ISSN 1883-1974
National Institute of Informatics News
No.35 Oct. 2010
(This English language edition of NII Today corresponds to No.49 of the Japanese edition)

NII Today

National Institute of Informatics News

[Special feature] **Artificial Intelligence**

From Personal to Societal Knowledge

Working to Unravel the Secrets of Biological Activity Through Inference

"Meaning" Links Information, Changing the Web

New Data Mining Made Possible by AI

Artificial Intelligence Reasons about the Courses of Trials

[Special feature]
Artificial Intelligence

From personal to societal knowledge

Computers replace human beings in leading to answers.
NII is carrying out extensive informatics-based artificial intelligence (AI) research.
The culmination of this research will be the fusion of AI and a variety of sciences, and the further evolution of Web services.
AI research now attracts considerable attention as a methodology for pursuing the "knowledge" that is vital to creating a prosperous society.

NII Interview

Working to Unravel the Secrets of Biological Activity Through Inference

Katsumi Inoue

Professor , Principles of Informatics Research Division, NII

Artificial intelligence establishes hypotheses, tests them, and confirms them

Yoshida The image computers that one may have is one of being good at performing calculations in accordance with defined rules, but being unable to handle things which do not follow rigidly predetermined lines. Professor Inoue, you're researching mechanisms which transcend that conventional wisdom — in other words, systems where computers infer, in the form of hypotheses, the rules necessary to reach a conclusion. That approach is currently being used in the field of systems biology. Could you start by telling us what it means to discover a hypothesis through inference?

Inoue A famous example of syllogism is: A "All men are mortal ", B "Socrates is a man", C "Socrates is also mortal". A is the major premise or rule, B is the minor premise, and C is the conclusion. Given general principle A, and individual fact B, syllogistic logic leads to conclusion C, which is mathematically correct. This is called deduction. If one only has A "All men are mortal", and C "Socrates is mortal", what sort of premise could be added to make this conclusion true?

Yoshida The minor premise that "Socrates is a man"?

Inoue Yes. One can infer the hypothesis that "Socrates is a man", and confirm whether or not that is correct. This is called abduction(*1). Conversely, starting with B "Socrates is a man", and C "Socrates is mortal", and discovering major premise A "All men are mortals" is called induction(*2).

Yoshida Discovering something like that seems to me to be the type of creative effort that is the domain of humans. Are you talking about a computer creating this new hypothesis?

Inoue We infer hidden relations and constraints that are logically possible, and in principle those inferred hypotheses can contain newly created theories for human beings.

Yoshida Is this field of inference-based hypothesis finding a new one?

Inoue No, it's been researched in the AI field for over 20 years. What's new about our research is that our method works well even with incomplete knowledge.

Discovered by foreign researchers, and applied to systems biology

Yoshida Professor Inoue, how did you come to be involved in the field of automated reasoning?

Inoue When I was a university student, I majored in algorithms. In order to quickly and efficiently solve problems, we need to consider the types of instructions to be used, how to combine and order them, and what procedures to follow. I was also very interested in game searching. You have to select from a massive range of continuously changing options of what to do in order to achieve your ultimate objective of victory over your opponent, while the opponent is also assumed to take the best strategy. The model for this is human intelligence.

Yoshida So you started in what could best be called the field of mathematical computation applied to intelligent search, but this served as the background to your entry into the world of logical inference

Inoue Indeed, when we consider a dynamic situation in which our knowledge and observations change from time to time, the problem that we need to solve also changes accordingly. In such a case we need to infer about past and future states using a more intelligent methodology. I thus became interested in basic theories of reasoning under incomplete knowledge.

Yoshida When did you start on the path to your current research?

Inoue It was around 2001. The hypothesis finding method that I introduced in 1991 was noticed by a British post-genomic research group, who asked me some questions. They were involved in some ambitious research, using a robot to generate hypotheses regarding gene functions, plan experiments based on those hypotheses, perform the experiments, and produce feedback based on those results, all automatically. They believed that my method could be used as the theoretical background for their hypothesis generation. I later had similar inquiries from French researchers, who I performed joint research with. We began full-fledged application of our "SOLAR(*3)" method. The theme we chose was identifying which enzymatic reactions in metabolic pathways are accelerated and which others are inhibited

Yoshida There are many computer learning and hypothesis finding techniques. What sets SOLAR apart?

