

# **MIT-Tsukuba Joint-Workshop on Data Systems Science towards Social and Business Innovations**

## **Program & Book of Abstracts**

**January 20<sup>th</sup> & 21<sup>th</sup>, 2020**

**University of Tsukuba, Tokyo Campus  
Bunkyo-School Building, Room 134/119  
Tokyo, Japan**

**Faculty of Business Sciences, University of Tsukuba  
MIT Institute for Data, Systems, and Society (IDSS)  
Graduate School of Business Sciences, University of Tsukuba**

### **Organizing committee:**

**Yuji Yamada (University of Tsukuba)  
Setsuya Kurahashi (University of Tsukuba)  
Naoki Makimoto (University of Tsukuba)  
Yasufumi Saruwatari (University of Tsukuba)  
Ali Jadbabaie (MIT)**

### **Supported by:**

- Pre-Strategic Initiative (Research Base-Building) program  
adopted project in FY2018, University of Tsukuba**
- Grant-in-Aid for Scientific Research (A) 16H01833, JSPS**
- Challenging Research (Exploratory) 19K22024, JSPS**

**Day1 Room134****Monday, January 20<sup>th</sup>****Opening address**

10:00-10:10

**■ 1<sup>st</sup> Session: Data management and social influence / Chair: Yasufumi Saruwatari**

10:10-11:10

**Dean Eckles**

Massachusetts Institute of Technology, USA

**“Seeding with limited or costly network information”**

11:15-11:55

**Yukie Sano**

Faculty of Engineering, Information and Systems, University of Tsukuba, Japan

**“Information spreading on social media: Network analysis and simulation”**

11:55-12:35

**Masataka Ban**

Faculty of Business Sciences, University of Tsukuba, Japan

**“A new inventory classification criterion for retail CRM”**

12:35-13:45

**\*\*\*\* lunch \*\*\*\*****■ 2<sup>nd</sup> Session: Networks and markets / Chair: Setsuya Kurahashi**

13:45-14:45

**Ali Jadbabaie**

Massachusetts Institute of Technology, USA

**“Misinformation spreading in social networks”**

14:50-15:30

**Kenji Tanaka**

Department of Technology Management for Innovation, The University of Tokyo, Japan

**“A Concept Proposal for Peer-to-Peer Power Exchange for Internet-of-Energy era”**

15:30-16:10

**Yuji Yamada**

Faculty of Business Sciences, University of Tsukuba, Japan

**“Creating prediction error insurance market for renewable energy trading”**

16:10-16:30

**\*\*\*\* coffee break \*\*\*\*****■ 3<sup>rd</sup> session: Data science and business innovations / Chair: Yuji Yamada**

16:30-17:10

**Mika Sato-Ilic**

Faculty of Engineering, Information and Systems, University of Tsukuba, Japan

**“Statistical Data Science Based on Soft Computing”**

17:10-17:50

**Yukio Ohsawa**

Department of Systems Innovation, The University of Tokyo, Japan

**“Create an Innovators' Marketplace with Data Jackets”**

18:10-20:00

**\*\*\*\* Reception (Kanobi Meikei-Kan)\*\*\*\***

**Day2 Room119****Tuesday, January 21<sup>th</sup>**

---

**■4<sup>st</sup> Session: Data-driven decisions / Chair: Ali Jadbabaie****10:00-10:15**      **Ali Jadbabaie**

Massachusetts Institute of Technology, USA

**“Introduction of IDSS”****10:15-10:55**      **Fujio Toriumi**

Department of Systems Innovation, The University of Tokyo, Japan

**“How to design rewarding systems for Consumer Generated Media”****11:05-11:45**      **Setsuya Kurahashi**

Faculty of Business Sciences, University of Tsukuba, Japan

**“Model-based Policy Making: Health policy, Electricity market and Urban dynamics.”****11:45-13:00****\*\*\*\* lunch \*\*\*\*****■5<sup>nd</sup> Session: Sports analytics and management / Chair: Naoki Makimoto****13:00-14:00**      **Peko Hosoi**

