Special Session 9: Mathematics for Information Processing and Management

Jianhong Wu, York University, Canada
Zongben Xu, Xian Jiaotong University, China

ELECTRE ranking approach for benchmarking analysis in marketing sector

Ai Ling Amy Poh
Meiji University, Malaysia
amypoh.al@gmail.com

The marketing mix is the lens through which the contemporary customer perceives value in retail stores on 4Ps is examined. From the model, we analyze the best practice among the four elements derived from a consensus ranking, a ranking method to identify the best in class. The analysis mainly depended on the outcome of what customer perceive towards the four marketing tactics. This paper discusses the introduction and use of a methodology for project ranking in retail store and, in particular, illustrates the use of a particular solution method called ELECTRE. In the effort of avoiding the shortcomings of the traditional methods based on the average aggregate monocriterion, outranking methods make it possible to deal with multicriteria benchmarking.

⟶ ∞ ⋄ ∞ ⟵

Discovering most collaborative teams of experts in social networks

Aijun An
York University, Canada
aan@cse.yorku.ca
Mehdi Kargar

We study the problem of discovering teams of experts from a social network. Given a project whose completion requires a set of skills, our goal is to find a set of experts that together have all of the required skills and also have the minimal communication cost among them. We present two communication cost functions designed for two types of communication structures. We show that the problem of finding the team of experts that minimizes one of the proposed cost functions is NP-hard. Thus, a 2-approximation algorithm is designed. We introduce the problem of finding a team of experts with a leader. The leader is responsible for monitoring and coordinating the project, and thus a different communication cost function is used in this problem. To solve this problem, an exact polynomial algorithm is proposed. We show that the total number of teams may be exponential with respect to the number of required skills. Thus, two procedures that produce top-k teams of experts with or without a leader in polynomial delay are proposed. Extensive experiments on real datasets demonstrate the effectiveness and scalability of the proposed methods.

⟶ ∞ ⋄ ∞ ⟵

Time consistent multiperiod risk measure under generalized convex framework

Zhiping Chen
Xi'an Jiaotong University, Peoples Rep of China
zchen@mail.xjtu.edu.cn

It is now accepted that time consistency should be a necessary requirement for multi-period risk measures. After investigating the relationship among different definitions of time consistency and discussing properties that a reasonable multi-period risk measure should possess, we propose a new class of time consistent multi-period risk measures under generalized convex framework. Basing on the new risk measure, we construct a corresponding multi-period portfolio selection model. Scenario technique is then used to efficiently compute the new multi-period risk measure and solve the multi-period portfolio selection problem. Preliminary numerical results are finally provided to illustrate the reasonability and practicality of our new multi-period risk measure and the corresponding multi-period portfolio selection model.

⟶ ∞ ⋄ ∞ ⟵

Firm clustering using standard-based financial statements

Dazhi Chong
Old Dominion University, USA
dchon001@odu.edu
Dazhi Chong, Hongwei Zhu

Classification and clustering of firms can produce useful insights about various aspects about the firms. Most existing approaches to classifying businesses only pay attention to the operation processes and outputs of firms. In this paper, we introduce a firm clustering method that uses firms' financial information. Publicly listed companies in the U.S. has been required to use a common data standards called the U.S. GAAP Taxonomy to submit their financial statements to the Securities and Exchange Commission. When constructing their financial statements, firms select data elements from the Taxonomy to tag their financial data. In this research, we construct the "social network" of firms using the shared data

elements in firms' financial statements. By implementing a spectral clustering method and applying it to the "social network", we identify clusters that differ from frequently used classification schemes. Firms within the same cluster share similar structures in their financial statements. Other firm characteristics are being explored to obtain additional insights about the clusters.

⟶ ∞ ◇ ∞ ⟵

Network utility optimization and effect to response time

Wenying Feng
Trent University, Canada
wfeng@trentu.ca
Christopher Johns, Kevin Mak

We study the optimization of network utilities under the constraints of limited resources. New models are developed and algorithms for theoretically solving the problems are presented. Particular cases are discussed. Applying simulation techniques, we also investigated the effects of different utility functions on network response time when the total utilization of the network is optimized.

⟶ ∞ ◇ ∞ ⟵

The influence of media on social behaviour in an influenza pandemic

Jane Heffernan
York University, Canada
jmheffer@yorku.ca
Shannon Collinson

Media reports affect social behaviour during epidemics and pandemics. Changes in social behaviour, in turn, affect key epidemic measurements such as peak magnitude, time to peak, the beginning and end of an epidemic. The extent of this effect has not been realized. We have developed mathematical models of influenza spread based on a Susceptible-Exposed-Infected-Recovered (SEIR) model including the effects of mass media. The models are used to evaluate different functions representing media impact and how these functions affect key epidemic measurements. We have also developed an agent based Monte Carlo (ABMC) simulation of influenza infection between hosts to determine variability in key epidemic measurements.