Inoue It can produce logically possible hypotheses for a wide range of problem classes. It is especially unique in that it possesses features for filling in missing knowledge or data in the form of hypotheses. In order to do this, computationally efficient methods to generate candidate hypotheses must be developed. When arriving at conclusions from incomplete premises, there are many conceivable candidates which should be filled in. However, checking each one, one by one, would take a tremendous amount of time. In some cases, the number of candidates explodes, and cannot be narrowed down within a practical time. In order to avoid this situation, we incorporate a statistical hypothesis evaluation method so that only top ranked hypotheses are verified in subsequent experiments.

Yoshida Computer technologies such as algorithms and game search concepts are utilized in a comprehensive manner. So this SOLAR system is used in systems biology, but what exactly is systems biology?

Searching for the missing link of biological activity

Inoue Systems biology is concerned with an integrated understanding of the different interrelated elements that make up life activities, which it treats as a single system.

Yoshida Biology has traditionally taken a functional explanation approach, saying that mitochondria are related to cell respiration, or that ribosomes copy DNA and create new proteins.

Inoue Biology has until now focused on separating individual structures and parts, and explaining what they do, and this is still the primary approach used in bio research. The goal of systems biology, on the other hand, is exploring the relationships between these separate components, which are then put back together again into a cohesive whole.

Yoshida It takes different levels of biological components, both large and small—organs, the cells that make them up, nerves, DNA, enzyme, proteins and metabolites—and looks at them as a single system?

Inoue Yes, but gathering everything into a single entity is not immediately feasible. Instead, for, say, a single function, it looks at how its component elements behave, how they interrelate, and what function is produced through those interrelationships. For example, within cells, signals are conveyed to trigger metabolic changes or reactions, such as respiration, or to suppress reactions, and gene expression and inhibition are controlled. These function as a complex network composed of many individual elements. Some elements also belong to other networks, and these networks affect each other in a complex layered structure. Systems biology models how individual elements interrelate and behave in a biological system, and uses computer simulations to verify these models.

Yoshida How are computer used effectively in systems biology?

Inoue We are discovering a great deal about how genes work, and producing an enormous amount of data. Computers are essential for handling this huge volume of data, but there are still many "missing links", that is, gaps in our understanding, that prevent us from being able to integrate this data into a cohesive whole.

Yoshida That's why using computers to draw inferences and discover hypotheses is important. Do you have any specific examples?

Inoue One example would be the mechanism for cancer resulting from ultraviolet light, for instance. When exposed to ultraviolet light, the skin's DNA gets damaged. If the number of abnormal skin cells increases, it may lead to cancer, so the body has an inhibitor gene, p53. p53 plays the essential role, but in actuality there are many reactions chained together, and if p53 binds with another protein and mutates, it may actually promote cancer formation. This structure is complex, and if the control network is completed we may be able to discover a new cancer inhibitor gene.

Yoshida Lastly, please tell us about your future goals.

Inoue The ultimate goal in this type of research is the discovery of useful hypotheses that have been heretofore unknown. In order for this to happen, the system must be used far and wide. Talking to biology researchers, I get the impression that they feel the need to process large volumes of data, but they are not sufficiently aware of systems biology. With systems biology, which handles large volumes of data, hypothesis finding systems can serve as effective tools for discovering hidden mechanisms. If there are more opportunities for the use of these systems in the future, I think they will ultimately produce results.

*1 Abduction: Used in logic to refer to inferring an explanatory hypothesis.
*2 Induction: Inference of general rules from cases. As with abduction, induction is another form of inference under incomplete knowledge.
*3 SOLAR: A system for efficiently deriving conclusions from a collection of given knowledge (logical formulas). These can be transformed to infer missing knowledge (hypotheses) from conclusions.