Massachusetts Institute of Technology, USA

**“Sports in the Age of Data”****14:05-14:45**      **Yoshio Takahashi**

Faculty of Health and Sport Sciences, University of Tsukuba, Japan

**“Innovation of Sports by New Technologies and Data Systems Science in Japan”****Closing address****14:45-15:00**      **Hideo Kigoshi,**

Vice President and Executive Director for Research, University of Tsukuba, Japan

Registration



Event site



# Book of Abstracts

## ■ Dean Eckles

### “Seeding with limited or costly network information”

When a behavior (e.g., product adoption) may spread through a social network, it can be advantageous to consider network structure when deciding where to seed that behavior. But what if the network is not yet observed and doing so is costly? Some recent empirical work employs methods that only rely on limited network information (e.g., by exploiting the friendship paradox). We consider these and other more sophisticated stochastic seeding strategies that likewise involve sampling information about the network.

In the first part, we develop nonparametric methods for empirically evaluating stochastic seeding strategies, including by reusing existing data. This draws on the policy evaluation, dynamic treatment regime, and importance sampling literatures. We show that the proposed estimators and designs can dramatically increase precision while yielding valid inference. We apply our proposed estimators to two field experiments on insurance marketing in rural China and anti-conflict interventions in New Jersey schools.

In the second part, we develop novel seeding strategies that come with guarantees about the loss compared with seeding using complete knowledge of the network. These algorithms make a bounded number of queries of the network structure and provide tight approximation guarantees for arbitrary networks. We test our algorithms on empirical network data to quantify the trade-off between the cost of obtaining more refined network information, and the benefit of the added information for guiding improved seeding strategies.

(The first paper <https://arxiv.org/abs/1809.09561> is joint work with Alex Chin and Johan Ugander, and the second is <https://arxiv.org/abs/1905.04325> joint work with Hossein Esfandiari, Elchanan Mossel, and M. Amin Rahimian.)

## ■ Yukie Sano

### “Information spreading on social media: Network analysis and simulation”

Information spreading is a crucial issue for society and companies. For example, under emergencies, incorrect and uncertain information may coexist, causing social confusion and sometimes causing serious damage to companies. However, it is generally very difficult to accurately capture the information spreading process. This is because unlike simple spreading processes like HIV, the relationship between the sender and the receiver plays a vital role in information sharing, and makes the spreading process more complicated.

In this presentation, we first present the results of a large-scale analysis of the actual data in Japan and China on both true and false information spreading. As a common feature in both countries, the network distance in false information spreading is larger than that of true information. The deviation of the degree distribution in false information spreading networks is smaller than that of the true information. When this feature is used to determine whether the information is false or true from the network topology alone

without the content, it is possible to make a determination at an early stage, e.g., five hours after the first re-postings.

In addition to network analysis, we also present the results of simulations using virtual scenarios on actual networks. By conducting simulations, it becomes possible to quantitatively discuss how the spreading speed and the scale change in the absence of an influential sender (influencer).

## ■ Masataka Ban

### “A new inventory classification criterion for retail CRM”

In this research, we propose a new criterion for goods inventory classification to formulate retailer's CRM (Customer Relationship Management) strategies. In general, a retail store manager classifies their inventory according to sales volume, average unit cost, annual dollar usage, lead time and so forth. In most cases, they are short-term performance criteria for store profitability, although a retailer using forward CRM strategy needs to evaluate the profitability based on the long-term criterion such as a customer LTV (Life-Time Value).

On the other hand, retailers often manage their customers using the so-called RFM metrics, in which “R” denotes purchase recency, “F” purchase frequency, and “M” average monetary amount of purchases. The retail managers classify the customers by the RFM to identify that customers are highly/lowly profitable, and are ready to churn. Although the RFM strongly depends on data periods, Shumittlein and Peterson (1994, Marketing Science) and their subsequent studies have developed models to estimate customer LTV with a small dependence on the RFM data period.

In this work, we construct a model that the customer LTV is regressed on their purchased items, assuming that a profitable customer purchases an item leading to a high performance for store profit. Then, each regression coefficient indicates how much an item does contribute to the customer LTV. Additionally, Shumittlein's LTV is calculated by a PLS statistics, where the PLS stands for “Purchase rate”, “Lifetime”, and “Spending per transaction” estimated by the RFM. We investigate how much an item does work for the PLS, and conclude that the model enables us to evaluate the long-term performance of inventory being consistent to the customer LTV. Note that more detailed explanations of the model and empirical studies will be shown in the presentation.

