International Kick-off Symposium of Graduate Program in Data Science

22nd - 23rd, February 2018

Tohoku University

Sendai, Japan

Table of contents

Overview

❖ Schedule overview

Day1			Day2		
Thursday, February 22			**Friday, February 23**		
9:00 -		Registration	9:00 -		Registration
9:30 -	9:40	Opening Remarks	9:30 -	11:30	Session 4: Invited Talks (Economics and Management)
9:40 -	9:50	Opening Address			
10:00 -	11:30	Session 1: Invited Talks (Medicine, Life Science and Bioinformatics)	11:30 -	11:45	Presentation Awards Ceremony
11:30 -	14:00	Lunch break	11:45 -		Closing Remarks
14:00 -	16:00	Session 2: Invited Talks (Data Science and Engineering)			
16:15 -	20:00	Session 3: Poster presentation and Get-toghether party			

❖ Location

Tohoku University, Aobayama campus, Aramaki-aza-Aoba, Aoba, Sendai, Miyagi, 980-8579
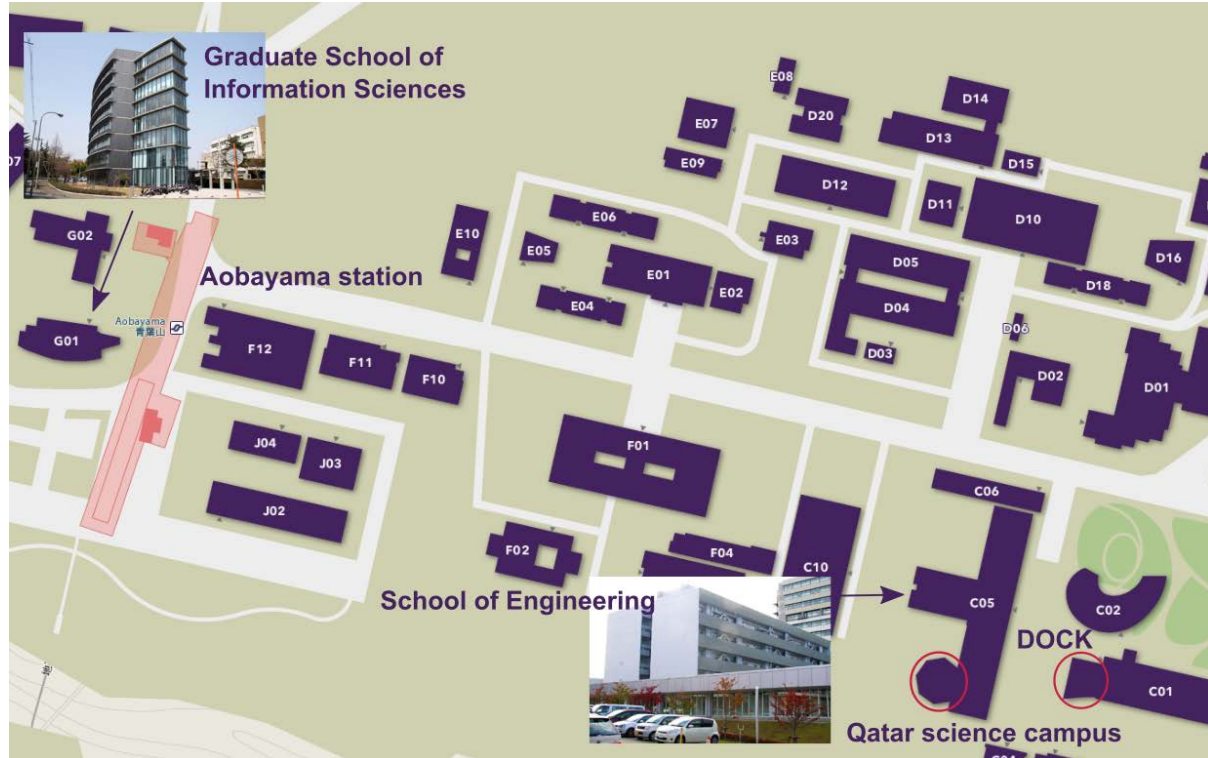
Main venue: **Qatar Science Campus Hall**

Poster presentation and Get-together party: Restaurant **DOCK** (Starting time: 17:30)

❖ Wireless LAN at Qatar Science Campus Hall

Please check a placard in the reception area.

Access and contact information

❖ Aobayama campus map

❖ Contact

Graduate School of Information Sciences, Tohoku University,
6-3-09, Aramaki-aza-Aoba, Aoba, Sendai, 980-8579, Japan
TEL: +81-22-795-4691
Mail: gpds_office@is.tohoku.ac.jp / kyamada@ecei.tohoku.ac.jp

Program

❖ February 22

09:00-	**Registration**
09:30-09:40	**Opening Remarks**
	Prof. Kimio Hanawa, Executive Vice President, Tohoku University
09:40-09:50	**Opening Address**
	Prof. Takeshi Tokuyama, Dean of Graduate School of Information Sciences, Tohoku University

Session 1: Medicine, Life Science and Bioinformatics (Chair: Prof. Takashi Makino)

10:00-10:30	**Invited talk**: **Network Biology for Advancing Drug Discovery and Human Health**
	Prof. Mark Chance, School of Medicine, Case Western Reserve University
10:30-11:00	**Invited talk**: **Construction of Gene Coexpression Databases with a Viewpoint of Evolutionary Systems Biology**
	Prof. Takeshi Obayashi, Graduate School of Information Sciences, Tohoku University
11:00-11:30	**Invited talk**: **Exploring Evolutionary Potential Using Genomic Big Data**
	Prof. Masakado Kawata, Graduate School of Life Sciences, Tohoku University

11:30-14:00	**Lunch Break**

Session 2: Data Science and Engineering (Chair: Prof. Shingo Kagami)