Noriyuki Yoshida
Staff writer, Science news department, Yomiuri Shimbun Tokyo Head Office

Comment from the Interviewer

I thought inference and hypotheses were the exclusive domain of humans. However, computers have increased both their speed and precision. Might there be no room for humans in the future? Professor Inoue clearly dispelled that worry. "There will always be a place for humans, looking at things from a bigger picture, and getting hints from different places." Those are words born from research into areas that are tied to the very essences of computation, intelligence and life. As we unravel ever more about the systems called "biological processes", I'm certain their profundity and splendor will become even more apparent.

"Meaning" Links Information, Changing the Web

Computers sort through information gathered from the Web, providing only useful information.
They behave as if they understand what people are thinking.
3 researchers working together on the same project discuss the allure and possibilities of the Semantic Web.

What is the Semantic Web, Which Reads People's Needs?

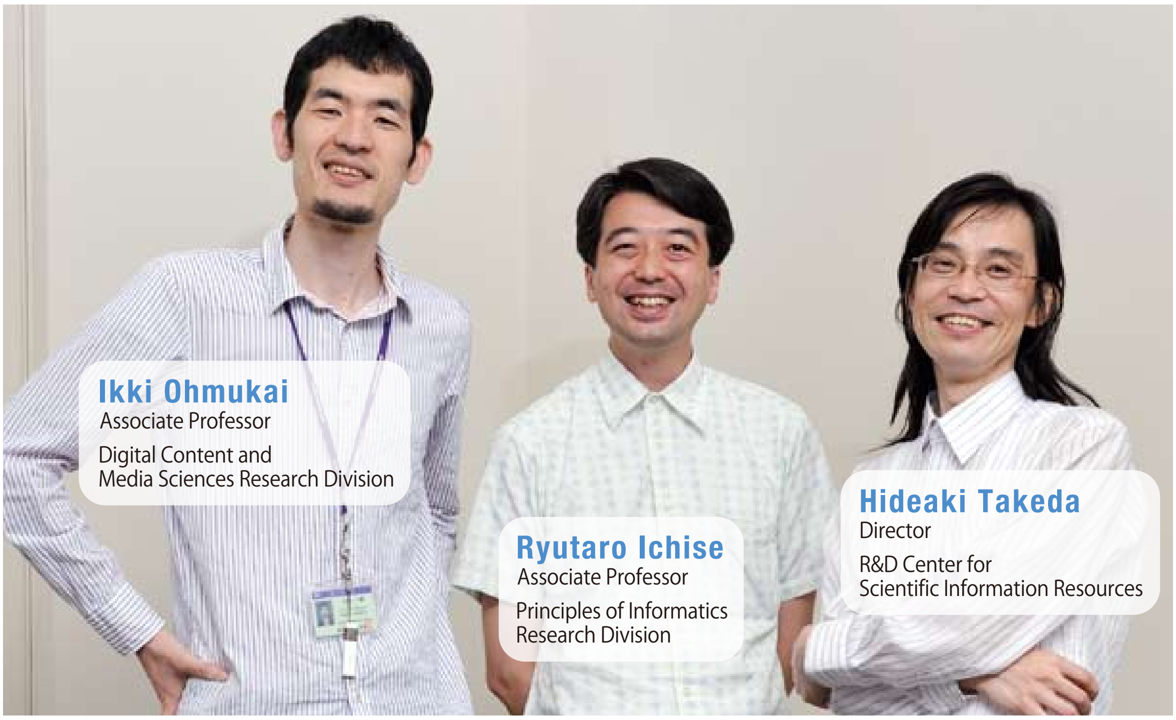
"Imagine that, during this discussion, I tossed out onto the Web the fact that I'm hungry. The computer then offered information about places to eat, taking into account even elements that I hadn't stated, such as when this meeting is scheduled to end, where I am, what I've eaten lately, and the like. This may sound like science fiction, but it may become reality in the not so distant future," explains NII Digital Content and Media Sciences Research Division Associate Professor Ikki Ohmukai, discussing the possibilities of the Semantic Web.

If computers can provide integrated, multifaceted information, it will decrease the time and effort expended by people in acquiring information from the Web, and making decisions based on it. The Semantic (*1) Web is an attempt to do just that. This concept was born of the fact that, despite collecting information from the Web becoming entirely commonplace, there is still no structure for selecting the information needed for making decisions from the massive volume of data on the Web. "It is said that 30% of white collar work consists of information retrieval. The current situation isn't all that much more efficient than before the Web existed, when we used to go to libraries to look up information."