## ■ Ali Jadbabaie

### “Misinformation spreading in social networks”

## ■ Kenji Tanaka

### “A Concept Proposal for Peer-to-Peer Power Exchange for Internet-of-Energy era”

The last decades have seen a huge increase of distributed energy resources. Managing this growing number of intermittent resources requires both a highly interconnected grid as well as distributed storage for balancing supply and demand. As the trend for decentralized infrastructure and hardware continues, so does the need for decentralized control systems and energy management software.

We argue that market mechanisms will be key to incentivizing both consumer and prosumer to compromise in electricity trading in order to alleviate peak demands and solving the demand-response problem. However, being able to conceive a secure and decentralized control and billing system adapted to such autonomous, peer-to-peer exchanges is one of the biggest challenges of this century.

We propose the applicability of Blockchain as new decentralized tool to enable P2P trading within prosumers of microgrids. To test the feasibility of such a system, we are conducting two real scale microgrid projects with households, shopping centre, and plug-in EVs. The result shows the future potential for harmonizing the gap between demand and supply with this mechanism.

## ■ Yuji Yamada

### “Creating prediction error insurance market for renewable energy trading”

Predicting future weather conditions is important for electricity industries with renewable energy generators to quote a next-day sales contract (i.e., day-ahead sales contract). If a prediction error exists, the market-monitoring agent has to prepare another power generation resource to immediately compensate for the shortage, resulting in an additional cost for the monitoring agent. In this context, a penalty may be required depending on the size of the prediction error, which may lead to a significant loss for renewable electric power industries. Because the main source of such losses is from prediction errors of weather conditions, they can instead effectively utilize an insurance contract (or a derivative contract) based on prediction errors of weather conditions to hedge against loss caused by prediction errors of power output for renewable energy trading.

The objective of this work is to introduce an insurance market based on prediction error weather derivatives, where we focus on solar power energy trades that particularly increased in Japan. To the end, we construct a cross-hedging strategy consisting of solar radiation and temperature derivatives and propose to apply nonparametric regression techniques that enable us to find optimal derivative contracts for solar prediction errors and/or optimal contract volumes of temperature derivatives. It is thought that the development and knowledge sharing of our hedging methods are effective not only for risk management of individual power industries but also for policy decision making as to efficient risk imposition on power industries.

## ■ Mika Sato-Ilic

### “Statistical Data Science Based on Soft Computing”

Today’s vast and complex societal data require adaptable methods to meet the rapidly changing needs of data structures in many areas. Statistical data analysis conventionally played a core role in dealing with any type of data and formed the theoretical foundation for a wide range of data analysis. However, conventional data analysis dependent on statistical methods is not always adequate to handle the often complex data types that make up this data.

Soft computing has a key framework to tolerate imprecise and uncertain problems with a robust and computational optimal low-cost solution. This opens the door to a new paradigm based statistical

multivariate analysis called soft data analysis that is capable of solving these statistical challenges.

In soft data analysis, in which conventional statistical methods and machine learning or data mining methods are combined, we have developed several models that utilize the latent scale captured by soft computing techniques to explain data. While the original scale does not have the capacity to act as the scale for complex data, a scale extracted from the data itself can deal with this data.

This presentation focuses on challenging issues of statistical data analysis caused by the new vast and complex data, the general problems of conventional statistical data analysis, how our soft computing based latent-scaled models are related to these issues, and how the models solve the problems with some applications.

## ■ Yukio Ohsawa

### “Create an Innovators' Marketplace with Data Jackets”

Data Jackets are human-made metadata for each dataset, reflecting peoples' subjective or potential interests. By visualizing the relevance among DJs, participants in the market of data think and talk about why and how they should combine the corresponding datasets.

Even if the owners of data may hesitate to open their data to the public, they can present the DJs in the Innovators Marketplace on Data Jackets that is a platform for innovations. Here, participants communicate to find ideas to combine/use/reuse data or future collaborators. Furthermore, explicitly or implicitly required data can be searched by the use of tools developed on DJs, which enabled, for example, analogical inventions of data analysis methods.

