

Conference Program

ISSAT International Conference on Data Science in Business, Finance and Industry

July 3-5, 2019

Dong A University, Da Nang, Vietnam

Conference Sponsor



The International Society of
Science and Applied Technologies
www.issatconferences.org

In Cooperation with



Dong A University



WORDS OF APPRECIATION

We wish to express our appreciation and thanks to Dr. Nguyen Thi Anh Dao, President of Dong A University, Mr. Luong Minh Sam, Vice-Rector and Ms. Nguyen Thi Ngoc Thanh for hosting and supporting DSBFI 2019 Conference and for all your help in coordinating all banquet and social events.

We would like to thank the Organizing Committee, Professor Hoang Pham, Professor Cedric Heuchenne, Professor Phuc Kim Tran and Dr. Thong Quoc Nguyen for devoting your time and tremendous efforts in organizing and making this conference a success!

We would like to thank our International Program Committee members who helped us in the rigorous and timely review process.

We would like to thank all speakers and authors for participating in the conference and making it a success. Your papers and presentations are valuable contributions to the conference.

With gratitude,
The International Society of Science and Applied Technologies



Organizing Committee Members

Conference Chairs

Hoang Pham
Rutgers University, USA

Thong Quoc Nguyen
Dong A University, Vietnam
Université Bretagne Sud, France

Program Chairs

Cedric Heuchenne
University of Liege, Belgium

Phuc Kim Tran
ENSAIT & GEMTEX, France

Arrangement Chairs

Luong Minh Sam
Dong A University, Vietnam

Ngoc Thanh Thi Nguyen
Dong A University, Vietnam

Program Committee Members

N. Balakrishnan (Canada)	Cédric Heuchenne (Belgium)	Peihua Qiu (USA)
Philippe Castagliola (France)	Jason J. Jung (South Korea)	Miklós Rásonyi (Hungary)
Giovanni Celano (Italy)	Michael B.C. Khoo (Malaysia)	Jin Seohoon (South Korea)
S. Chakraborti (USA)	Seoung Bum Kim (South Korea)	François Septier (France)
Shyi-Ming Chen (Taiwan)	Ludovic Koehl (France)	Isabelle Thomas (Belgium)
J.-C. Delvenne (Belgium)	Célestin Kokonendji (France)	Sebastien Thomassey (France)
A. Di Bucchianico (Netherlands)	Huong Bui Mai (Vietnam)	Thu Huong Truong (Vietnam)
Bay Vo Dinh (Vietnam)	Fadel M. Megahed (USA)	Fugee Tsung (Hong Kong)
Alexandre Dolgui (France)	Lars Meyer-Waarden (France)	Hoang Son Vo (Vietnam)
Jean-Marc Ferrandi (France)	Loan T. T. Nguyen (Vietnam)	Min Xie (Hong Kong)
Emmanuel Frénod (France)	Abderrahim Oulhaj (UAE)	Zach Zacharia (USA)
Thong Ngee Goh (Singapore)	Changsoon Park (South Korea)	Xianyi Zeng (France)

Conference at a Glance

WEDNESDAY, JULY 3	THURSDAY, JULY 4	FRIDAY, JULY 5
8:00 - 2:00 Registration	8:45 - 11:30 Registration	8:45 - 10:00 Registration
9:00 - 9:15 Welcome - Award Presentation	9:00 - 9:50 Keynote III	9:00 - 10:15 Panel Session
9:15 - 10:05 Keynote I	9:50 - 10:15 Coffee Break	10:15 - 10:30 Coffee Break
10:05 - 10:30 Coffee Break	10:15 - 11:45 Technical Sessions	10:30 - 12:15 Tutorial
10:30 - 11:20 Keynote II	11:45 - 12:45 Conference Luncheon	12:15 - 1:15 Farewell Lunch
11:30 - 1:00 Conference Lunch	1:00 - 2:15 Technical Sessions	1:15 Adjourn!
1:15 - 2:30 Technical Sessions	2:45 - 9:30 Hoi An Cultural Tour with Dinner and Show	
2:30 - 2:45 Coffee Break		
2:45 - 4:15 Technical Sessions		
5:30 - 8:00 Welcome Reception (Dinner Cruise)		

Agenda at a Glance

Wednesday, July 3		
9:00 - 9:15	Welcome and Opening Remarks / Award Presentation Dr. Nguyen Thi Anh Dao, President of Dong A University, Vietnam Dr. Hoang Pham, Rutgers University, USA	
9:15 - 10:05	Keynote I: From Big Data Mining to Building of Innovative Business Models in Classical Industries: A Case Study of Fashion/Textile Supply Chain Chair: Prof. Peihua Qiu (<i>University of Florida, USA</i>)	
10:30 - 11:20	Keynote II: Statistical Process Control and Its Potential for Handling Big Data Applications Chair: Prof. Cedric Heuchenne (<i>University of Liege, Belgium</i>)	
1:15 - 2:30	Session 1	Statistical Machine Learning Methods in Data Science
	Session 2	Quantitative Modeling in Finance and Industry
2:45 - 4:15	Session 3	Predictive Modeling and Analytics
	Session 4	Machine Learning and Its Applications
Thursday, July 4		
9:00 – 9:50	Keynote III: Reliability Modelling and Assessment of a Heterogeneously Repaired System with Partially Relevant Recurrence Data Chair: Prof. Hoang Pham (<i>Rutgers University, USA</i>)	
10:15 - 11:45	Session 5	Machine Learning Applications in Finance and Healthcare
	Session 6	Feature Selection and Recommender Systems
1:00 - 2:15	Session 7	Modeling and Prediction and Their Applications
	Session 8	Predictive Modeling and Recommender Systems for Business and Industry
Friday, July 5		
9:00 - 10:15	Panel Session: Data Science in Industry 4.0 Chair: Prof. Hoang Pham (<i>Rutgers University, USA</i>)	
10:30 – 12:15	Tutorial: Statistical Machine Learning & Deep Learning with Python: Methods and Applications	

