

DMSS 2006 Workshop Program

The International Workshop on Data-Mining and Statistical Science (September 25-26, 2006, Century Royal Hotel, Sapporo, Japan)
JSAI SIG-DMSM, JSAI SIG-FPAI, and MEXT Grant-in-aid for Specially Promoted Research on "Semistructured Data Mining"

Monday, September 25

Monday Morning, Room A (20F, "Hakucho / Swan")

9:00 Opening

9:00 - 9:50 **Session I1** (Monday Morning, Room A)

Invited Talk by Koji Tsuda (MPI for Biological Cybernetics.)

Graph Mining Applications to Machine Learning Problems [abstract]
Koji Tsuda

9:50-10:00 Break

10:00 - 10:50 **Session I2** (Monday Morning, Room A)

Invited Talk by Alexandre Termier (The Institute of Statistical Mathematics)

From Closed Tree Mining to Closed DAG Mining [abstract]
Alexandre Termier, Yoshinori Tamada, Seiya Imoto, Takashi Washio, Tomoyuki Higuchi

10:50-11:00 Break

11:00 - 11:50 **Session I3** (Monday Morning, Room A)

Invited Talk by Ryo Yoshida (Univ. Tokyo, IMS, Human Genome Center)

Unsupervised Learning of $n \ll p$ Data and Its Applications to Bioinformatics [abstract]
Ryo Yoshida, Seiya Imoto, Tomoyuki Higuchi, Satoru Miyano

11:50-13:20 Lunch

Monday Afternoon, Room A (20F, "Hakucho / Swan")

13:20 - 14:40 **Session A1** (Monday Afternoon, Room A)

Clustering and Unsupervised Learning

General Consistency Theorems for Proper Discrete Bayesian Learning
Jan Poland

Clustering Without Data: the Relevant-Set Correlation Model
Michael E. Houle

Clustering Without Data: the GreedyRSC Heuristic
Michael E. Houle

A Probabilistic Fuzzy Clustering Algorithm with Application to Web Document Collections
Nataliya Lamonova, Yuzuru Tanaka

14:40-15:00 Break

15:00 - 16:00 **Session A2** (Monday Afternoon, Room A)

Mining and Searching Semi-structured Data I

A Spectrum Tree Kernel

Tetsuji Kuboyama, Kouichi Hirata, Hisashi Kashima, Kiyoko F. Aoki-Kinoshita, Hiroshi Yasuda

On Designing Trust Calculation Algorithms on P2P Networks with Turnover
Kouki Yonezawa

Online Construction of Truncated Suffix Tree with Word Count Limitation
Takuya Kida, Takashi Uemura, Hiroki Arimura

16:00-16:20 Break

16:20 - 18:00 **Session A3** (Monday Afternoon, Room A)

Mining Rules and Models I

Modeling Dynamics of Massive Dimensional and Complex Systems
Viet Phuong Nguyen, Takashi Washio

Sampling for Imbalanced Data Learning
Canh Hao Nguyen, Tu Bao Ho

Parallel EM Learning for Symbolic-Statistical Models
Yusuke Izumi, Yoshitaka Kameya, Taisuke Sato

A Knowledge Discovery from POS Data using State Space Models: An Analysis of Change in Price Elasticities by New Product's Entry to Market (in Japanese)

Sato Tadahiko, Higuchi Tomoyuki

Casualty Insurance Pure Premium Estimation Using Two-Stage Regression Tree (in Japanese)
Kumiko Nishi, Ichiro Takeuchi

Monday Afternoon, Room B (03F, "Elegance")

13:20 - 14:40 **Session B1** (Monday Afternoon, Room B)

Text Mining I

Opinion Mining from Weblogs: Extraction and Structurization
Kobayashi Nozomi, Inui Kentaro, Yuji Matsumoto

Discovering Concepts from Word Cooccurrences with a Relational Model
Kenichi Kurihara, Yoshitaka Kameya, Taisuke Sato

Scientific Emerging Trend Detection
Minh-Hoang Le, Tu-Bao Ho, and Yoshiteru Nakamori

Text mining and visualization system for numerical information from a large number of documents

Masaki Murata, Koji Ichii, Qing Ma, Tamotsu Shirado, Toshiyuki Kanamaru, Sachiyo Tsukawaki, Hitoshi Isahara

14:40-15:00 Break

15:00 - 16:00 Session B2 (Monday Afternoon, Room B)

Text Mining 2

Towards Biomedical Information Extraction Using Local Alignments (in Japanese)

Kazuo Hara, Masashi Shimbo, Yuji Matsumoto

Research on Multiple Information Resource Analysis by Using Correlation Change (in Japanese)