In this talk, the speaker shows some results on the applications of DJs to the explanation of latent dynamics of , consumption market, stock market, earthquakes, and soccer games. Thus, we show a data-mediated platform to cultivate plans in business, science, and daily activities that are now being extended to the redesigning of living and working environments.

## ■ Fujio Toriumi

### “How to design rewarding systems for Consumer Generated Media”

After web 2.0 era, thousands of systems that provide a platform to bridge user to user are developed which called Consumer Generated Media (CGM) such as social news sites, review sites, information sharing sites and video sharing sites. Various incentive systems are implemented in CGM to encourage those content-providers. Both functions of comments and the "like" buttons are well-known systems to reward the providers. Here comes a research question: What kind of incentives encourage to post articles on CGM?

In this talk, I'll explain our agent-based model to confirm the effect of incentive systems implemented to CGM.

## ■ Setsuya Kurahashi

### “Model-based Policy Making: Health policy, Electricity market and Urban dynamics”

Many significant policies of our society and economy are determined by someone day after day. However, most of the plans have been discussed and decided based on past experiences and data. Many of them estimate policy effects by analysing actual phenomena and data using statistical methods. For this method called evidence-based policymaking (EBP), this lecture proposes model-based policymaking (MBP). The MBP is designed with an agent-based model and data science techniques, and it is also called as social simulation. The model-based approach enables to design realistic phenomena as a model and predict the effect on unfolding future events due to hypotheses or activities that are difficult to experiment using computer experiments.

In the field of business and sociology, data analysis as an induction method and strategy planning as a deductive method are connected. In the lecture, I will introduce the analyses of health policy for infectious diseases, electricity market and urban dynamics.

## ■ Peko Hosoi

### “Sports in the Age of Data”

In light of recent advances in data collection, sports possess a number of features that make them an ideal testing ground for new analyses and algorithms. In this talk I will describe a few studies that lie at the intersection of sports and data. The first centers on fantasy sports which have experienced a surge in popularity in the past decade. One of the consequences of this recent rapid growth is increased scrutiny surrounding the legal aspects of the games, which typically hinge on the relative roles of skill and chance in the outcome of a competition. While there are many ethical and legal arguments that enter into the debate, the answer to the skill versus chance question is grounded in mathematics. In this talk I will analyze data from daily fantasy competitions and propose a new metric to quantify the relative roles of skill and chance in games and other activities. This metric is applied to FanDuel data and to simulated seasons that are generated using Monte Carlo methods; results from real and simulated data are compared to an analytic approximation which estimates the impact of skill in contests in which players participate in a large number of games. We then apply this metric to professional sports, fantasy sports, cyclocross racing, coin flipping, and mutual fund data to determine the relative placement of all of these activities on a skill-luck spectrum.

The second project I will describe is a collaboration with Major League Baseball to determine the physics behind the recent increase in the rate of home runs. In this part of the talk I will enumerate different potential drivers for the observed increase and evaluate the evidence in the data (box score, ball tracking, weather, etc) in support of each theory.

