity of data generated by sensors, surveillance, transactions, clickstreams, and communication technologies precipitates the need for predictive analytics to run faster (in real time), more accurately, and using larger heterogeneous information sources of varying data quality and complexity. For instance, state-of-the-art healthcare analytics incorporates open data, social media, and multimedia content to predict health outcomes. Additionally, organizations effectively detect financial fraud using real-time anomaly detection engines capable of efficiently perusing through millions of daily transactions.

1541-1672/15/$31.00 © 2015 IEEE
Published by the IEEE Computer Society

Furthermore, we're seeing predictive analytics applied at both "micro" and "macro" levels of granularity. For instance, new forms of predictive analytics are being developed to anticipate human behavior, social dynamics, political outcomes, financial market trends, and security-related events at the individual, group, community, national, and international levels. These differences in prediction scope, coupled with the big data dimensions, present an array of exciting possibilities. In the first of a two-part special issue on predictive analytics, we feature three articles that apply analytics at the "macro" level.

In their article entitled "Predicting Elections for Multiple Countries Using Twitter and Polls," Adam Tsakalidis and his colleagues use time-series analysis of tweets and poll results over several weeks to predict the outcomes of the 2014 EU elections in Germany, the Netherlands, and Greece. While several recent studies have used Twitter to make predictions about election outcomes, the study in this issue incorporates some interesting nuances. First, the authors incorporate sentiment and volume-related attributes derived from more than 1 million tweets, coupled with the results from 48 different polls. Second, they make predictions for three different countries. Third, they evaluated their results in comparison with reputable election result prediction websites and multiple baseline methods. Fourth, whereas some studies have predicted binary outcomes (that is, winners of two-party elections), the authors make relatively accurate predictions for between six and 10 political parties per country. Overall, the study presents several interesting findings.

The second article is entitled "Financial Crisis Forecasting via Coupled Market State Analysis" by Wei Cao and Longbin Cao. Whereas several studies have tackled the problem of predicting financial market movements (and attained good results), a critical problem remains effectively forecasting financial crises. These anomalous, black swan-type happenings tend to occur rather infrequently, yet wreak havoc with many analytical forecasting models. By operationalizing the key insight that financial markets are interrelated (that is, that to varying degrees, financial markets serve as lag/lead indicators for one another and themselves), the authors consider inter- and intramarket hidden connections for forecasting major financial crises. The methodological novelty of the study lies in the proposed forecasting framework, which utilizes coupled market state analysis in conjunction with a coupled hidden Markov model. Applying their framework on a test bed encompassing more than 20 years of data from six major financial markets, the authors are able to significantly outperform several baseline methods.

The third article in this special issue is entitled "Dynamic Business Network Analysis for Correlated Stock Price Movement Prediction," by Wenping Zhang and his colleagues. Following a similar intuition to Cao and Cao's aforementioned study, these particular authors consider the interrelations between companies' financial performance. Based on the extended five forces model, they incorporate the impact of cooperative networks (based on supply-chain linkages or complementary products/services) and competing firms' dynamics. This key algorithm intuition is operationalized through a dynamic business network and energy cascading model that captures interfirm business influence using graph propagation. Using five years of data for 43 to 83 firms in each of four industry sectors—technology, energy, finance, and consumer staples—the authors' method more accurately predicts sector stock price movements than comparison techniques, thereby demonstrating the utility of incorporating network relations.

All three articles are nice exemplars of predictive analytics, encompassing novel insights, key nuances, rigorous analytical methods, large-scale experimentation, and interesting findings and takeaways for important application domains. We hope that readers

THE AUTHORS

Donald E. Brown is director of the Data Science Institute and W.S. Calcott Professor in the Department of Systems and Information Engineering at the University of Virginia. His research focuses on data fusion, statistical learning, and predictive modeling with applications to security and safety. Brown is a fellow of the IEEE, past president of the IEEE Systems, Man, and Cybernetics Society, and former EIC of *IEEE Transaction on Systems, Man, and Cybernetics, Part A: Systems and Humans*. He's the recipient of the Norbert Wiener Award and IEEE Millennium Medal. Contact him at deb@virginia.edu.

Ahmed Abbasi is associate professor of information technology, director of the Center for Business Analytics, and a member of the Predictive Analytics Lab at the University of Virginia. His work on security, health, social media, and big data has been funded through multiple NSF grants, an IBM Faculty Award, and an AWS Research Grant. Abbasi is a senior member of IEEE and an associate editor for *ACM Transactions on MIS, IEEE Intelligent Systems, Information Systems Research, and Decision Sciences*. Contact him at abbasi@comm.virginia.edu.

Raymond Y.K. Lau is an associate professor in the Department of Information Systems and associate director of the Center for Intelligent Information Systems at City University of Hong Kong. His research on information retrieval, social media, and big data analytics has been funded by multiple Hong Kong SAR and China regional grants. Lau is a senior member of IEEE and a life member of the Hong Kong Computer Society. Contact him at raylau@cityu.edu.hk.

will enjoy this issue! In the subsequent issue, we'll present articles that apply analytics at the "micro" level to individual customers, users, and firms.

References

1. H. Chen, "AI and Security Informatics," *IEEE Intelligent Systems*, vol. 25, no. 5, 2010, pp. 82–90.
2. X. Wang, D.E. Brown, and M.S. Gerber, "Spatio-Temporal Modeling of Criminal Incidents Using Geographic, Demographic, and Twitter-Derived Information," *Proc. IEEE Int'l Conf. Intelligence and Security Informatics* (ISI), 2012, pp. 36–41.
3. A. Abbasi et al., "Enhancing Predictive Analytics for Anti-Phishing by Exploiting Website Genre Information," *J. MIS*, vol. 31, no. 3, 2014.
4. D. Adjeroh et al., "Signal Fusion for Social Media Analysis of Adverse Drug Events," *IEEE Intelligent Systems*, vol. 29, no. 2, 2014, pp. 74–80.
5. T. Fu et al., "Sentimental Spidering: Leveraging Opinion Information in Focused Crawlers," *ACM Trans. Information Systems*, vol. 30, no. 4, 2012, article no. 24.
6. R.Y. Lau, C. Li, and S.S. Liao, "Social Analytics: Learning Fuzzy Product Ontologies for Aspect-Oriented Sentiment Analysis," *Decision Support Systems*, vol. 65, 2014, pp. 80–94.
7. Y. Xie et al., "MuSES: A Multilingual Sentiment Elicitation System for Social Media Data," *IEEE Intelligent Systems*, vol. 29, no. 4, 2014, pp. 34–42.
8. A. Abbasi et al., "MetaFraud: A Meta-learning Framework for Detecting Financial Fraud," *MIS Quarterly*, vol. 36, no. 4, 2012, pp. 1293–1327.
9. C.C. Yang, J. Yen, and J. Liu, "Social Intelligence and Technology," *IEEE Intelligent Systems*, vol. 29, no. 2, 2014, pp. 5–8.
10. R.Y.K. Lau et al., "Web 2.0 Environmental Scanning and Adaptive Decision Support for Business Mergers and Acquisitions," *MIS Quarterly*, vol. 36, no. 4, 2012, pp. 1239–1268.