⟶ ∞ ◇ ∞ ⟵

Beyond traditional search: probabilistic approaches and their applications

Jimmy Huang
York University, Canada
jhuang@yorku.ca

Most of the traditional Information Retrieval models are based on the assumption that query terms are independent of each other and a document is represented as a bag of words. Nevertheless this assumption may not hold in practice. In this talk, I will discuss how the query terms associate with each other and how to incorporate the term proximity information into the classical probabilistic IR models. Through extensive experiments on standard large-scale TREC Web collections, I will show that the extended models are able to markedly outperform the BM25 baseline and at least comparable to the state-of-the-art model. The talk will conclude with a discussion of novel challenges raised in extending probabilistic Information Retrieval and several applications such as promoting diversity in ranking for biomedical IR, sentiment analysis for predicting sales performance and EMR data analysis for effective health care.

⟶ ∞ ◇ ∞ ⟵

Estimating a density in the presence of interval censoring

Hanna Jankowski
York University, Canada
hkj@mathstat.yorku.ca

In this talk I will compare several methods of density estimation, including parametric and nonparametric approaches. I will discuss the benefits and drawbacks of each approach. Finally, I will illustrate the various methods on data coming from the 2009 H1N1 pandemic in Ontario, Canada.

⟶ ∞ ◇ ∞ ⟵

A divide-and-conquer approach to effective and efficient L1 norm matrix factorization

Deyu Meng
Xi'an Jiaotong University, Peoples Rep of China
dymeng@mail.xjtu.edu.cn
Zongben Xu

The L1 norm low-rank matrix factorization has been recently attracting much attention due to its intrinsic robustness to outliers and missing data. However, most of the existing algorithms suffer the problems such as high complexity and/or low accuracy. In this paper, we propose a novel solution, which is essentially a divide-and-conquer approach, to robust L1 norm matrix factorization. The main idea is to break

the original difficult problem into many pieces of simple sub-problems, each involving only one unknown scalar parameter and having a closed-form solution. By recursively solving these small problems without time-consuming inner loop numerical optimization, an efficient algorithm can be readily constructed to tackle the original problem. The computational complexity of the proposed algorithm is approximately linear in both data size and dimensionality, making it be able to handle large-scale L1 norm matrix factorization problems. The extensive experimental results validate that our method outperforms state-of-the-arts in term of both computational time and accuracy, especially on large-scale applications such as face recognition and structure from motion.

⟶ ∞ ◇ ∞ ⟵

Agent-based modelling frameworks for developing public health policies

Seyed Moghadas
York University, Canada
Moghadas@yorku.ca

Advances in Information and Communications Technologies (ICTs) during the past two decades have provided a major conduit for knowledge exchange, which conveys the notion of human connectivity within a complex network. These advances, combined with scientific discoveries, have had profound impacts on many aspects of real life, and led to dramatic changes in approaches to addressing societal issues. The underlying processes by which our knowledge influences public policy and decision-making involves retrieval and analysis of vast amounts of heterogeneous and complex data collected through the use of ICT tools. We show how an Agent-Based Modelling framework can integrate such data, and inform public health policies in the event of an emerging disease. We use the 2009 pandemic influenza H1N1 as a case study to illustrate how intervention strategies can be targeted to reduce the impact of disease on the population.

⟶ ∞ ◇ ∞ ⟵

Tensor regression model for crime prediction

Yang Mu
University of Massachusetts, USA
yangmu@cs.umb.edu
Yang Mu, Henry Lo, Wei Ding

Crime incidents are affected by both spatial and temporal factors as well as other predictor variables. As such, crime prediction requires that we take spatiotemporal attributes into account when processing crime data. Yet, the structure and order inherent in space and time dimensions are not considered adequately using other mathematical techniques; it is typical for prediction models to treat each spatial point as distinct and independent, even though in reality variables tend to be correlated with values in neighboring cells. In this work, we use a tensor regression model to predict a crime dataset that spans a northeastern city in last 5 years. This model aims to minimize the Ordinary Least Square (OLS) error and accepts tensor inputs directly. It greatly reduces the appearance of Small Sample Size (SSS) problem, which could easily occur in a vector based OLS case. Tensor features are extracted and used to train a tensor regression model in order to predict future crime. The tensor-based regression model is evaluated with comparison to a traditional point-prediction model, which considers only historical data but no spatial neighborhood data. Results show that the tensor model is far more effective in predicting hotspots of crime than the traditional model.