14:00-14:30	**Invited talk**: **Materials Data Science: Using Distributed/High Performance Computing for Lifetime Performance of Photovoltaic Systems**
	Prof. Roger French, Department of Material Science and Engineering, Case Western Reserve University
14:30-15:00	**Invited talk**: **Research Infrastructures for Data Science**
	Prof. Ramin Yahyapour, Institute of Computer Sciences, Georg-August-Universität Göttingen
15:00-15:30	**Invited talk**: **Systems Science of Animal Navigation: Trajectory Data Mining and Neural Control of Movement**
	Prof. Koichi Hashimoto, Graduate School of Information Sciences, Tohoku University
15:30-16:00	**Invited talk**: **Data Analysis for Precise Smart Agriculture based on Deep Learning Algorithms**
	Prof. Nen-Fu Huang, College of Electrical Engineering and Computer Science, National Tsing Hua University

Session 3: Poster presentation and Get-together party (Chair: Prof. Naho Orita)

16:15-17:15	**Poster preview**

17:30-20:00   **Poster presentation and Get-together party – Restaurant *DOCK***

01: Kentaro Ogawa (Graduate School of Life Sciences)

02: Huan Sun (Graduate School of Life Sciences)

03: Agness Lakudzala (Graduate School of Life Sciences)

04: Komaki Ninomiya (Graduate School of Life Sciences)

05: Ishara Perera (Graduate School of Life Sciences)

06: Turgut Akyol (Graduate School of Life Sciences)

07: Daiki Sato (Graduate School of Life Sciences)

08: Lida Sanchez (Graduate School of Life Sciences)

09: Yuichi Aoki (Graduate School of Information Sciences)

10: Hayato Anzawa (Graduate School of Information Sciences)

11: Yuki Kagaya (Graduate School of Information Sciences)

12: Kazunori Yamada (Graduate School of Information Sciences)

13: Shun Kodate (Graduate School of Information Sciences)

14: Yutaro Okano (Graduate School of Information Sciences)

15: Siwalee Choilek (Graduate School of Information Sciences)

16: Xiyue Wang (Graduate School of Information Sciences)

17: Cherdsak Kingkan (Graduate School of Information Sciences)

18: Pongsate Tangseng (Graduate School of Information Sciences)

19: Nguyen Duy Kien (Graduate School of Information Sciences)

20: Naoya Chiba (Graduate School of Information Sciences)

21: Kohei Asano (Graduate School of Information Sciences)

22: Shu Kawamata (Graduate School of Information Sciences)

23: Ryo Takahashi (Graduate School of Information Sciences)

24: Zhen Wang (Graduate School of Information Sciences)

25: Yuta Watanabe (Graduate School of Information Sciences)

26: Xinmeng Li (Graduate School of Information Sciences)

27: Yinxing Li (Graduate School of Economics and Management)

28: Aijing Xing (Graduate School of Economics and Management)

29: Linh Nguyen (Graduate School of Economics and Management)

30: Mirai Igarashi (Graduate School of Economics and Management)

31: Takaki Sato (Graduate School of Economics and Management)

February 23

09:00-   **Registration**

Session 4: Economics and Management (Chair: Prof. Nobuaki Obata)

09:30-10:00   **Invited talk**: **Examining the Impact of Customer Mobile App Adoption on their Multi-Channel Purchase**
Prof. PK Kannan, Robert H. Smith School of Business, University of Maryland

10:00-10:30   **Invited talk**: **Resilience of Transportation to Extreme Weather Events: From the Big Data Perspective**
Prof. Zhenhua Chen, College of Engineering, Ohio State University

10:30-11:00   **Invited talk**: **Agglomeration Economics to be Examined by Big Data**
Prof. Dao-Zhi Zeng, Graduate School of Information Sciences, Tohoku University

11:00-11:30   **Invited talk**: **Spatio-temporal ARCH Models**
Prof. Yasumasa Matsuda, Graduate School of Economics and Management, Tohoku University

11:30-11:45   **Presentation Awards Ceremony**

11:45-   **Closing Remarks**
Prof. Mitsuyuki Nakao, Vice Dean of Graduate School of Information Sciences, Tohoku University

Day 1 Thursday, February 22
10:00 – 11:30

Session 1: Invited Talks
Medicine, Life Science and Bioinformatics

Venue: Qatar Science Campus Hall

Invited Talk

Network Biology for Advancing Drug Discovery and Human Health

Mark Chance
School of Medicine, Case Western Reserve University

Keywords: network analysis, drug discovery, Crosstalker, visualization

Network analysis methods have become commonplace research tools due to their proven ability to interrogate and organize lists of molecular targets of interest identified by basic statistics alone, and use of network analysis to refine classifier feature sets has been shown to provide superior performance compared to targets identified singly. We introduce Crosstalker as a freeware platform for academic use that is web based and incorporates multiple public interaction and gene set databases to perform network analysis, enrichment testing and visualization in a modern HTML5+JS interface. The use of open databases and algorithms coupled to convenient user choices allows cross comparison of findings and permits easy replication of results by any laboratory improving reproducibility and rigor. In addition, new –omics data sets are constantly being developed, in particular phosphoproteomics analysis provide clues to cellular signaling that can drive novel biomarker identification and identify important targets for drug development.

Invited Talk

Construction of Gene Coexpression Databases with a Viewpoint of Evolutionary Systems Biology

Takeshi Obayashi
Graduate School of Information Sciences, Tohoku University

Keywords: systems biology, evolution, gene coexpression, database

Gene coexpression information, which is a guilt-by-association approach based on a large amount of publicly available gene expression profiles, provides a fundamental view of gene functional networks. It enables various applications from gene prioritization to the delineation of global relationships among network modules within a species. In addition, comparisons of gene networks among multiple species can be used to identify conserved and specific gene modules in an evolutionary context. We have been continuously developing coexpression databases for plants (ATTED-II, http://atted.jp), animals (COXPRESdb, http://coxpresdb.jp) and microalgae (ALCOdb, http://alcodb.jp). In this talk, I present our research on gene coexpression information to understand the cellular systems encoded in a genome.

Exploring Evolutionary Potential Using Genomic Big Data

Masakado Kawata
Graduated School of Life Sciences, Tohoku University

Keywords: Genome data, evolution, evolvability

Evolution is an accumulating change in genomic contents over generations. Recent rapid increasing amount of genomic data could provide a great insight into evolutionary biology. In this talk, I will introduce how and why understanding evolutionary mechanisms is important not only for basic biology, but also for many areas of applied science such as conservation, agriculture and medicine. Then, I will show how genomic big data could make it possible to solve evolutionary problems which have remained unsolved by showing our recent research approaches using genomic data. Especially, I will focus on an unsolved issue as to why some organisms such as invasive species can evolve and adapt to various environments with high probabilities.