Keynote Addresses

From Big Data Mining to Building of Innovative Business Models in Classical Industries: A Case Study of Fashion/Textile Supply Chain

Xianyi Zeng

Full Professor

Ecole Nationale Supérieure Des Arts Et Industries Textiles (ENSAIT), France

ABSTRACT - New business models generated in the environment of big data will enable to provide all kinds of customized services to enterprises. These data-based services include: 1) development of virtual product visualization-based digital platforms for sharing product data at different levels, breaking industrial boundaries and performing interactive design and production with direct participation of consumers; 2) building overall supply chain monitoring systems for controlling all product data from a very initial stage, such as fiber manufacturing to end products at clothing and shopping stages; 3) building product recommendation systems by relating technical parameters to consumer perception on fashion and functional properties, 4) building data-based optimized flexible manufacturing and supply chain, for realizing various customized production tasks by a series of networked small factories, 5) building networked marketing systems for realizing online forecasting consumers' behaviors and predicting final markets.

Xianyi Zeng received the B.Eng. degree from the Department of Science and Technology, Tsinghua University, Beijing, China, in 1986, and the Ph.D. degree from the Centre d'Automatique, Université des Sciences et Technologies de Lille, Villeneuve d'Ascq, France, in 1992. He is currently a Full Professor with the French Engineer School, Ecole Nationale Supérieure des Arts et Industries Textiles (ENSAIT), Roubaix, France, and Director of the GEMTEX (textile engineering and materials) National Research Laboratory. Since 2000, he has been leader of the HCD research team (Human Centered Design) in GEMTEX. He has published three scientific books, more than 110 papers at peer-reviewed international journals, and about 200 papers at international conferences. His research interests include: 1) intelligent decision support systems for fashion and material design and recommendations, 2) modeling and analysis of human perception and cognition on industrial products and their integration into CAD processes, and 3) development of intelligent wearable systems.

Dr. Zeng is currently a senior IEEE member, an Associate Editor of International Journal of Computational Intelligence System and Journal of Fiber Bioengineering and Informatics, a Guest Editor of Special Issues for sept international journals, and delivered more than 20 plenary presentations at international conferences. He has organized more than ten international conferences and workshops since 1996. Since 2000, he has been the leader of three European projects, four national research projects funded by the French government, three bilateral research cooperation projects, and more than 20 industrial projects.

Statistical Process Control and Its Potential for Handling Big Data Applications

Peihua Qiu

Professor and Founding Chair

University of Florida, USA

ABSTRACT - "Big data" is a buzzword these days due to an enormous amount of data-rich applications in different industries and research projects. In practice, big data often take the form of data streams

in the sense that new batches of data keep coming over time. One fundamental research problem for analyzing big data in a given application is to sequentially monitor the underlying process behind the observed data to see whether it is longitudinally stable, or how its distribution changes over time. To monitor a sequential process, one major statistical tool is statistical process control (SPC), which has been used mainly for monitoring production lines in manufacturing industries during the past several decades. With many new and versatile methods developed in recent SPC research, SPC can provide a powerful tool for handling many big data applications in business, finance and industry. In this talk, I will introduce some traditional and more recent SPC concepts and methods, and discuss their potential to solve some big data problems. Certain challenges in the interface of the existing SPC methods and some big data applications will also be discussed.

Peihua Qiu received his Ph.D. in statistics from the Department of Statistics at the University of Wisconsin at Madison in 1996. He worked as a senior research consulting statistician of the Biostatistics Center at Ohio State University during 1996-1998. He then worked as a faculty member of the School of Statistics at the University of Minnesota for 15 years during 1998-2013. Since 2013, he has been the professor and founding chair of the Department of Biostatistics at the University of Florida. Qiu is an elected fellow of the American Statistical Association, an elected fellow of the Institute of Mathematical Statistics, and an elected member of the International Statistical Institute. He served as associate editor for several top statistical journals, including the Journal of the American Statistical Association, Biometrics, and Technometrics. He was the editor-in-chief of the flagship statistical journal Technometrics. Qiu has made substantial contributions in the areas of jump regression analysis, image processing, statistical process control, survival analysis, and dynamic screening systems. So far, he has published over 100 research papers in refereed journals, many of which appeared in top journals, such as Technometrics, Journal of the American Statistical Association, Annals of Statistics, Annals of Applied Statistics, Journal of the Royal Statistical Society (Series B), Biometrika, Biometrics, IEEE Trans. on Pattern Analysis and Machine Intelligence, and IJSE Transactions. His research monograph titled "Image Processing and Jump Regression Analysis" (2005, John Wiley) won the inaugural Ziegel prize in 2007. His paper Qiu et al. (2010) was chosen as one of the classic papers published in Technometrics during the past 60 years.

