

Preface

This volume contains the research reports of the Discovery Science project in Japan (No. 10143106), in which more than 60 scientists participated. It was a three-year project sponsored by Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan. This project mainly aimed to (1) develop new methods for knowledge discovery, (2) install network environments for knowledge discovery, and (3) establish Discovery Science as a new area of study in Computer Science / Artificial Intelligence.

In order to attain these aims we set up five groups for studying the following research areas:

- (A) Logic for/of Knowledge Discovery
- (B) Knowledge Discovery by Inference/Reasoning
- (C) Knowledge Discovery Based on Computational Learning Theory
- (D) Knowledge Discovery in Huge Databases and Data Mining
- (E) Knowledge Discovery in Network Environments

These research areas and related topics can be regarded as a preliminary definition of Discovery Science by enumeration. Thus Discovery Science ranges over philosophy, logic, reasoning, computational learning, and system developments.

In addition to these five research groups we organized a steering group for planning, adjustment, and evaluation of the project. The steering group, chaired by the principal investigator of the project, consists of leaders of the five research groups and their subgroups as well as advisors from outside of the project. We invited three scientists to consider Discovery Science and the five above mentioned research areas from viewpoints of knowledge science, natural language processing, and image processing, respectively.

Group A studied discovery from a very broad perspective, taking account of historical and social aspects, and computational and logical aspects of discovery. Group B focused on the role of inference/reasoning in knowledge discovery, and obtained many results on both theory and practice in statistical abduction, inductive logic programming, and inductive inference. Group C aimed to propose and develop computational models and methodologies for knowledge discovery mainly based on computational learning theory. This group obtained some deep theoretical results on the boosting of learning algorithms and the minimax strategy for Gaussian density estimation, and also methodologies specialized to concrete problems such as algorithms for finding best subsequence patterns, biological sequence compression algorithms, text categorization, and MDL-based compression. Group D aimed to create computational strategies for speeding up the discovery process in total. For this purpose, group D was made up of researchers from other scientific domains and researchers from computer science so that real issues in the discovery process could be exposed and practical computational techniques could be devised and tested for solving these real issues. This group handled many kinds of data: data from national projects such as genomic data and satellite observations, data generated from laboratory experiments, data collected from personal interests such as literature and medical

records, data collected in business and marketing areas, and data for proving the efficiency of algorithms such as the UCI repository. So many theoretical and practical results were obtained on such a variety of data. Group E aimed to develop a unified media system for knowledge discovery and network agents for knowledge discovery. This group obtained practical results on a new virtual materialization of DB records and scientific computations to help scientists make a scientific discovery, a convenient visualization interface that treats web data, and an efficient algorithm that extracts important information from semi-structured data in the web space.

We would like to express our immense gratitude to the members of the Discovery Science project, listed on the subsequent pages. The papers submitted to this volume were reviewed both by peers and external referees. We would like to express our sincere gratitude to the following external referees:

Jean-Francois Boulicaut	Kazunori Matsumoto
Koji Eguchi	Aart Middeldorp
Toshirou Ejima	Tsuyoshi Murata
Tatsuaki Hashimoto	Tadas Nakamura
Yutaka Hata	Kazuki Nishi
Masafumi Hirahara	Osamu Nishizawa
Koichi Hirata	Takashi Okada
Koichi Hishida	Seishi Okamoto
Takashi Iida	Ryuji Omori
Daisuke Ikeda	Chiaki Sakama
Katsumi Inoue	Hiroshi Sakamoto
Makio Ishiguro	Takafumi Sakurai
Osamu Katai	Hiroyuki Sato
Hajime Kato	Tetsuo Shibuya
Iwane Kimura	Tetsuro Shimizu
Yoshiki Kinoshita	Wataru Sunayama
Kazuo Kitahara	Jun-ichi Takeuchi
Satoshi Kobayashi	Kiyotaka Uchida
Shozo Makino	Shinji Yoshioka

We would also like to thank to the external advisors, Raymond Greenlaw, Carl Smith, and Thomas Zeugmann, for their valuable comments.