With the advent of the Semantic Web, how will computers find the "information needed for decision-making" which proves so elusive for people?

Intelligent Computers Which Learn about the Relationships between Information

There are, broadly speaking, two approaches to creating a mature Semantic Web. The first is to use machine learning (*2) techniques to improve the information processing capabilities of computers. "Specifically, this involves considering the relationships between the meanings of words, and having computers memorize them," explains NII Principles of Informatics Research Division Associate Professor Ryutaro Ichise. "For example, the relationship between 'The Tale of Genji' and 'Murasaki Shikibu' is that between a work of literature and an author, but they are not equal. However, we collected information from the Web concerning these, and analysis of that information showed that on the web they are used with roughly the same meaning. It would help a user's decision-making process, then, if when they searched for 'The Tale of Genji' on the Web, they were also provided with search results for 'Murasaki Shikibu'. This processing can be made possible by performing computer analysis of information, and using those results to store 'The Tale of Genji = Murasaki Shikibu' in the computer." When computers perform this kind of information processing on the hundreds of millions of web pages on the Web, what is important is how well it can be done without human involvement, without mistakes or oversights, and how efficiently it can be performed. The key to these is held by the other approach.



Computers Reading and Linking Information

The second approach consists of distributing as much computer readable information over the Web as possible. Specifically, it consists of using XML (*3) or the like to append metadata (*3) which specifies data types. The most common description format of information on the Web, hyperlinks, are easy for humans to read, but not for computers. If the amount of information on the Web in easily machine-readable form were to increase, it would be possible for computers to automatically collect and analyze information, and store the relationships between data items.

At present, there is no requirement that data be uploaded to the Web with XML or similar metadata. One might think at first that in order to increase the amount of information in easily machine-readable form, some sort of new rules should be put in place. However, that may not always be the case.

"Are you familiar with the term 'social tagging'? The basic idea is to tag data uploaded to the Web with the assumption that it will be shared with other users. It has become a major movement, used by flickr, a photo sharing site, delicious, a social bookmarking site, and others. Their respective users are not necessarily concerned with the Semantic Web, but when clusters of tagged information are looked at from a higher vantage, one can see that they are systematized. These types of clusters of information collected by multiple users are referred to as the 'wisdom of the crowd'," NII R&D Center for Scientific Information Resources Director Hideaki Takeda explained. "Artificial intelligence research was initially targeted at individual intelligence. With the appearance of the Web, it has come to be directed at the intelligence of societies. The amount of data involved is immense, and constantly changing, making it extremely challenging for researchers. The wisdom of the crowd can be said to symbolize the intelligence of society."

"With the current rapid acceleration of social media, a range of styles of human relationships are being built on the Web with unprecedented speed. I look forward to research into the wisdom of crowds growing ever more interesting, as these relationships between people become intertwined with high volumes of information," said Associate Professor Ohmukai with a smile.