Day 1 Thursday, February 22
14:00 – 16:00

Session 2: Invited Talks
Data Science and Engineering

Venue: Qatar Science Campus Hall

Invited Talk

Materials Data Science: Using Distributed/High Performance Computing for Lifetime Performance of Photovoltaic Systems

Roger H. French
SDLE Research Center, Materials Science & Engineering, Case Western Reserve University

Keywords: photovoltaics, spatio-temporal modeling, network modeling, hadoop, materials science

The merger of "Peta-byte" scale Distributed Computing systems based on Hadoop/Hbase/Spark, with traditional "Peta-flop" scale High Performance Computing, has introduced a new Dist/HP Computing paradigm that enables new categories of Data Science applications. We use these large-scale data analytics techniques to investigate the degradation of photovoltaic power systems over their 25 to 50 year lifetimes and identify root cause mechanisms of degradation. Using real-world I-V and Pmp time-series datastreams from 3.4 GW of PV systems, we develop spatio-temporal models of their power rate of change (ROC). Cross-correlation of lab-based exposure testing with real-world degradation, using <Stress|Mechanism|Response> network modeling, then allows us to identify mechanisms causing real-world degradation. Cross-correlation of data science models for outdoor/real-world performance and indoor/accelerated exposures is based on a cross-correlation scale factor, which re-normalizes the indoor time base to the real-world time by least-squares fitting the indoor model to the real-world model. We also use image processing and machine learning of electroluminescence images from Silicon PV cells to quantify degradation and electrical performance over lifetime. These materials data science methods are enabling in-depth quantitative understanding of real-world performance.

Invited Talk

Research Infrastructures for Data Science

Ramin Yahyapour
GWDG, University of Göttingen, Germany

Keywords: data science, parallel computing, research infrastructure

Research infrastructures have always been evolving to accommodate the needs of the researchers. From clusters to high-performance computing and the cloud, changes in architectures enabled new possibilities for research, previously though impossible. Now, with the advent of "Big Data", a number of modern tools and architectures emerged to solve these new challenges. Access to high performance computing and data analytics infrastructure is a crucial element for many scientific disciplines. Similarly, access to data has been recognized as an essential aspect for competitive research. As such, establishing research data infrastructures are strategic goals for European as well as German activities. It is a challenge to find suitable organizational and structural configurations that enables many users to apply data analytics in an effective way. It requires the combination of method, domain and infrastructure experts to work jointly on a particular problem. The talk gives examples for current use-cases and provides input on the discussion how data science can be introduced to different scientific disciplines.

Invited Talk

Systems Science of Animal Navigation: Trajectory Data Mining and Neural Control of Movement

Koichi Hashimoto

AI center, Graduate School of Information Sciences, Tohoku University

Keywords: logbot, trajectory classification, robot microscope, optgenetics

We started a Grants-in-Aid for Scientific Research project "Systems Science of Bio-Navigation" from 2016. Navigation is a fundamental behavior of animals including human. In navigation, the following three functions are required: the acquisition of dynamically-changing information from external and internal environment, the choice of route and destination based on the information, and the behavioral regulation to reach the destination. We aim for systems science of bio-navigation to understand the "algorithms" for the navigation of animals. To this end, we bring together experts from control engineering, data science, animal ecology, and neuroscience; and jointly work on how to measure, analyze, understand, and verify bio-navigation. In this talk I introduce several examples of this project: logging device, trajectory data mining of birds, and optgentetic feedback control of animal movement.

Invited Talk

Data Analysis for Precise Smart Agriculture based on Deep Learning Algorithms

Nen-Fu Huang

Distinguished Professor and Dean, College of Electrical Engineering and Computer Science, National Tsing Hua University, Taiwan

Keywords: deep learning, smart agriculture, precise agriculture, IoT

In this talk I will introduce the IoT-based precise smart agriculture service platform in Taiwan. The service platform includes IoT-based sensor-hub data collections and transmission technologies (LoRa/NB-IoT), AI-based data analyzing technologies (deep learning algorithms), agriculture experts knowledge database, cloud-based automatic control technologies, IoT/Bigdata/AI/FinTech online and offline (OnO) MOOC courses for talent training, art technologies for smart health farms, and blockchain for agriculture. We will also demo real system deployments on several farms with high economic value fruits, including dragon fruit and coffee farm.

Day 1 Thursday, February 22
16:15 – 20:00

Session 3:
Poster presentation and Get-together party

Venue:
Qatar Science Campus Hall (Poster preview)
Restaurant DOCK (Poster presentation and
Get-together party)

01: Alterations of Neural Activity in Monkeys Induced by Repetitive Transcranial Magnetic Stimulation (rTMS) – an ECoG Study

Kentaro Ogawa[1], Shinya Nakamura[1], Takayuki Hosokawa[1], Toshio Iijima[1], Yukio Nishimura[2], Ken-Ichiro Tsutsui[1]

[1]Laboratory of Systems Neuroscience, Tohoku University Graduate School of Life Sciences, [2]Neural Prosthesis Project, Tokyo Metropolitan Institute of Medical Science

Keywords: brain stimulation, brain signal analysis, frequency analysis

Repetitive transcranial magnetic stimulation (rTMS) is a non-invasive method for stimulating the brain. It is widely believed that high-frequency rTMS facilitates the neural activity, whereas low-frequency rTMS suppresses. However, little is known about the working mechanisms of rTMS. In this study, we recorded signals from electrode array on the surface of the brain (electrocorticogram; ECoG) of monkeys before and after rTMS. Frequency analysis revealed increase in high-gamma band activity after high-frequency rTMS and decrease in beta band activity after low-frequency rTMS. These results suggest that different mechanisms support the facilitatory and inhibitory effects of rTMS.