Masaharu Yoshioka, Tsuyoshi Taniguchi, Makoto Haraguchi

An Automatic Indexing Method by Interval Estimation (in Japanese)

Shuusuke Fujii, Koichiro Doi, Akihiro Yamamoto

16:00-16:20 Break

16:20 - 17:20 Session B3 (Monday Afternoon, Room B)

Organized Session on MEXT Specially-Promoted Research on "Semi-structured Data Mining"

Semi-Structured Data Mining Project 2005-2008": Towards Knowledge Infrastructure Formation from Large Semi-Structured Data

Hiroki Arimura, Kimihito Ito, Takuya Kida, Shin-ichi Minato

Symmetric Item Set Mining Method Using Zero-suppressed BDDs and Application to Biological Data

Shin-ichi Minato, Kimihito Ito

Data Mining in Amino Acid Sequences of Human Influenza A Viruses

Kimihito Ito

19:00-21:00 Banquet (Monday, Evening, Room B - 3F, "Elegance")

Tuesday, September 26

Tuesday, Morning, Room C (20F, "Shinju / Pearl")

9:00 - 9:50 Session I4 (Tuesday, Morning, Room C)

Invited Talk by Hisashi Kashima (IBM Tokyo Research Laboratory)

Methods for Network Structure Prediction [\[abstract\]](#)

Hisashi Kashima

9:50-10:00 Break

10:00 - 10:50 Session I5 (Tuesday, Morning, Room C)

Invited Talk by Shuji Kijima (Univ. Tokyo, Dept. of Math. Informatics)

Random Sampling via Markov Chain [\[abstract\]](#)

Shuji Kijima, Tomomi Matsui

10:50-11:00 Break

11:00 - 12:00 Session C1 (Tuesday, Morning, Room C)

Mining and Searching Semi-structured Data 2

Discovering Constrained Frequent Closed Ordered Subtrees (in Japanese)

Tomonobu Ozaki, Takenao Ohkawa

Fast Reachability Test on DAGs for XML (in Japanese)

Yusaku Nakamura, Tetsuya Maita, Hiroshi Sakamoto

Anti-unification of Semi-structured Documents Compressed with Tree Grammars (in Japanese)

Koichiro Doi, Jun Onuma, Akihiro Yamamoto

12:00-13:10 Lunch

Tuesday, Afternoon, Room C (20F, "Shinju / Pearl")

13:10 - 14:30 Session C2 (Tuesday Afternoon, Room C)

Mining Rules and Models 2

Nantonac Collaborative Filtering

Toshihiro Kamishima, Shotaro Akaho

Incipient Fault Diagnosis of Pressure Regulator for High Pressure Gas by Using SVM with the Kullback-Leibler Kernel (in Japanese)

Tsukasa Ishigaki, Tomoyuki Higuchi, Kajiro Watanabe

Mining Sectorial Episodes from Bacterial Culture Data (in Japanese)

Takashi Katoh, Kouichi Hirata, Masateru Harao

Efficient Variable Selection Method for Exposure Variables on Binary Data (in Japanese)

Manabu Ohno, Tomoyuki Tarumi

14:30-14:50 Break

14:50 - 16:10 Session C3 (Tuesday Afternoon, Room C)

Novel Applications of Knowledge Discovery and Data Mining

Filter for Detecting Unknown Computer Viruses Using Graham Bayes Learning Algorithm for Spam Detection

Ryuiti Koike, Naoshi Nakaya, Yuuji Kouji

Design of discussion information sharing system between Municipalities (in Japanese)

Masahiro Ehara, Michio Ito

Natto: A Tool for Exploratory Data Analysis with Information Theoretic Approach (in Japanese)

Ryota Suzuki, Tatsuhiro Nagai, Tomoya Taniguchi

Frequent Closed Item Set Mining Based on Zero-suppressed BDDs

Shin-ichi Minato, Hiroki Arimura

16:10 Closing

Additional Information

Rooms:

Century Royal Hotel, Sapporo

Room A: 20F, "Hakucho / Swan" (Invited talks and technical talks on 9/25, Monday)

Room B: 03F, "Elegance" (Technical talks and banquet on 9/25, Monday)

Room C: 20F, "Shinju / Pearl" (Invited talks and technical talks on 9/26, Tuesday)

Presentation

Invited talk: 50min = 40min talk + 10min discussion per talk

Technical talk: 20min = 15min talk + 5min discussion per talk

PC projectors and OHPs are available at every room.

The latest information will be found on:

DMSS2006 homepage: <http://www-ikn.ist.hokudai.ac.jp/dmss2006/>

Program: <http://www-ikn.ist.hokudai.ac.jp/dmss2006/program.html>

Please contact us if you have any problem with the above program:

E-mail: arim@ist.hokudai.ac.jp