Reliability Modelling and Assessment of a Heterogeneously Repaired System with Partially Relevant Recurrence Data

Narayanaswamy Balakrishnan

Distinguished University Professor
McMaster University, Canada

ABSTRACT - In this talk, I will first consider a reliability data to provide a basic motivation for the reliability problem considered in this work. Next, I will explain the stochastic modelling of the reliability problem and then describe the assessment methods for reliability. I will then revisit the data and illustrate the model and the assessment methods developed here. Finally, I will conclude the talk with some brief remarks and further suggestions!

Dr. Narayanaswamy Balakrishnan is a Distinguished University Professor in the Department of Mathematics and Statistics at McMaster University, Hamilton, Ontario, Canada. He is a Fellow of the American Statistical Association, a Fellow of the Institute of Mathematical Statistics and an Elected Member of the International Statistical Institute. He is the author/editor of many books and monographs, including the highly-cited four volume-set on "Distributions in Statistics" (co-authored with N.L. Johnson and S. Kotz), published by John Wiley & Sons, and the 16-volume set on "Encyclopedia of Statistical Sciences" (co-edited with S. Kotz, C.B. Read and B. Vidakovic), published by John Wiley & Sons. He was recently awarded an Honorary Doctorate Degree by The National University of Athens, Greece.

Statistical Machine Learning Methods in Data Science

SESSION 1 - Chair: Prof. Ping Shing Ben Chan (*The Chinese University of Hong Kong, China*)

An Outer–Inner Linearization Method for Nonconvex and Nondifferentiable Composite Statistical Learning Problems

Minh Pham (Rochester Institute of Technology, USA)

Xiaodong Lin (Rutgers University, USA)

Andrzej Ruszczyński (Rutgers University, USA)

ABSTRACT – We propose a new outer–inner linearization method for nonconvex statistical learning problems involving nonconvex structural penalties and nonconvex loss. Many important statistical problems fall in this category, including the robust M-estimators, generalized linear models, and different types of structured learning problems. Our method exploits the fact that many such loss and penalty functions can be represented as compositions of smooth concave functions and nonsmooth convex functions. It linearizes the outer concave functions and solves the resulting convex, but still nonsmooth, subproblems by a special alternating linearization method. Numerical examples involving nonconvex structural penalties and nonconvex loss functions demonstrate the efficacy and accuracy of the method.

Wearable Sensor Data Based Human Activity Recognition Using Machine Learning: A New Approach

H. D. Nguyen (Vietnam National University of Agriculture, Vietnam)

K. P. Tran (ENSAIT, GEMTEX Laboratoire de Génie et Matériaux Textiles, France)

X. Zeng (ENSAIT, GEMTEX Laboratoire de Génie et Matériaux Textiles, France)

L. Koehl (ENSAIT, GEMTEX Laboratoire de Génie et Matériaux Textiles, France)

G. Tartare (ENSAIT, GEMTEX Laboratoire de Génie et Matériaux Textiles, France)

ABSTRACT – Recent years have witnessed the rapid development of human activity recognition (HAR) based on wearable sensor data. One can find many practical applications in this area, especially in the field of health care. Many machine learning algorithms such as Decision Trees, Support Vector Machine, Naive Bayes, K-Nearest Neighbor and Multilayer Perceptron are successfully used in HAR. Although these methods are fast and easy for implementation, they still have some limitations due to poor performance in a number of situations. In this paper, we propose a novel method based on the ensemble learning to boost the performance of these machine learning methods for HAR.

Anomaly Detection by Time Series Motif Discovery for Vietnamese Customs

Dong Van Hoang (Ha Noi University of Science and Technology, Vietnam)

Quang Anh Ngoc Pham (Ha Noi University of Science and Technology, Vietnam)

Nam Thanh Vu (Ha Noi University of Science and Technology, Vietnam)

Ngoc Anh Thi Nguyen (Ha Noi University of Science and Technology, Vietnam)

ABSTRACT – Financial transactions exploded recently by mobile, online transactions in the whole world. Fraud detection problems have been important great interests of many researchers. In data analytics, fraud detection is used machine learning methods. This paper proposes a supervised method by time series motif discovery for customs. A new proposed model is given using time series, motif discovery and machine learning for the real-life data of account company from customs data. The results of this paper achieve the new models for motif discovery and experiment results.

Quantitative Modeling in Finance and Industry

SESSION 2 - Chair: Dr. Viktor Prokop (University of Pardubice, Czech Republic)

Bootstrap Confidence Interval for Regression Coefficients of PCR

Jieun Kim (Korea University, South Korea)
Seohoon Jin (Korea University, South Korea)

ABSTRACT – There is a problem of multicollinearity arising from the correlation between independent variables in multiple regression analysis. Principal components regression (PCR) analysis is one of the methods used to solve this problem. Since PCR uses the principal component as an independent variable, it is difficult to identify the significance of the original variables. We used the bootstrap to build confidence interval of regression coefficients of the original variables and interpreted the meaning of the variables.