From personal to societal knowledge

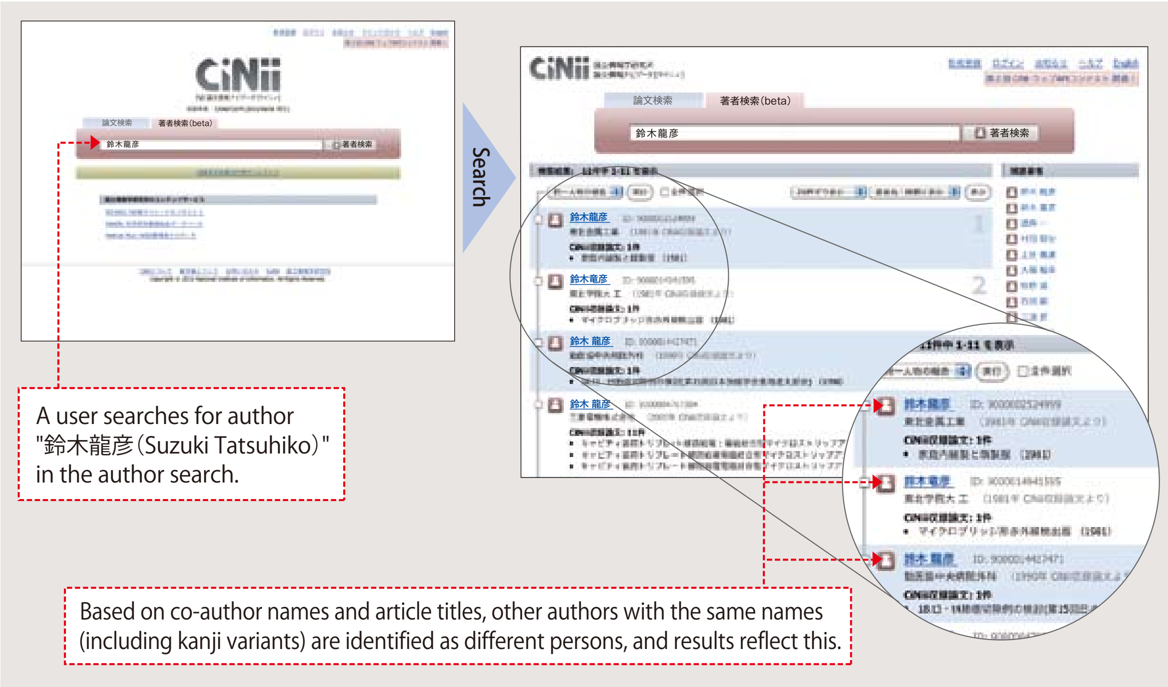


Figure Academic search service "CiNii"

NII's Semantic Web Approach

"I think it is effective as a means of discovering new, heretofore hidden knowledge, but I also think there's value in placing the information produced over a long period of time by a multitude of people in the context of the Semantic Web. The NII, by distributing its extensive archive of academic data in easily machine-readable form, hopes that the information will be used in entirely new ways," says Associate Professor Ohmukai, who is involved in Semantic Web support for NII's academic article search service, "CiNii" (see figure). The first step in this was the release, in April 2010, of search functionality that distinguished between people with the same first and last names. Normally, names are just strings of characters, and as such people who share names cannot be distinguished between, but Associate Professor Ohmukai explains that he developed the new search function by taking advantage of the traits particular to academic articles. "Academic articles are often written by multiple people, and their titles are strongly affected by the themes of the research they describe. This trait sets them apart from general books. By storing the relationships of 'author names' to 'co-author names' and 'article titles' in computers, people whose names are identical strings of characters are distinguished between."

This is an effective and cutting edge effort to use the two aforementioned approaches in order to make effective use of the massive volume of useful data already on the Web in the world of the Semantic Web.

The "LOD.AC (Linked Open Data for Academia)" project, in which Takeda, Ichise, and Ohmukai are involved, is one of NII's representative projects. This project distributes information in easily machine-readable form, and makes a wide range of data connections, in order to advance searching from merely "looking for a page" to "looking for things". For example, until now, if one wanted comprehensive information regarding the facilities where a certain artist's work was housed, one would have to manually follow each hyperlink. With "LOD.AC", one would be able to obtain information regarding those facilities in a single sweep. One of the defining traits of the Semantic Web is the presence of countless links between information items keyed to specific items such as this.

Transcending the Barriers of Language and Linking Information

When the words "Web" or "IT" are mentioned, there are surely many people for whom America leaps to mind. However, research into the Semantic Web is, surprisingly, more popular in Europe. "European livelihoods depend on multiple countries gathering together and acting as a single body, the EU. In order for this to happen, it is essential that information be shared across language barriers. The Semantic Web, which creates links between information using 'meaning' added via XML or the like to information, is an information sharing tool perfectly tailored to their needs," says Associate Professor Ichise. "In countries like Japan, in which communication is possible using a shared language, information can be shared without paying special attention to 'meaning', so many people may not have a clear understanding of what 'meaning' is. That is precisely why I think there needs to be an emphasis on Semantic Web research," says Professor Takeda.