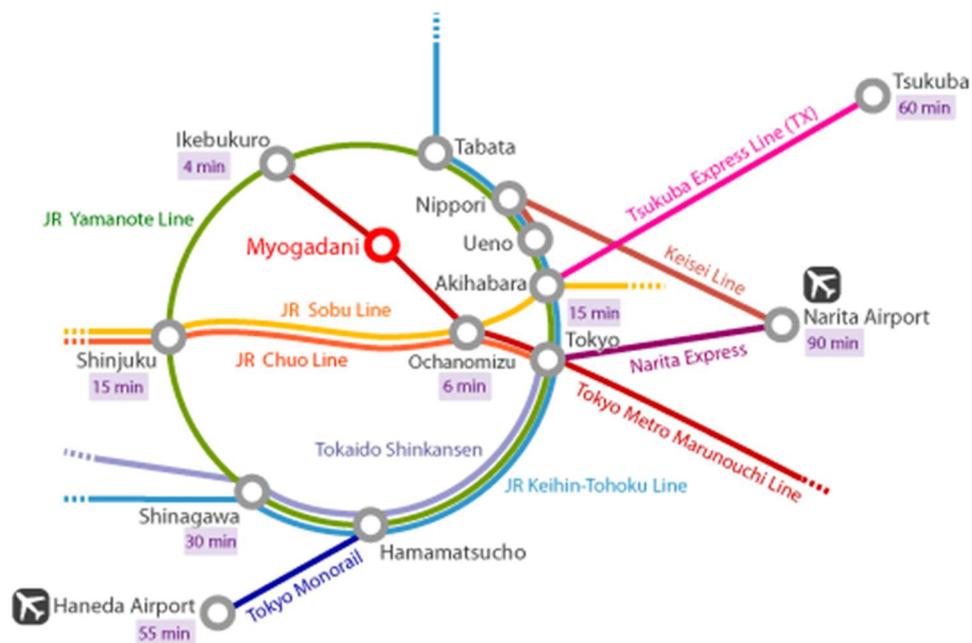
**■ Yoshio Takahashi****“Innovation of Sports by New Technologies and Data Systems Science in Japan”**

In the 21st century, the IT industry began to drive the economy in Japan. In the sports world, Rakuten decided to enter professional baseball in 2005, and DeNA acquired the baseball team in 2012. Originally, Japanese sports have spread mainly to club activities as part of school education. Companies also owned sports clubs for employee benefits and advertising, but they did not use them for technology development. This is very different from American sports, which have developed as entertainment business, and in Japan, there was no incentive to introduce new technology into the sports world. Against this background, data analysis for winning the Olympics and world championships was the first to draw attention by the Japan Institute of Sports Sciences. The data analyst who led the national volleyball team to a bronze medal in London Olympics launched the Japan Sports Analysts Association. On the other hand, IT continued to develop in Japan, and as a result of the successful bid for the Tokyo Olympics, the number of companies that disseminated Japan's technological capabilities to the world increased in the wake of the Tokyo Olympics. The Sports Agency, which was established in 2015, positions the sports industry as a growth industry and supports sports innovation, the development of new technologies, and data science as a national sports policy. In the presentation, I would like to introduce the use of New Technologies and Data Systems Science in Japanese sports.

## Access

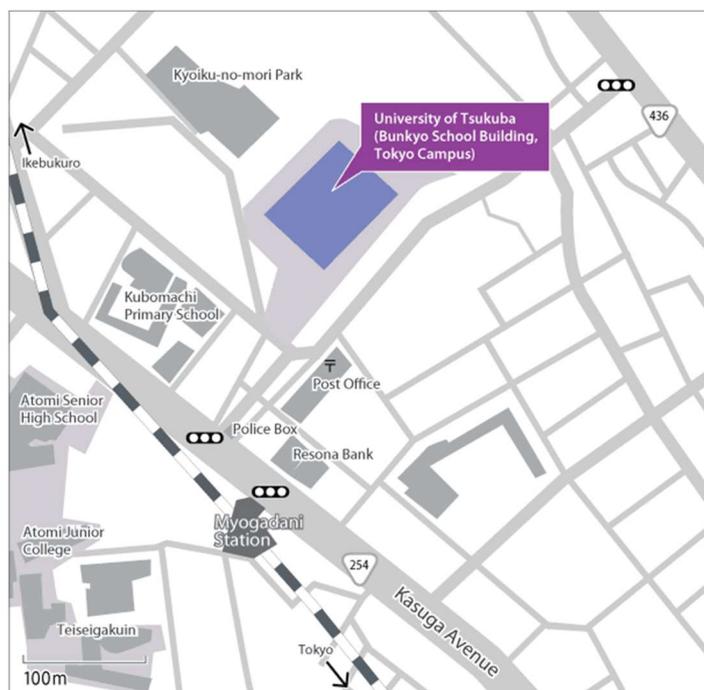
### ○By Train

Tokyo Metro Marunouchi Line (M23): Myogadani Station (about two-minute walk)



### ○Address

3-29-1 Otsuka, Bunkyo-ku, Tokyo



Google Map  
(Tokyo campus)