Modelling the Relationship between the CSR Investment and the Financial Performance of Capital Constrained Firms

Dinh Anh Phan (Univ Rennes 1, CNRS, CREM - UMR 6211, France)
Thi Le Hoa Vo (Univ Rennes 1, CNRS, CREM - UMR 6211, France)
Anh Ngoc Lai (Univ Rennes 1, CNRS, CREM - UMR 6211, France)

ABSTRACT – The recognition of the corporate social responsibility (CSR) effects on firm's financial performance has important implications for financial managers in making CSR decisions. Unfortunately, CSR has both negative and positive relationship with the financial performance in empirical findings. To provide a more nuanced explanation for the relationship between CSR and a firm's financial performance, we developed an analytical model for the impacts of CSR investment on the cost of bank loans and on the profitability of the capital constrained firms. Our analysis shows how the influence of CSR on the demand for firm's products affects the cost of bank loans and the firm's profitability. In addition, we explore the conditions under which investing in CSR can reduce the cost of bank loans and enhance the firm profitability. The results potentially clarify the complexity of the mechanisms through which CSR affects the firm's financial performance and suggest some directions for empirical studies on this link.

Importance of Basic Attributes in Innovation Environment - Efficiency Analysis in the EU28

Viktor Prokop (University of Pardubice, Czech Republic)

Jan Stejskal (University of Pardubice, Czech Republic)
Petr Hájek (University of Pardubice, Czech Republic)
Michaela Stříteská (University of Pardubice, Czech Republic)

ABSTRACT – Nowadays, measuring efficiency within countries' innovation environment seems to be incremental in the process of gaining competitive advantage. Therefore, this study is aimed to evaluate efficiency in patent creation within EU28 countries. We are using specialized tool for assessing the effectiveness, performance and productivity of comparable production units Data Envelopment Analysis and data from Eurostat. Moreover, we are analyzing countries' efficiency according to their innovation performance measured by European Commissions' European Innovation Scoreboard 2017. Results show that only 5 out of the 28 European countries are effectively using basic attributes of Innovation Environment (investment in science and research; human resources in science and technology; cooperation with external research and development firms). All of these countries belong to the group of Innovation Leaders. We also propose practical implications (for each country) on how to improve and how to change their inputs and outputs to become (more) efficient and provide information about countries that could be benchmark for less efficient countries.

Predictive Modeling and Analytics

SESSION 3 - Chair: Dr. Quang Hung Do (*University of Transport Technology, Vietnam*)

Predicting Win-Loss of League of Legends at the Start of the Game Using Machine Learning Techniques

Cheolgi Kim (Soongsil University, South Korea)
Soowon Lee (Soongsil University, South Korea)

ABSTRACT – E-sports is not only a simple game that individuals enjoy but also a world-class sport that is adopted as a demonstration sport in the 2018 Asian Games. The League of Legends is one of the e-sports games. In this paper, we propose a win-loss prediction model of League of Legend at the start of the game using machine learning techniques. The prediction accuracy of 60.25% was obtained experimentally with the artificial neural network model. By using the prediction model proposed in this paper, it is expected that the team-based strategy can be established in choosing a champion at the start of the game.

One-Sided Run Rules Control Charts for Coefficient of Variation with Measurement Errors

P. H. Tran (HEC Liège - Management School of the University of Liège, Belgium)
C. Heuchenne (HEC Liège - Management School of the University of Liège, Belgium)
H. D. Nguyen (Vietnam National University of Agriculture, Vietnam)

ABSTRACT – We investigate in this paper the effect of measurement errors on the performance of Run Rules control charts monitoring the coefficient of variation (CV) squared. The previous Run Rules - CV in literature is improved slightly by monitoring the CV squared instead of the CV itself. The numerical results show that this improvement gives better performance for Run Rules charts. Moreover, we will show through simulation that the precision and

accuracy errors do have negative effect on the performance Run Rules charts. We also find out that multiple measurements per item does not have high efficiency in reducing these negative effects.

Prediction of KOSPI Fluctuation Based on S&P500 Index Using a Neural Network Model

Taeseung Kim (Soongsil University, South Korea)

Soowon Lee (Soongsil University, South Korea)

ABSTRACT – Stock price forecasting is an interesting subject that is studied extensively in various fields such as economics, mathematics, statistics, and artificial intelligence. However, it is not easy to predict stock price because stock price movements contain a lot of noise and generally have nonlinear characteristics. In the past, there have been studies to quantify the degree of correlation between various variables and stock prices and to predict future volatility through time series analysis of past stock price patterns. There is also a method based on artificial intelligence that generates a predictive model by learning a machine learning model using data affecting stock price formation. In this paper, we propose a stock price prediction model based on LSTM that predicts the next day's stock price fluctuation of Korea's stock index, KOSPI, using S&P500, the US stock index. Comparing the prediction accuracy of the proposed method with the prediction accuracy of the comparison models, the prediction accuracy of the proposed method is about 2% higher than the other models.