02: Dopamine and Raf Signaling for Long Term Memory in Fruit Fly

Huan Sun

Department of Developmental Biology and Neurosciences, Graduate School of Life Sciences, Tohoku University

Keywords: dopamine, RNAi, learning and memory

Drosophila melanogaster, known as the fruit fly, has been successfully established as a model organism to study the mechanism of learning and memory. In the olfactory learning of *Drosophila*, flies can associate an unconditioned stimulus (US) reinforcement such as punishment or reward, with an odor as a conditioned stimulus (CS), to form an aversive or appetitive memory, respectively. The convergence of US and CS occurs in the mushroom body (MB). It has been demonstrated that both aversive and appetitive reinforcements to the MB are highly relying on dopamine (Aso et al., 2010; Liu et al., 2012). However, little is known about the intracellular signaling downstream to dopamine involved in learning and memory. To identify the critical molecules, I tested 38 *Drosophila* genes whose homologues in mouse are phosphorylated or dephosphorylated upon application of D1 or D2 agonists (Nagai et al., 2016). I knocked down these genes in the MB by RNAi technics and measured the aversive and appetitive memory performance, respectively. 28 of them showed behavioral phenotypes in either aversive or appetitive memory. To correlate the results of screening to the dopamine receptors in behavioral level, I also systematically assessed the role of the four *Drosophila* dopamine receptors in olfactory learning. Furthermore, I found that protein Raf is specifically required in the appetitive long term memory, and phosphorylated upon learning. Therefore, I not only identified many candidate genes that can be the basis of learning but also suggest that phosphorylation of Raf is a critical step in appetitive LTM formation.

03: Anti-viral Activities of Lectins, PPL and DB1, Isolated from Pteria Penguin Pearl Shell and *Dioscoria batata*

Lakudzala Agness Ethel, Matsui Takashi, Tanaka Yoshikazu, Ogawa Tomohisa
Applied Biological Molecular Science Laboratory, Graduate School of Life Sciences, Tohoku University

Keywords: anti-virus, carbohydrate-binding, HIV-gp120, lectin, recombinant expression

Lectins are a group of proteins that recognize carbohydrates specifically. Some lectins inhibit HIV-1 entry by binding to carbohydrate structures of the viral envelope glycoprotein (gp120). Compared to other plant lectins, there is relatively little information about animal lectins, particularly with respect to their anti-viral activity. In this study we comprehensively analyze carbohydrate binding specificities of marine invertebrate lectins, PPLs. Based on the glycan array analysis the carbohydrate specificities of PPL2A and PPL3 showed the similar profiles that bind glycoproteins. Furthermore, we constructed the expression system of GFP-conjugated gp120 in mammalian cells for analyzing the interaction between lectins and gp120.

04: Functions of Rho-GEF, PLEKHG4B in Actin Cytoskeletal Remodeling and Cell-cell Junction Formation

Komaki Ninomiya, Kensaku Mizuno, Kazumasa Ohashi
Department of Biomolecular Sciences, Graduate School of Life Sciences, Tohoku University

Keywords: Rho GEF, Rho-GTPase, actin cytoskeleton, cell-cell junction

Rho-GEFs - Rho-GTPases pathways show functional diversity and orchestrate actin cytoskeletal dynamics. Precise remodeling of actin cytoskeleton is essential for numerous cellular events, such as embryogenesis, organogenesis and tumorigenesis. Thus, much attention has been paid to the study of signaling pathways including Rho-GEFs and Rho GTPases. In this study, performing biochemical and morphological analysis, I revealed that a Rho-GEF, PLEKHG4B induces unique actin structures and plays a crucial role on cell-cell junction in epithelial cells. Furthermore, I show that PLEKHG4B regulates the two Rho GTPases, Rac1 and Cdc42, on this role.

05: Importance of Temporal Scales in the Variation of the Phenology of Zooplankton

Ishara U Perera, Jotaro Urabe, Wataru Makino, Natsumi Maruoka
Community and Ecosystem Ecology Laboratory, Graduate School of Life Sciences, Tohoku University

Keywords: temporal variation, phenology, zooplankton

Plankton are often viewed as sensitive to environmental changes. However, their sensitivity may change depending on time scales. Here, we analyzed the magnitude of temporal changes in a zooplankton community at three different scales; month, season and year, to examine the phenological stability and time resolution in their environmental responses. Statistical tests showed that monthly and seasonal variations in the community was higher than annual variation, and that seston P and temperature were selected as the best external variables explaining these variations. With other analyses, we discuss implications of the temporal variations in plankton communities in relation with ecosystem stability.

06: Effect of Arbuscular Mycorrhizal Fungi Inoculation on the Root Fungal and Bacterial Communities of Bunching Onion

Turgut Yigit Akyol[1], Rieko Niwa[2], Hideki Hirakawa[3], Shigenobu Yoshida[2], Tatsuhiro Ezawa[4], Shusei Sato[1,3]
[1]Graduate School of Life Sciences, Tohoku University, [2]Central Region Agricultural Research Center, National Agriculture and Food Research Organization (NARO), [3]Kazusa DNA Research Institute, [4]Graduate School of Agriculture, Hokkaido University

Keywords: root microbiome, arbuscular mycorrhizal fungi, bunching onion, AMF inoculation, next-generation sequencing

Root microbiome is crucial for plant health. Arbuscular mycorrhizal fungi (AMF) are important members of root microbiome as they provide soil nutrients (particularly phosphorus) to the plant. AMF have been used as biofertilizers for sustainable agriculture. In this study, we used high-throughput community analysis to perform a combined investigation of root bacteria and fungi of AMF-inoculated bunching onion (*Allium fistulosum* L.) grown under field conditions. We aimed to reveal the effect of AMF inoculation on the root native bacterial and fungal communities. Additionally, the associations between the inoculated AMF species and native microbial communities were explored.