⟶ ∞ ◇ ∞ ⟵

L1/2 regularization theory for sparse machine learning

Zongben Xu (special session keynote speaker)
Xi'an Jiaotong University, Peoples Rep of China
zbxu@mail.xjtu.edu.cn
Baili Chen

Machine Learning (ML) is a fundamental tool in data-driven information technologies, which aims at modeling a likely existing but unknown relationship between the input and output spaces based a given set of samples. In applications, ML leads to ill-posed problems and it is solved then through constraining the solutions to be found (via regularization). The constraints normally take the form of certain type of norms of the solution, say, the 2-norm, reflecting the expectation that the solution to be found is as smooth as possible, or the 1-norm, expecting the solution to be sparse and as simple as possible. I propose to apply the 1/2 quasi-norm in this talk in performing regularization. I develop an L(1/2) regularization based sparse ML theory, showing that the new regularization scheme preserves the classical Moreau's forward-backward splitting representation, can be very fast solved and is capable of yielding nearly sparest solution with good generalization performance. I present also two applications to demonstrate the powerfulness of the developed theory. The first application is on image processing (the image representation, denoising, impainting, restoration and identification in particular), which shows how the new theory can yield the state of the art techniques for those problems. The second application is on microwave imaging. In this later case I show that the new theory underlies a new microwave imaging principle, and furthermore, leads to a very useful, new SAR system.

⟶ ∞ ◇ ∞ ⟵

Risk assessment of China's commercial banks: assessing data quality

Feng Xu
School of Management, Xi'an Jiaotong University, Peoples Rep of China
xf_xjtu@foxmail.com
Tan Zhang,Yahui Li,Hongyun Zhang

This research studies the function of analytical modeling in risk assessment of China's commercial banks. This mathematical model aims to solving the risk assessment problems, especially in data quality (Kaplan et al. 1998). We will use this analytical model as a systematic exploration from data quality perspective, which can be used to improve credit card businesses of China's commercial banks in the future. In addition, a decision support system is suggested so that financial regulators can use it for assessing data quality more effectively and efficiently. This research constructs a mathematical model and algorithms that support financial regulators to determine the target error classes as well as the minimum set of control procedures when assessing financial risks, which is the key to ensure high level of data quality, especially high level of reliability.

$\longrightarrow \infty \diamond \infty \longleftarrow$

Delay differential equations and its application in developmental biology

Xuguo Zhang
Xi'an Jiaotong University, Peoples Rep of China
x.zhang@mail.xjtu.edu.cn
Steven Bishop, Julian Lewis

Delay Differential Equations (DDEs) play a fundamental role in the various fields such as Biology, Engineering and even Economics over recent years. This paper, with its focus on DDEs, investigate a specific model for cellular formation in Zebrafish developed by Lewis (2003), hereafter called Lewis model to illustrate the dynamic behaviour and the properties of Ordinary Differential Equations (ODEs) and DDEs extracted from Lewis work (Lewis, 2003). By using linearisation and some simple analytic skills such as Bendixsons negative criterion and Liapunov function, it has been showed that system without delay cannot exhibit self-sustained oscillation. When model involves delays, we mainly analyse and explore the form of f function (Rate of synthesis for new mRNA molecules) extensively using Matlab and a new f function has been found to satisfy the DDE system. It has been showed that the amplitude of the oscillation against delay T or the power of protein n is monotonically changing when parameters are varied and meantime the appropriate bifurcation diagram is showed. This investigation have made a strong theoretical surport for Lewis experimental work and the significant analytic skills improved the results as well.

$\longrightarrow \infty \diamond \infty \longleftarrow$

An improved model of anonymous entity resolution in the public sector

Zhou Zhang
School of Management, Xi'an Jiaotong University, Peoples Rep of China
joecanoe@126.com
Jing Chen, Hongyun Zhang

Entity resolution (ER) is a body of knowledge and practice related to the activities supporting a process to decide whether two entity references are equivalent or not equivalent; entity reference refers to a collection of identity attribute values that describe a particular entity. A previous model was developed to deal with ER in the public sectors, where the information stored in different databases is prevented from data mining by a data warehouse because of policies and regulations (Talburt et al. 2005). In this paper, an improved model is proposed to deal with the problem of anonymous entity resolution in the Intranet. In addition, the model is more efficient with the use of Bloom Filter when accessing and comparing information between different sectors. Based on the improved model, this research is anticipated to improving data quality in the public sectors successfully.

$\longrightarrow \infty \diamond \infty \longleftarrow$