Forecasting Crude Oil Price Based on Mutual Information Technique and ANFIS Trained by Biogeography-Based Optimization (BBO) Algorithm

Quang Hung Do (University of Transport Technology, Vietnam)

Duc Son Nguyen (University of Transport Technology, Vietnam)

Thanh Hang Thi Doan (University of Transport Technology, Vietnam)

Thu Ha Thi Vu (University of Transport Technology, Vietnam)

ABSTRACT – Crude oil price forecasting is a crucial task in the field of energy research because crude oil is a world's major commodity with a high volatility level. This study proposes the Adaptive Neuro-Fuzzy Inference System (ANFIS) with parameters optimized by Biogeography-Based Optimization (BBO) algorithm and mutual information (MI) technique for forecasting crude oil price. The MI is utilized to maximize relevance between inputs and output and minimize the redundancy of the selected inputs. The proposed approach combines the strengths of fuzzy logic, neural network and the heuristic algorithm to detect the trends and patterns in crude oil price data, and thus have been successfully applied to crude oil price forecasting. Other different forecasting methods, including artificial neural network (ANN) model, ANFIS model, and linear regression method are also developed to validate the proposed approach. In order to make the comparisons across different methods, the performance evaluation is based on root mean squared error (RMSE), mean absolute error (MAE), relative absolute error (RAE), root relative squared error (RRSE), and correlation coefficient (R). The performance indexes show that the ANFIS-BBO model achieves lower MAE, RMSE, RAE and RRSE, as well as higher R, indicating that the ANFIS-BBO model is a better method.

Machine Learning and Its Applications

SESSION 4 - Chair: Prof. In Hong Chang (*Chosun University, South Korea*)

Modeling Color Fading Ozonation of Dyed Textile Using Artificial Intelligence

Zhenglei He (ENSAIT, GEMTEX, France)
Kim-Phuc Tran (ENSAIT, GEMTEX, France)
Sébastien Thomassey (ENSAIT, GEMTEX, France)
Xianyi Zeng (ENSAIT, GEMTEX, France)
Changhai Yi (Wuhan Textile University, China)

ABSTRACT – Textile products with faded effect achieved via ozonation are increasingly popular recently. In this study, the effects of ozonation in terms of pH, temperature, water pick-up, time and original colors on the color fading performance of reactive-dyed cotton are modeled using Extreme Learning Machine (ELM), Support Vector Regression (SVR) and Random Forest Regression (RFR) respectively. It is found that RF and SVR perform better than ELM in this issue, but SVR is more recommended to be used in the real application due to its balance predicting performance and less training time.

Integrating Sentiment Analysis in Recommender Systems

Thanh Hung Bui (Thu Dau Mot University, Vietnam)

ABSTRACT – Customer product reviews play an important role in the customer's decision to purchase a product or use a service. Providing a useful suggestion of products to online users to increase their consumption on websites is the goal of many companies nowadays. In this paper, we propose a recommender system based on sentiment analysis. The system is built by integrating sentiment analysis to recommender system in order to generate the most accurate. We use hybrid deep learning method CNN-LSTM for sentiment analysis based on vector of words in the customer product reviews. The result in the sentiment analysis is used to combine the neighbor's item ratings to produce a prediction value for the target user. This helps the recommender system to generate efficient recommendations for that user. We do experiment in Amazon food review dataset. The proposed model shows interesting results on the impact of integrating sentiment analysis in the recommender systems.

Modelling Innovation Paths of European Firms Using Fuzzy Balanced Scorecard

Petr Hájek (University of Pardubice, Czech Republic)
Jan Stejskal (University of Pardubice, Czech Republic)
Michaela Strítěská (University of Pardubice, Czech Republic)
Viktor Prokop (University of Pardubice, Czech Republic)

ABSTRACT – Modelling innovation paths is considered a difficult task because innovation processes are complex, uncertain and highly dimensional. As traditional regression models fail to address these issues, here we propose a novel approach for the modelling. The approach integrates Balanced Scorecard, a method used for strategic performance assessment, and fuzzy set qualitative comparative analysis. In addition to key performance indicators,

strategic goals are taken into consideration in the modelling. We provide empirical evidence for the effectiveness of the approach on a large dataset of European firms. We show that several innovation pathways can be identified for these firms, depending on their strategic goals. These results may be of relevance for the decision making of innovative firms and other actors of innovation system.

Comparison Activation Function of Deep Neural Network for Class Recognition Classification of the Elderly

Dong Su Lee (Chosun University, South Korea)
Youn Su Kim (Chosun University, South Korea)
In Hong Chang (Chosun University, South Korea)

ABSTRACT – This paper identifies the degree of subjective class recognition of the elderly. We examine the degree of class recognition classification of the elderly by various variables that the general characteristics of elderly people, subjective expectation, life satisfaction, household income, living cost, etc. There is a significance to suggest a new dimension to the quality of life of the elderly by examining the influence factors of the elderly's life not only on the individual dimension but also on the subjective class recognition of the individual in the community. This paper presents the types of activation functions during the deep learning analysis and compares the classification rates by type. In this paper, 4,485 elderly people aged 55 years and over using the 6th data of the aging research panel of the Korea Employment Information Service in 2018 are subjected to using the deep neural network during the deep learning analysis. The results show that the PReLU among functions that the ReLU function, the Leaky ReLU function, the PReLU function, and ELU function have the highest class recognition classification of the elderly.