Organization

Steering Group

In addition to the five groups listed below we had a steering group (Soukatsu-han in Japanese) for planning, adjustment, and evaluation of the project, consisting of the following members:

Setsuo Arikawa (Chair, Kyushu University)
Masahiko Sato (Leader of Group A, Kyoto University)
Taisuke Sato (Leader of Group B, Tokyo Institute of Technology)
Akira Maruoka (Leader of Group C, Tohoku University)
Satoru Miyano (Leader of Group D, University of Tokyo)
Yasumasa Kanada (Leader of Group E, University of Tokyo)
Yuichiro Anzai (Keio University)
Setsuo Ohsuga (Waseda University)
Kinji Ono (NII)
Hiroakira Ono (JAIST)
Takuya Katayama (JAIST)
Yahiko Kambayashi (Kyoto University)
Tadao Saito (Chuo University)
Hidehiko Tanaka (University of Tokyo)
Yuzuru Tanaka (Hokkaido University)
Jun'ichi Tsujii (University of Tokyo)
Jun-ichoro Toriwaki (Nagoya University)
Makoto Nagao (Kyoto University)
Teruyuki Nakajima (University of Tokyo)
Shozo Makino (Tohoku University)
Keiichi Noe (Tohoku University)
Koichi Furukawa (Keio University)
Hiroshi Motoda (Osaka University)
Ayumi Shinohara (Kyushu University)

Group A: Logic of Knowledge Discovery

Masahiko Sato (Leader, Kyoto University)
Tetsuo Ida (University of Tsukuba)
Hiroakira Ono (JAIST)
Mitsuhiro Okada (Keio University)
Keiichi Noe (Tohoku University)
Masami Hagiya (University of Tokyo)
Yukiyoshi Kameyama (University of Tsukuba)
Shun Tsuchiya (Chiba University)

Group B: Knowledge Discovery by Inference

Taisuke Sato (Leader, Tokyo Institute of Technology)
Hiroki Arimura (Kyushu University)
Mutsumi Imai (Keio University)
Masako Sato (Osaka Prefecture University)
Takeshi Shinohara (Kyushu Institute of Technology)
Makoto Haraguchi (Hokkaido University)
Hiroshi Tsukimoto (Toshiba)
Chiaki Sakama (Wakayama University)
Ken Sato (Hokkaido University)
Yukio Ohsawa (University of Tsukuba)
Akihiro Yamamoto (Hokkaido University)
Koichi Furukawa (Keio University)
Hiroshi Tanaka (Tokyo Medical and Dental University)
Masahiko Yachida (Osaka University)
Katsumi Inoue (Kobe University)

Group C: Knowledge Discovery by Learning Algorithms

Akira Maruoka (Leader, Tohoku University)
Naoki Abe (C&C, NEC)
Hiroshi Imai (University of Tokyo)
Ayumi Shinohara (Kyushu University)
Atsuhiko Takasu (NII)
Osamu Watanabe (Tokyo Institute of Technology)
Eiji Takimoto (Tohoku University)
Sadao Kurohashi (Kyoto University)
Kuniaki Uehara (Kobe University)

Group D: Knowledge Discovery in Databases

Satoru Miyano (Leader, University of Tokyo)
Thoru Araki (Kyoto University)
Genshiro Kitagawa (Institute for Statistical Mathematics)
Shinichi Morishita (University of Tokyo)
Hiroshi Motoda (Osaka University)
Ryohei Nakano (Nagoya Institute of Technology)
Koichi Nijima (Kyushu University)
Katsutoshi Yada (Kansai University)
Yuji Ikeda (Kobe University)
Yoshiya Kasahara (Kyoto University)
Einoshin Suzuki (Yokohama National University)
Takehiko Tanaka (Kyushu University)

Shusaku Tsumoto (Shimane Medical University)
Hiroshi Hasegawa (Ibaragi University)
Rie Honda (Kochi University)
Kouhei Yamaguci (Shizuoka University)
Toyofumi Saito (Nagoya University)
Masayuki Takeda (Kyushu University)
Osamu Maruyama (Kyushu University)
Kenichi Yoshida (Hitachi)
Shozo Makino (Tohoku University)
Shogo Nishida (Osaka University)
Tomoyuki Higuchi (Institute for Statistical Mathematics)
Shinobu Machida (Kyoto University)

Group E: Knowledge Discovery in Network Environments

Yasumasa Kanada (Leader, University of Tokyo)
Setsuo Arikawa (Kyushu University)
Shigeki Goto (Waseda University)
Etsuya Shibayama (Tokyo Institute of Technology)
Yuji Takada (Fujitsu Lab.)
Yuzuru Tanaka (Hokkaido University)
Hiroyuki Kawano (Kyoto University)
Sachio Hirokawa (Kyushu University)
Fumio Mizoguchi (Science University of Tokyo)
Hiroki Ishizaka (Kyushu Institute of Technology)