07: Evolutionary Mechanism of Psychiatric Disorders: Interactive Effects of a Gene and Natural Disasters on Depression

Daiki Sato, Masakado Kawata
Graduate School of Life Sciences, Tohoku University

Keywords: human evolution, GxE interaction, VMAT1, the Great East Japan earthquake, ToMMo

Evolutionary mechanism to keep psychiatric disorders in population has been an attracting issue in a broad range of medical and biological sciences. In a recent study (Sato & Kawata, *submitted*), we have found that a psychiatric-disorders-relevant-polymorphism Thr136Ile in *VMAT1* gene, which encodes vesicular monoamine transporter 1, has been positively maintained by a certain sort of selective pressure (balancing selection) during the human evolution. In the present study, we aimed to reveal the mechanism of the balancing selection and investigated its context (environment)-dependent genotypic effects on depressive symptoms using massive dataset collected by Tohoku Medical Megabank Organization (ToMMo).

08: Echo-based orientation and high-frequency calls in the long-clawed shrew *Sorex unguiculatus* (Eulipotyphla: Soricidae)

Lida Sanchez Sanchez[1], Shinichiro Maruyama[1], Satoshi Ohdachi[2], Atsushi Kawahara[3], Masakado Kawata[1]
[1] Graduated School of Life Sciences, Tohoku University, [2] Hokkaido University, [3] Hokkaido Regional Environment Office

Keywords: *Sorex unguiculatus*, echo-based orientation, tonal calls

Echo-based orientation has been proposed for shrew species as an orientation system to obtain additional information when surveying their environments. Nevertheless, no specifications regarding which variables of the vocalizations are used or in what kind of behavior shrews are more likely to emit these sounds, have been reported on this system. In this study, we conducted behavioral experiments with acoustic recordings on the Japanese long-clawed shrew species, *Sorex unguiculatus,* to test how this orientation system could work. Three experimental conditions were set up and in two of them, obstacles were presented to shrew individuals. Automatic detection of recorded calls within the sound files and extraction of their acoustic variables were performed using R packages. Short click, noisy and tonal calls were detected, ranging from the audible to the ultrasonic frequency ranges. Individuals emitted more significantly frequently these vocalizations when they faced obstacles or engaged in the exploring behavior. Among tonal calls, ultrasonic calls were frequently emitted when animals faced obstacles. On the other hand, call duration and dominant frequency values were not significantly varied between animals facing the obstacles and those showing other behaviors. Our data showed that uniform and monotonal calls were used when the animals sensed obstacles within a certain distance. This suggests that *S. unguiculatus* employed a simple echo-based orientation system to support their somatosensory system with acoustical information, unlike more complex orientation systems found in echolocating animals such as bats and dolphins.

09: Microbiome Profile-based Accurate Phenotype Prediction with Small Size Training Data

Yuichi Aoki[1,2], Takeshi Obayashi[2], Kengo Kinoshita[1,2]
[1]Tohoku Medical Megabank Organization, Tohoku University, [2]Graduate School of Information Sciences, Tohoku University

Keywords: microbiome, phenotype prediction, feature clustering, positive and unlabeled learning

The microbiome state around animal internal organs or plant roots in soil greatly affect the host phenotype, such as human disease or crop yields. Although the recent advances in next-generation sequencing methods have produced a numerous number of microbiome profiling data, the higher dimensionality of microbiome profile and the limited training data size due to the missing of corresponding phenotypic data prevent the development of accurate and robust phenotype prediction models. To overcome this limitation, we have tried to explore an effective machine learning approach by applying the feature clustering technique and the positive and unlabeled (PU) learning algorithm.

10: Effects that Decoy Sequences Bring to ChIP-Seq Analysis

Hayato Anzawa[1], Kengo Kinoshita[1,2,3]
[1]Guradate School of Information Sciences, Tohoku University, [2]Institute of Development, Aging and Cancer, Tohoku University, [3]Tohoku Medical Megabank Organization

Keywords: NGS, ChIP-Sequencing, decoy sequence

The human reference genomes that include 'decoy sequences' have been used for variant calling assay in order to obtain more accurate results. Although decoy sequences compensate for the human reference genome incompleteness, it is not applied to any next generation sequencing (NGS) based assays commonly other than variant calling. Here we report the effects of decoy sequences on Chromatin immunoprecipitation (ChIP) followed by high-throughput DNA sequencing (ChIP-Seq) analysis. About 4% reads were influenced by decoy sequence introduction. However, many of these reads had low mapping quality scores. Our observation suggests decoy sequences are only partially effective for ChIP-Seq data but reinforce the importance of quality control using mapping quality.

11: Initial Analysis of the Japanese IBS Population Using a Newly Developed iPhone Application

Yuki Kagaya[1], Yukari Tanaka[2,3], Kenji Tagawa[1], Haruki Hirata[1], Shin Fukudo[3], Kengo Kinoshita[1,2]

[1]Department of Applied Information Science, Tohoku University Graduate School of Information Sciences, [2]Tohoku Medical Megabank Organization, Tohoku University, [3]Department of Behavial Medicine, Tohoku University Graduate School of Medicine

Keywords: irritable bowel syndrome, iPhone application, disease analysis with smartphone

Irritable bowel syndrome (IBS) is a group of symptoms that is thought to be common among young people in developed countries, and it is considered that stresses are some of the major causes. Although it is estimated that 10~15% of the Japanese are suffering from IBS, it is a little-known disease because of its complexity of the symptoms and difficulty of diagnosis. In this study, by developing an app which investigates IBS, we are conducting a nationwide survey for the disease using smartphones. We also show the initial analysis of the data obtained from the Japanese population using the app.

12: De Novo Protein Profile Generation with LSTM Network

Kazunori D Yamada, Kengo Kinoshita

Graduate School of Information Sciences, Tohoku University

Keywords: long short-term memory, sequence alignment, software development

Amino acid sequence profiles are widely used for bioinformatics studies, such as sequence similarity searches, multiple alignments, and evolutionary analyses. Currently, many biological sequences are becoming available, and the rapidly increasing amount of sequence data emphasizes the importance of scalable generators of amino acid sequence profiles. We employed a long short-term memory network and developed a novel profile generator to construct profiles without any assumptions, except for input sequence context. Our method could generate better profiles than existing de novo profile generators on the basis of profile-sequence similarity search performance with linear calculation costs against input sequence size.