Machine Learning Applications in Finance and Healthcare

SESSION 5 - Chair: Prof. Jaehee Cho (Kwangwoon University, South Korea)

Insights from Using Macroeconomic Indicators to Predict the Monthly Closing Price of Major U.S. Stock Indices

Bin Weng (Auburn University, USA)
Waldyn Martinez (Miami University, USA)
Yao-Te Tsai (Feng Chia University, Taiwan)
Chen Li (Auburn University, USA)
Lin Lu (Auburn University, USA)
James R. Barth (Auburn University, USA)
Fadel M. Megahed (Miami University, USA)

ABSTRACT – In this talk, we propose a two-stage approach that can be used to investigate whether the information hidden in macroeconomic variables (alone) can be used to accurately predict the one-month ahead price for major U.S stock and sector indices. Stage 1 is constructed to evaluate the hypothesis that the price for different indices is driven by different economic indicators. It consists of three phases. In phase I, the data is automatically acquired using freely available APIs (application programming interfaces) and prepared for analysis. Phase II reduces the set of potential predictors without the loss of information through several variable selection methods. The third phase

employs four ensemble models and three time-series models for prediction. The prediction performance of the seven models are compared using the Mean Absolute Percent Error (and two additional metrics). If the hypothesis were to be true, one expects that the performance of the ensemble models to outperform the time-series models since the information in the economy is more important than the information in previous prices. In Stage 2, a hybrid approach of the recurring neural network used for time-series prediction (i.e., the LSTM) and the ensemble models is constructed to examine the secondary hypothesis that the residuals from the time-series models are not random and can be explained by the macroeconomic indicators. To test the two hypotheses, the monthly closing prices for 13 U.S. stock and sector indices and the corresponding values for 23 macroeconomic indicators were collected from 01/1992–10/2016. Based on the case study, the four ensembles prediction performance were superior to that of the three timeseries models. The MAPE of the best model for a given index was $< 1.87\%$. The Stage 2 results also show that the three evaluation metrics (RMSE, MAPE and MAE) can be typically improved by 25–50% by incorporating the information hidden in the macroeconomic indicators (through the ensemble approach). Thus, we show that, for the analysis period and the indices studied, the macro-economic indicators are leading predictors of the price of 13 U.S. sector indices.

Industrial Internet of Things, Big Data, and Artificial Intelligence in the Smart Factory: A Survey and Perspective

H. D. Nguyen (Vietnam National University of Agriculture, Vietnam)

K. P. Tran (ENSAIT, GEMTEX Laboratoire de Génie et Matériaux Textiles, France)

X. Zeng (ENSAIT, GEMTEX Laboratoire de Génie et Matériaux Textiles, France)

L. Koehl (ENSAIT, GEMTEX Laboratoire de Génie et Matériaux Textiles, France)

P. Castagliola (Université de Nantes, LS2N UMR CNRS 6004, France)

P. Bruniaux (ENSAIT, GEMTEX Laboratoire de Génie et Matériaux Textiles, France)

ABSTRACT – Thanks to the rapid development and applications of advanced technologies, we are experiencing the fourth industrial revolution, or Industry 4.0, which is a revolution towards smart manufacturing. The wide use of cyber physical systems and Internet of Things leads to the era of Big Data in industrial manufacturing. Artificial Intelligence algorithms emerge as powerful analytics tools to process and analyze the Big Data. These advanced technologies result in the introduction of a new concept in the Industry 4.0: the smart Factory. In order to fully understand this new concept in the context of the Industry 4.0, this paper provides a survey on the key components of a smart factory and the link between them, including the Industrial Internet of Things, Big Data and Artificial Intelligence. Several studies and techniques that are used to enable smart manufacturing are reviewed. Finally, we discuss some perspectives for further researches.

Proposal of Feature Matching Technique Using Similarity Features Filtering for Image Alignment

Thanh Trung Nguyen (R&D Lab, Sun* Inc, Vietnam)

Hong Anh Thi Pham (R&D Lab, Sun* Inc, Vietnam)

Toan Van Pham (R&D Lab, Sun* Inc, Vietnam)

Thanh Minh Ta (Le Quy Don Technical University, Vietnam)

An Van Tran (Le Quy Don Technical University, Vietnam)

ABSTRACT – Recently, stemming from the storage of raw data into database systems for the purpose of further analysis and statistics has made digitizing paper invoices become one of the important tasks in many traditional businesses. The task is to automatically extract information from user-provided invoice image such as customer's name, customer's address, detail of transaction and so on, then to store to database. In general, this task consists of three phases: (1) align the image with the template, (2) detect and extract the information regions in the image then, (3) recognize the optical character with deep learning technique. Image alignment is the most important step in the pipeline and directly affects the result, the accuracy of next step. In this paper, we propose a new approach for feature matching method called similarity features filtering and some techniques applying on invoices image pre-processing to improve the image alignment accuracy. The experimental results show that our proposed approach can achieve better results than other feature-based methods.

Dartboard-Shaped Spatial Segmentation: Depicting Distribution of Geo-Tweets and Defining Urban Boundaries

Jaehee Cho (Kwangwoon University, South Korea)

Euiyoung Baik (Kwangwoon University, South Korea)

ABSTRACT – This study compares and analyzes the spatial distribution of people in three Vietnamese cities using the geolocated Twitter data. Ho Chi Minh, Ha Noi and Da Nang were selected as the target cities. The data collection period is 123 days in 2016 and 125 days in 2018. The spatial distribution of the three cities is analyzed according to two periods and residence status (local vs foreigner). In this study, we have found popular places based on a newly devised spatial segmentation method and we also estimated the urban boundaries by reflecting the location as well as the volume of geo-tweets. Instead of histograms, CDF (Cumulative Distribution Function) curves are employed to show the distribution of distances between the city center point and actual tweet location.