Globalization will continue to speed up in a variety of fields, such as government and economics. Japan must be capable of gathering, and disseminating, the latest information, in order to avoid becoming isolated amidst this global tide. That must be considered as a given fact. Semantic Web research is strongly tied to Japan's international competitiveness.

The Expanding Possibilities of the Semantic Web

The Semantic Web is not limited to information acquisition; it also has potential in a variety of other fields. I discussed with each researcher the future they envision for the Semantic Web. Professor Takeda said he was interested in the relationship between the Semantic Web and the "life logs" of people's daily lives on the Web. "The devices which are used to record our daily lives are growing steadily more advanced, and it is possible to record everything we do. However, I think life logs will become useful information when it becomes possible to record not only our behavior, but why we did what we did – the "meaning" of our actions." If life logs could contain not only actions but their meanings, that would be revolutionary. Far-off friends could create a record in a diary-like form just by living their lives, and upload it to the Web to share, in real time, what they were thinking and doing. It would also show great promise as a way to perform regular checkups on the state of far away elderly persons, and in crime prevention measures.

Associate Professor Ichise believes that computers will automatically link information on the Web, and discover new knowledge. "In the world of the Semantic Web, data with different properties are becoming linkable. For example, there is experimental data regarding chemical substances from pharmaceutical companies, and data from academic bodies regarding those same chemicals. I think computers will become able to automatically determine that these data sets are about the same theme, analyze their contents, and provide methods for developing new drugs." It might also become possible to link the Semantic Web with camera robots, sharing video taken by the robot over the Web in a system which automatically searches for new biological organisms.

Associate Professor Ohmukai believes there will be demands placed on humans as the Semantic Web expands. "The spread of the Semantic Web means that computers will be automatically gathering information and providing only the parts which they have determined to be necessary. However, computers will not be able to determine if that information is in fact correct. Each of us will have to decide that on our own. Unless we ourselves change along with the changes in information acquisition systems, I don't think there is hope for the true progress promised by the Semantic Web."

The spread of the Semantic Web depends on both "computer capabilities" and "information description formats". In order for the Semantic Web which grows as a result of these to have a positive effect on society, we humans will also need to grow.

(Written by Satoko Isogai)

＊1 Semantic: "Of or related to meaning"
＊2 Machine learning: Technologies and methods for granting computers abilities equivalent the ability of humans to learn. These are founded on analysis of groups of data, and identification of patterns within them, making them deeply intertwined with statistics.
＊3 XML, metadata: Information regarding data, such as what it contains, when it was created, who created it, what format it is in, etc. is referred to as metadata. XML is a language used to describe metadata.

Seiji Yamada
Professor
National Institute of Informatics,
Digital Content and Media Science Research Division

Masayuki Okabe
Assistant Professor
Toyohashi University of Technology,
Information and Media Center

New Data Mining Made Possible by AI

Gathering information by searching for it on the Web has become commonplace, but it is difficult for the average user to find the information they need from the constantly growing volume of data. NII is currently at the center of research into interactive information acquisition and data mining using constrained clustering techniques. What is this revolutionary and innovative information acquisition technology based on fundamental AI technologies?

Applying Constraints to Clustering to Acquire Desired Information

NII Digital Content and Media Science Research Professor Seiji Yamada is currently at the center of research into a "framework for minimal user feedback (*) based interactive information acquisition and data mining". Professor Yamada has been involved in research into human interfaces and human agent interaction (see NII Today No.44), with the central theme of "machine-friendly people". In his research, he has constantly maintained that the conventional approach of improving individual system functions has its limits.

"I think that the first major breakthrough will come from providing easy, effective support to system users, so that the user and the system work in concert. In other words, by providing user feedback to the system, and performing re-clustering, better results can be provided to users. The main topic of this research is maximizing results while imposing a minimum amount of burden on the user.
Our goal is not just to create an instinctive, easy-to-use system, but a system backed by informatics theory," says Professor Yamada.

The key to this is "clustering". "Clusters" generally refers to collections of objects of the same type. "Clustering" is a fundamental technology used in data mining – the extraction of knowledge from large volumes of data. Clustering can be used, for example, to select photographs with common characteristics, such as mountain scenes or portrait shots, from a set of several thousand digital camera photos, and automatically divide them into several groups. Clustering can be applied not only to images, but any data which can be expressed with vectors, including text, audio, and video. For example, for text, one can examine how many of which words appear, expressing the results as a high-dimension vector.