13: Sleeping Beauty Papers are Strongly Connected in Co-citation Network

Shun Kodate[1], Kengo Kinoshita[1, 2, 3]
[1]Graduate School of Information Sciences, Tohoku University, [2]Tohoku Medical Megabank Organization, Tohoku University, [3]Institute of Development, Aging and Cancer, Tohoku University

Keywords: citation analysis, bibliometrics, scientometrics, science of science, computational social science

A Sleeping Beauty (SB) refers to a scientific paper, which is not much cited after its first publication and then suddenly gets a lot of citations. Prior research suggests SBs are not a few exceptions, but the general mechanisms of generating and awakening them are still unknown. We investigated SBs using co-citation network analysis and found they were strongly connected in the network than normal papers. The results suggest that even though SBs existed in large amounts, it would be diffusion of a small number of concepts or ideas that produced them.

14: Prediction of Any Disease-associated Genes Using Biological Literature and Mining Those Diseases Whose Predictive Performance is Higher

Yutaro Okano[1] and Kengo Kinoshita[1,2,3]
[1]Guraduate School of Information Sciences, Tohoku University, [2]Institute of Development, Aging and Cancer, Tohoku University, [3]Tohoku Medical Megabank Organization

Keywords: prediction of disease-associated genes, any disease, text mining

Predicting any disease-associated genes with the higher predictive performance is important to develop the future diagnostics and therapeutics. Previous studies focused on only specific disease and they improved the predictive performance of association between genes and the disease by using text-mining methods with existing and curated disease-gene association information such as protein-protein interactions. In this study we applied the text-mining methods to any disease, rather than focusing on a specific disease, and attempted to mine diseases, for which the text-mining methods were of higher performance.

15: Health Index Formation based on Analysis of Accelerometer Information

Siwalee Choilek, Norihiro Katayama, Mitsuyuki Nakao
Laboratory of Biomodelling, Department of Applied Information Sciences, Graduate School of Information Sciences, Tohoku University

The goal of this study is to investigate relationships between involuntary movements while sleeping and human's health condition by using wrist-worn tri-axis accelerometer. In comparison to other methods (Polysomnography, waist-worn accelerometer), wrist-worn accelerometer is more cost-friendly and gives more natural recording condition. Two-weeks of accelerometer data and responses of sleep questionnaire were collected from different subject groups (workers and students). Movement detection algorithm with multiple time-scale analysis is implemented to capture movements correctly. Current results obtained from correlation analysis shows that significant relationships between parameters obtained from modelling of distribution of movement intervals during sleep and subjective sleep questionnaire existed.

16: Machine Learning Enhanced Novel Sensing with Smart Toy Blocks for Children's Action Recognition

Xiyue Wang[1], Kazuki Takashima[1], Tamoaki Adachi[2], Yoshifumi Kitamura[1]
[1]Research Institute of Electrical Communication, Graduate School of Information Sciences, Tohoku University, [2]Department of Education, Miyagi Gakuin Women's University

We propose a machine learning approach for enhancing the sensing capability of a novel toy blocks interface. The blocks are designed to sense motions and the touch from hand or from each other. However, the action recognition rate of our customized sensing interface suffers from data lost and overflow. In this poster, we present an exploratory study that using machine learning methods to enhance the performance of actions recognition, as well as a comparison of the action recognition performances with and without the machine-learning approach.

17: Generating Mesh-base Shape from Learned Latent Spaces of Point Clouds with VAE-GAN

Cherdsak Kingkan, Koichi Hashimoto
Intelligent Control Systems Laboratory, Graduate School of Information Sciences, Tohoku University

Keywords: deep Learning, 3D Reconstruction, 3D Vision

We propose a framework that generates mesh-based objects from point clouds in an end-to-end manner by using a combination of variational autoencoder and generative adversarial network. Instead of converting point cloud to other representations like voxels before input into the network, our network directly consumes the point cloud and generates the corresponding 3D object. Given point clouds of objects, our network encodes local and global geometry structures of point clouds into latent representations. These latent vectors are then leveraged to generate Signed Distance Functions (SDF) of objects corresponding to those point clouds. SDF preserves the inside-outside information of objects. Then we can easily reconstruct polygon mesh surfaces of objects.

18: Recommending Outfits from an Item Pool

Pongsate Tangseng[1], Kota Yamaguchi[2], Takayuki Okatani[1,3]
[1]Computer Vision Laboratory, Graduate School of Information Sciences, Tohoku University, [2]CyberAgent, Inc., [3]RIKEN Center for Advanced Intelligence Project

Keywords: artificial intelligence, computer vision for fashion, recommendation system

Our work is to recommend outfits from a pool of fashion items. The challenges are that the features of items are complex, the quality of outfit is subjective, and the number of possible solution is exponential. We create a system that learn from a large dataset to extract item features and judge the outfits, and use beam search to reduce the complexity. We create a large dataset from a fashion sharing website for both training and evaluation. According to our evaluation, from a pool of apparently random items, our recommender can create outfits that look significantly better than randomly-created ones.

19: Improved Fusion of Visual and Language Representations by Dense Symmetric Co-Attention for Visual Question Answering

Nguyen Duy Kien, Takayuki Okatani
Computer Vision laboratory, Graduate School of Information Sciences, Tohoku University

Keywords: multimodal neural network, attention mechanism, visual question answering

A key solution to visual question answering exists in how to fuse visual and language features extracted from an input image and question. We show that an attention mechanism that enables dense, bi-directional interactions between the two modalities contributes to boost accuracy of prediction of answers. Specifically, we present a simple architecture that is fully symmetric between visual and language representations. It can be stacked to form a hierarchy for multi-step interactions between an image-question pair. We show through experiments that the proposed architecture achieves a new state-of-the-art on VQA datasets despite its small size.

20: Fast Light Transport Matrix Estimation for 3D Shape Acquisition

Naoya Chiba, Koichi Hashimoto
Graduate School of Information Sciences, Tohoku University

Keywords: 3D shape estimation, light transport matrix, compressive sensing

Various 3D measurement methods which use a projector-camera system have been proposed; however, there is still difficulty measuring metallic objects and translucent objects. We use Light Transport (LT) Matrix, which includes all camera response for each projector pixels, as an optical model. LT Matrix makes it easy to measure 3D shape of various objects because of its information richness. We developed fast LT Matrix estimation method which uses compressive sensing technique. We also modify a compressive sensing algorithm ADMM (Alternating Direction Method of Multipliers) to adapt to the light saturation on camera pixels.