Feature Selection and Recommender Systems

SESSION 6 - Chair: Dr. Thong Quoc Nguyen (*Université de Bretagne-Sud & LMBA UMR CNRS 6205, France*)

Extract Features Using Gaussian Filter and Its Application for Improvement of an Accuracy of the Epileptic Detection Using SVM

Ha Manh Hoang (Thu Dau Mot University, Vietnam)

Quang Tung Ha Hoang (International University, Vietnam)

Ngoc Giau Ho (Thu Dau Mot University, Vietnam)

ABSTRACT – In diagnostic of Epilepsy, an important issue is to infer epileptic seizure presence. Until recently, the SVM is the main tool for the epileptic seizure recognition purpose. The accuracy of this approach base on features extraction ability of the pre-processing. An important issue is to extract phenomena that correspond to the thorn of epileptic. This paper introduces method for improving the robustness of classification problems using one dimensional Gaussian filter. The experimental results show that our method has advantage in case of complicated epileptic seizure.

Extended Sentence Similarity Based on Word Relations for Document Summarization

Heechan Kim (Soongsil University, South Korea)

Soowon Lee (Soongsil University, South Korea)

ABSTRACT – Automatic text summarization is a method to extract key sentences from input documents for readers to understand the documents with little effort. Automatic text summarization is largely divided into follows. Abstractive summarization is a method to create a new human-like natural language to provide a summary. Extractive summarization is a method to select salience sentences. It is one of the mainly studied in the fields of natural language processing research. One of the represented extractive summarization methods is TextRank, in which sentences in a document are represented as a graph and the similarity between sentences is calculated based on the frequencies of co-occurring words. This similarity measure has a drawback in that it does not sufficiently consider the semantic similarity between words in a sentence. To overcome this drawback, in this paper, we propose a similarity measure between words by defining co-occurrence relations of all word pairs in a sentence. Further, we propose a novel sentence vector function to apply the co-occurrence relations between words while calculating the similarity between sentences. The experiments revealed that the proposed method was more accurate than TextRank.

Conceptual Design of Incremental Matrix Consolidation Operation for Recommendation Systems

Nam D. Vo (Chung-Ang University, South Korea)

Hoang Long Nguyen (Chung-Ang University, South Korea)

Gen Li (Chung-Ang University, South Korea)

Jason J. Jung (Chung-Ang University, South Korea)

David Camacho (Iniversidad Autónoma de Madrid, Spain)

ABSTRACT – In this paper, we propose a method for consolidating multiple matrices incrementally for recommendation system. Based on mathematical approach methods, we unify multiple matrices together based on the overlap of the utility matrix. Specifically, we suggest a procedure to consolidate matrices together. Then the recommendation system can use the unification matrix to produce the item predictions for users.

Real-Time Production Monitoring Approach for Smart Manufacturing with Artificial Intelligence Techniques

Q. T. Nguyen (Université de Bretagne-Sud & LMBA UMR CNRS 6205, France)

H. D. Nguyen (Vietnam National University of Agriculture, Vietnam)

K. P. Tran (ENSAIT, GEMTEX Laboratoire de G'énie et Mat'eriaux Textiles, France)

P. Castagliola (Université de Nantes & LS2N UMR CNRS 6004, France)

E. Frénod (Université de Bretagne-Sud & LMBA UMR CNRS 6205, France)

ABSTRACT – Production monitoring in real-time is a very important problem in smart manufacturing. It helps enterprises to timely detect abnormalities in the production process and then guarantee the product quality and reduce waste. In this paper, we develop a novel method to monitor the real-time production based on the

Convolution Neural Network and the Support Vector Data Description algorithm. The numerical result shows that our proposed method leads to highly efficient on the testing data.

Modeling and Prediction and Their Applications

SESSION 7 - Chair: Dr. Fadel Megahed (*Miami University, USA*)

Prediction of English Premier League Game Results by Using Deep Learning Techniques

Jaehyun Yi (Soongsil University, South Korea)
Soowon Lee (Soongsil University, South Korea)

ABSTRACT – Predicting the outcome of a sporting event can provide an overall flow of events and enable the team to improve its performance. Therefore, much research has been conducted on the prediction of the outcome of sports events through statistical techniques and deep learning techniques. In this paper, we propose a deep learning model predicting English Premier League game results based on the league rankings at the time of the competition and the previous game content of both teams. The proposed model uses the long short term memory model, considering the characteristics of the time series input data. Experimental results show that the average performance of proposed model was 3.6% higher than the existing models.

Link Prediction in Graph Stream by Distributed Computation Approach

Khanh-Duy Le-Trinh (Vietnam National University HCMC, Vietnam)
Anh-Thu Nguyen-Thi (Vietnam National University HCMC, Vietnam)
Tu-Anh Nguyen-Hoang (Vietnam National University HCMC, Vietnam)

ABSTRACT – Recently, graph stream has become an essential model to represent interactive elements in the massive networks. It is a promising key to performing many real-world applications, such as social networks, E-commerce networks, and telecommunication networks. However, the most exciting link prediction methods just focused on predicting the existence of links in snapshot graphs, while most recent applications in the form of the graph stream. When applying these methods to graph stream, we have two challenges: a) a large number of n nodes which produce n^2 possible links, resulting in significant complexity and b) the rapid evolvement of graph stream. In this paper, we introduce an effective method to predict real-time existence link in graph stream. We propose an efficient Graph Stream Distributed Computation framework (GSDC) which is immediately amendable to parallelization, facilitating a scale-able distributed implementation on Apache Spark platform. With our framework, we use the Distributed Computation Score feature (DCS) which is designed to parallel compute the similarity scores between nodes. Experiment results on three real-world social networks demonstrate the effectiveness of the proposed framework.