"For this research, we used "constrained clustering", which applies constraints to clustering. By having people provide a direction to clustering, instead of performing clustering without user feedback, we can improve the accuracy of the clustering (see figure). For example, when photos of mountains and rivers are separated into different groups, but users want them to be in the form of a single "natural scenery" group, or when photos of cats and dogs are gathered into a single "animal" group, and users want to separate them into two separate groups, these grouping instructions are issued by the user. I have teamed up with Sector Leader Onoda, who is in charge of clustering, Associate Professor Takama, who handles the GUI, and Assistant Professor Okabe, who works on constrained clustering algorithms, to research how to increase efficiency in order to minimize the amount of human feedback required, and what kind of GUI (graphical user interface) can be used to make it easy for people to apply constraints," says Professor Yamada.

Constrained Clustering Accuracy Improvement Efforts

Clustering is a technology which sprang from the field of artificial intelligence machine learning. Machine learning includes both "supervised learning", where the computer is provided information regarding the information to be output, and "unsupervised learning", which can produce results without the need to specify the information to be output. Clustering, which automatically identifies and groups similar objects, is one of the fundamental unsupervised learning technologies. However, the results produced by unsupervised learning are not always those which people desire, so in constrained

clustering, people apply constraints in order to adjust the direction of results. This is called "semi-supervised learning".

So, how precisely can the accuracy of constrained clustering itself be improved? Sector Leader Takashi Onoda of the Central Research Institute of Electric Power Industry, in charge of clustering development for the research team, explains, "When computers automatically group information, the results are always identical. However, when people group information, the results vary from person to person. In order to make it easy for people to apply constraints, instead of just showing items which are similar, similar but different objects are compared. By doing so, feedback can be provided to the computer regarding how the user wishes to divide information, with a minimal burden on the user. As there is no clear answer, this is incredibly difficult. In the future, we hope to research clustering methods that make this more "human". That is, methods which produce results whose clusters are instinctually comprehensible to users."

The Toyohashi University of Technology's Assistant Professor Masayuki Okabe is responsible for developing algorithms for producing maximal effects from as few constraints as possible.

"My role is to create functions which produce maximum results from minimum constraints, such as algorithms for selecting constraints which show the greatest promise for improving clustering accuracy in order to reduce the number of user-placed constraints, or algorithms for pseudo-expansion of constraints. I'm also carrying quantitative evaluations by experimentally demonstrating how well these algorithms perform."

Developing a GUI that Makes it Easy for People to Compare and Evaluate

Another pillar of this research is research into an easy to use GUI. Tokyo Metropolitan University Associate Professor Yasufumi Takama, responsible for GUI research, explains, "When applying constraints to clustering results, users are shown two items on the computer screen, and asked to specify whether they wish those two to be considered as members of the same group, or of different groups. My task is to develop a GUI that makes users specify this kind of constraints with low burden. For example, when comparing documents, if the documents are long, they will take a long time to read, but if the documents are too short, the user may not have the information they need to assess the documents. In order to categorize documents correctly, users must evaluate several document pairs, but this results in a significant burden being placed on the user. That's why we are experimentally verifying how much information must be shown to users to produce optimal evaluation results. For example, we're looking at how much of a difference in time in takes to evaluate long texts, keywords, and snippets (summaries a few lines long). We're also using eye-tracking to verify where users look when they make their evaluations. As a result, we're producing interesting experimental results, such as finding that the snippets used by standard search engines are also effective in our own experiments, with users looking for longer periods of time at keywords shared by texts regarding the same topic, and spending more time looking at keywords that only appear in one of the texts for texts regarding different topics. We hope to further systematize the knowledge we gain about how to develop GUIs that better reflect the sensibilities of users."

Professor Yamada hopes in the future to research how to lower interaction costs and create a more general-purpose system. The creation of a system that can instantly select the information users need out of an ocean of information will only be possible by combining the knowledge and wisdom of informatics, such as fundamental AI technologies and applied mathematics.