21: Challenge to Long-Term Prediction of Electricity Load

Kohei Asano, Kenya Kikuchi, Quentin Labernia, Yashio Kabashima, Takeshi Tokuyama
Graduate School of Information Sciences, Tohoku University

Keywords: curve regression, machine learning, forecasting, electricity loads

Last September, our student team joined the Electricity Load Forecasting Technology Contest organized by TEPCO. The contest requires prediction of hourly electric power consumption for the next two weeks. The hourly electric power consumption data for recent 8 years was provided by TEPCO, and participants can use public data such as weather etc. We decomposed the power consumption data into signals of three ranges of frequency: seasonal, daily and hourly. Each extracted signal was then analyzed by using machine learning and optimization methods. Although we could not have good scores in the contest, we obtained several experience through it.

22: Online Behavior Inference of Rescue Dogs from the Inertial Data

Shu Kawamata, Syunta Shimizu, Takeshi Tokuyama
Graduate School of Information Sciences, Tohoku University

Keywords: data analysis, machine learning, supervised learning, online estimation, multi label classification

We propose a novel construction of a system for guessing the behavior of rescue dogs while they perform searching for victims in a disaster site out of field of view of handlers. Our system infers what rescue dogs are currently doing from the inertial data obtained from acceleration sensors attached to the dogs. Precisely speaking, our system automatically finds suitable actions of the dog selected from a candidate set of "running, walking, stopping, barking, sniffing some objects or sniffing the air", and visualize it to the handlers and rescue supervisors. We have labeled batches of inertial data in the past and analyze this problem as a multi label classification problem to apply machine learning tools. We hope this research will empower the ability of rescue dogs in disaster relief tasks.

23: Knowledge Sharing by Joint Training with Autoencoders

Ryo Takahashi[1], Tian Ran[1], Kentaro Inui[1,2]
[1]Graduate School of Information Sciences, Tohoku University, [2]RIKEN Center for Advanced Intelligence Project (AIP)

Keywords: machine learning, knowledge base, information extraction

Knowledge bases (KBs) storing real-world facts about entities and their relationships are useful resources for various natural language processing tasks. However, since KBs are typically incomplete, it is helpful to be able to perform KB completion, i.e., predict whether a relationship not in the knowledge base is likely to be true. Embedding models for KB completion associate entities and relations with dense feature vectors or matrices. One of the research issues for such models is knowledge sharing among relations. We propose a framework to promote knowledge sharing by jointly training an embedding model and an autoencoder for relations.

24: Auto-tuning of Hyperparameters of Machine Learning Models

Zhen Wang
Takizawa lab, Graduate School of Information Sciences, Tohoku University

Keywords: machine learning, hyperparameter, auto-tuning

Most machine learning models use hyperparameters empirically defined in advance of their training processes. Even a classic machine learning model, so-called multilayer perceptrons, has a lot of hyperparameters. In the case of using such a model for a classification problem, one difficulty is that the achieved classification accuracy could drastically change every time even if the same hyperparameter values are used. Hence, it is challenging to determine an appropriate hyperparameter configuration at a low cost. The same problem has so far been discussed in the research field of software performance automatic tuning, or auto-tuning. Therefore, in this work, we employ one of such auto-tuning mechanisms, ATMathCoreLib, for auto-tuning hyperparameters of machine learning models, and discuss the feasibility of using such technologies in the field of machine learning.

25: On Algebraic Properties of the Combinatorial Structure of a Finite Projective Geometry

Yuta Watanabe
Graduate School of Information Sciences, Tohoku University

Keywords: projective geometry, incidence algebra, quantum algebra

It is known that the incidence algebra of a finite projective geometry is known to be a homomorphic image of the quantum algebra U_q(sl_2). We strengthen the connection between finite projective geometries and quantum algebras. We consider two natural extensions of the incidence algebra of a finite projective geometry by taking its local structures into account because the incidence algebra captures only its global structure. Then we show that each of new algebras extends the above relation to the level of the quantum affine algebra U_q(sl_2^).

26: A Footloose Capital Model and Labor Market Frictions

Xinmeng Li, Dao-Zhi Zeng
Graduate School of Information Sciences, Tohoku University

Keywords: unemployment, labor market frictions, capital, fixed cost

To investigate how labor product market, capital market and labor market interact among each other, we present a generalized version of a 2-country model with footloose capital incorporating Pissarides (2000) canonical model of equilibrium unemployment. We show that the unemployment rate curve in large country is U-shape relative to the trade freeness. Moreover, the unemployment rates in 2 countries increase with fixed cost in domestic market and decrease with fixed cost in exported market.

27: Social Media and the Diffusion of an Information Technology Product

Yinxing Li, Nobuhiko Terui
Graduate School of Economics and Management, Tohoku University

The expansion of the Internet has led to a huge amount of information posted by consumers online through social media platforms such as forums, blogs, and product reviews. This study proposes a diffusion model that accommodates pre-launch social media information and combines it with post-launch sales information in the Bass model to improve the accuracy of sales forecasts. The model is characterized as the extended Bass model, with time varying parameters whose evolutions are affected by the consumer's communications in social media. Specifically, we construct variables from social media by using sentiment analysis and topic analysis.

28: Interpretable Perceived Topics in Online Customer Reviews and Product Management

Aijing Xing , Nobuhiko Terui
Graduate School of Economics and Management, Tohoku University

Online customer reviews contain useful and important information for product management—because customers tend to praise or criticize certain features or attributes of goods in their reviews. We propose a model that provides interpretable perceived topics by assigning a priori labeled words to respective topics and connects them to satisfaction and helpfulness variations by supervised learning. The empirical analysis on user-generated content of food reviews shows that our proposed model performs better than alternative models, and it suggests product managers the necessity of improving some specific attributes and focus their advertising on these attributes as fulfilling customer needs.