A Novel Two-Stage Machine Learning Approach for Monotonic Heart Transplantation Survival Prediction

Hamidreza Ahady Dolatsara (Auburn University, USA)

Ying-Ju Chen (University of Dayton, USA)
Christy Evans (Auburn University, USA)
Fadel M. Megahed (Miami University, USA)

ABSTRACT – Accurate prediction of graft survival after a heart transplant is an important, yet challenging problem because: (a) it is the only treatment option for patients with end-stage heart failure; (b) the availability of hearts from deceased donors is scarce; (c) it requires an estimation of the matching suitability of patient-donor based on their medical information; and (d) its success is affected by the patient's adherence to strict medical instructions after transplant. Here, we propose a two-stage approach for estimating the graft survival probabilities at 1-10 years post surgery. First, we estimate the survival probability at different time points using machine learning methods. Then, we calibrate these probabilities using isotonic regression. Using a national registry of U.S. heart transplants from 1987-2016, we showed that our first stage produces an area under the receiver operating curve, AUC, between 0.75 and 0.92 for years 1-10. Additionally, the application of isotonic regression to smooth/calibrate the survival probabilities for each patient over the 10-year period guarantees monotonicity. This study provides two general implications: (a) accurate predictions can be obtained using the two-stage methodology; and (b) using isotonic regression to calibrate machine learning models in survival analysis can lead to more informative results.

Predictive Modeling and Recommender Systems for Business and Industry

SESSION 8 - Chair: Dr. Phuc Kim Tran (ENSAIT & GEMTEX, France)

Goodness-Of-Fit Tests for the Component Lifetimes Distribution Based on the System Failure Data with Known Signature

Ping Shing Chan (The Chinese University of Hong Kong, China)
Yaoxi Zhong (The Chinese University of Hong Kong, China)
Gaofeng Da (Nanjing University of Aeronautics and Astronautics, China)

ABSTRACT – Suppose the lifetimes of coherent systems of a given, fixed design with components which have i.i.d. lifetimes distributed according to a common distribution F are observed. We would like to test the hypothesis that if the distribution F is exponential or Weibull. Kolmogorov-Smirnov type statistics are developed to test such hypothesis. Power studies are carried out via simulations.

Crowdsourcing System for Measuring Cognitive Similarity in Recommendation System

Luong Vuong Nguyen (Chung-Ang University, South Korea)
O-Joun Lee (Chung-Ang University, South Korea)
Hoang Long Nguyen (Chung-Ang University, South Korea)
Sojung An (Chung-Ang University, South Korea)
Jason J. Jung (Chung-Ang University, South Korea)
Yue-Shan Chang (National Taipei University, Taiwan)

ABSTRACT – In this research, we develop OurMovieSimilarity (OMS) as a crowdsourcing system that can i) provide an effective process for collecting similar movies with high interoperability and ii) recommend similar movies based on users' cognition. With this approach, OMS can achieve better precision in discovering similar movies and overcome the cold start problem. In order to evaluate our system, we conducted experiments by using data from IMDB data and OMDB. The result shows that OMS effectively reduces the wasting time of users in discovering their similar movies. Besides, we can extend this research for clustering users by their cognition using our similar movie dataset.

Prediction of Fine Dust (PM2.5) Concentration Based on RBF Kernel SVM

Deukwoo Lee (Soongsil University, South Korea)

Soowon Lee (Soongsil University, South Korea)

ABSTRACT – Fine dust in the air causes respiratory problems. Particularly PM2.5 causes more serious health problems because it infiltrates deep blood vessels in the body, even the brain. Thus, it is important to predict days with 'Bad' level of PM2.5. In this paper, we propose a model to predict PM2.5 of Seoul city from air information. The prediction model uses RBF Kernel SVM with the adjusted gamma parameter. To evaluate the performance of the prediction model, we conducted comparative experiments using various models. As a result, the proposed model showed the best performance against the others in all labels.

Panel Session

Chair: Prof. Hoang Pham (*Rutgers University, USA*)

Data Science in Industry 4.0

Prof. Peihua Qiu (University of Florida, USA)

Prof. Cedric Heuchenne (University of Liege, Belgium)

Prof. Xianyi Zeng (ENSAIT, France)

Prof. Narayanaswamy Balakrishnan (McMaster University, Canada)

Tutorial

Statistical Machine Learning & Deep Learning with Python: Methods and Applications

Dr. Thong Quoc Nguyen (Université Bretagne Sud, France & Dong A University, Vietnam)

Dr. Huy Ngoc Chau (Alfred Renyi Institute of Mathematics, Hungary)

**Thank you for being part of the DSBFI 2019.
We hope you enjoy the conference and the
beautiful city of Da Nang**