29: Data Quality Management of Chain Stores based on Outlier Detection

Linh Nguyen, Tsukasa Ishigaki
Graduate School of Economics and Management, Tohoku University

Keywords: data quality management, outlier detection, isolation forest

In this paper, we present a data quality evaluation method for shops in chain businesses based on outlier detection and then, we apply this method to a dataset observed in real chain stores, which provide tire maintenance for vehicles. To evaluate the data quality of each shop, we use data about truck's tire information such as tread depth, tread pattern and distance, which was recorded by shop at maintenance time. The proposed method can support better maintenance services for customer as well as be able to get more correct data from these shops, which will be useful for next research.

30: A Model for Detecting Influencers in Social Media

Mirai Igarashi, Nobuhiko Terui
Graduate School of Economics and Management, Tohoku University

Keywords: topic models, influencers, text analysis, network models, social networks

In modern social media development, viral marketing, which aims at efficient information diffusion through word-of-mouth by socially influential people, or influencers, is an important area of study. The author proposes a new form of topic model that simultaneously analyzes the network structure representing the relationships between users and the text information posted on social media to detect the influencers and understand their interests. As a result, the proposed model succeeds in identifying the interests of users with a network structure that identifies an influencer, that is, revealing the influencer's image, which was not disclosed in the previous research.

31: Spatial GARCH models

Takaki Sato, Yasumasa Matsuda
Graduate School of Economics and Management, Tohoku University

This study proposes a spatial extension of time series generalized autoregressive conditional heteroscedasticity (GARCH) models. We call the spatial extended GARCH models as spatial GARCH (S-GARCH) models. S-GARCH models specify conditional variances given simultaneous observations and conditional variance, which constitutes a good contrast with time series GARCH models that specify conditional variances given past observations and conditional variance. S-GARCH models are transformed into Spatial autoregressive moving-average (ARMA) models and the parameters of S-GARCH models are estimated by the quasi maximum likelihood (QML) estimation method. We prove consistency and asymptotic normality of the proposed QML estimators and demonstrate empirical properties by real data analysis of land price data in Kanto area.

Day 2 Friday, February 23
09:30 – 11:30

Session 4: Invited Talks
Economics and Management

Venue: Qatar Science Campus Hall

Invited Talk

Examining the Impact of Customer Mobile App Adoption on their Multi-Channel Purchase

P. K. Kannan

Dean's Chair, Smith School of Business, University of Maryland

Keywords: mobile apps, multichannel, big data, difference-in-difference, Heckman selection

Mobile apps have become more than an additional channel for firms to engage with customers, provide services, and, according to conventional wisdom, positively impact customer spending and loyalty. This paper focus on the dark side of customer adoption of mobile apps. Using data from a leading hospitality firm covering a four-year period, we investigate customer purchase behavior before and after mobile app adoption of those customers, who adopt the app at different times. Our analysis uses propensity score method to match customers based on observables, uses different cohort matching to control for time invariant unobservables contributing to adoption and difference-in-differences model to account for time variant unobservables that are common to all individuals. We also combine difference-in-differences specification with Heckman-style selection model to control for self-selection problem. Our analysis indicates that app adoption has a negative impact on average spending per customer in this specific context. Results of a survey conducted with the firm's and competitors' app adopters suggest that early adopters of the app are also highly likely to adopt competitors' apps, to shop around using apps, and thus become less loyal to the focal firm, thus reducing their share of wallet.

Invited Talk

Resilience of Transportation to Extreme Weather Events: From the Big Data Perspective

Zhenhua Chen

City and Regional Planning, College of Engineering, The Ohio State University

Keywords: aviation, high-speed rail, resilience, extreme weather, big data

Climate change has made transportation systems become vulnerable given the rising extreme weather events, such as typhoon and winter storms. One direct consequence is that on-time performance of various transportation systems turns out to be much unpredictable due to the uncertainty of these extreme weather events. The objective of this research is to investigate a fundamental question related to the operational reliability of two transportation modes, aviation and high speed rail: how resilient are these systems to extreme weather events? Based on the analytics using artificial intelligence for a dataset that includes 5.7 million flight records and over 100 thousands of high-speed rail operational records for the period October 2016 - September 2017, the study reveals that although HSR tends to be more resilient than aviation to extreme weather events in general, the operational performance of HSR tends to be spatially different. Specifically, the HSR system operated in the South China tends to be less resilient to typhoon, thunderstorm and heavy rainfall, whereas the system operated in the Central and East China is less resilient to snowstorm.

Invited Talk

Agglomeration Economics to be Examined by Big Data

Dao-Zhi Zeng

Social-Economic Analysis of Urban System, Graduate School of Information Sciences, Tohoku University

Keywords: home market effects, spatial economics, consumer preferences

As a milestone result in spatial economics, the home market effects (HME) conclude that a manufacturing sector agglomerates in large regions (countries) and that the wage rates in larger regions are higher. A CES utility function is assumed to theoretically derive the HME. Applying a HARA utility function, which is more general than the CES, we find that the HME results crucially depend on the utility parameters. Thus, exact information about consumers' preferences in an economic space becomes important to disclose the economic structure there. An empirical study on consumers' preferences is possible if the consumer data (across regions and countries) is rich enough and available.

Spatio-temporal ARCH Models

Yasumasa Matsuda

Graduate School of Economics and Management, Tohoku University

Keywords: ARCH model, time series analysis, curse of dimension, stock price

Autoregressive Cobditional Heteroscadsticity (ARCH) model has been attracting much attention as one of the most influential models in financial time series analysis, since Robert Engle introduced it in 1982 to evaluate volatility of stock prices. In this talk, we review some brief history of ARCH models with the introduction, and try a multivariate extension that can evaluate volatility matrix for several tens of stocks. One of the main difficulty of multivariate extensions is in "curse of dimension", which means the number of parameters grow explosively as the dimension of stocks becomes larger, a typical problem in big data analysis. Our idea to tackle the difficulty is to employ spatial models to reduce the number of parameters, where spatial weight matrix plays a critical role to avoid the curse of dimension. Spatial weight matrix is originally the one with geographical distance among observation points of spatial data. We generalize it to that for financial time series by introducing a distance among stocks. We demonstrate the multivariate extension by applying to daily stock price series of topix core 30 in the period of 2008 financial crisis.