(Written by Madoka Tainaka)

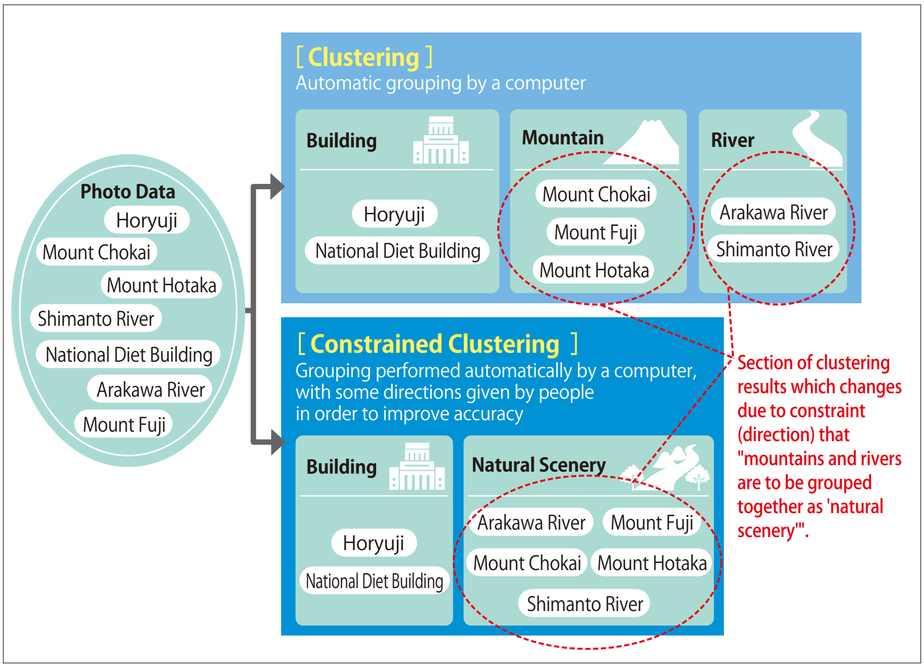


Figure　Difference Between Clustering and Constrained Clustering

*User feedback: Feedback provided to computers by people (users) working with information.

Artificial Intelligence
Reasons about the Courses of Trials

There is a need in the judicial world for the trial process to be accelerated. Research is underway to increase trial efficiency by applying logical processing to some civil litigation processes. Trial and legal documents contain expressions and words used by real people for communicating, and may seem quite removed from artificial intelligence and logic programming, but in fact they share much in common. We heard from Professor Ken Satoh, of the Informatics Principle Research Division, regarding the fruits of his research, and its future potential.

Avoiding the vagueness inherent in natural language through logic programming

Logic programming, explained simply, is the implementation of "mathematical logic" on computers. It is used as a description language for artificial intelligence. Professor Ken Satoh, of the NII Principles of Informatics Research Division, says, "The legal and judicial world has, so far, used natural language exclusively for all issues. I thought that adding a logical perspective, using symbolic notation and formulas, and applying logic programming, would result in the creation of new knowledge, and contribute to actual trials."

Ken Satoh
Professor
Principles of Informatics Research Division

There are three primary advantages of applying logical analysis to law. The first is that it enables one to discover implicit assumptions buried in legal documents. When using the logical approach, all assumptions must be explicit in order to produce results. However, because legislation is written in natural language, there are matters which are implied. These can be made explicit by logical expression. The second advantage is that if legal documents are logically analyzed, and it becomes possible to express them in strict logical forms, it will become possible to handle legal concepts themselves via computer. This will make it possible to resolve even more complex legal problems, and detect inconsistencies in the law. The third advantage is that by separating logical problems and legal problems, legal scholars will be able to concentrate exclusively on legal problems.

In other words, if logical approaches are applied to jurisprudence, things which were unclear will come to light, and previously complex work can be made more efficient.

The "presupposed ultimate fact theory" process which proves assertions

Professor Satoh is currently researching the deployment of logic programming in civil suits.

In civil suits, judges arrive at their conclusions by comparing the assertions of plaintiffs and defendants against substantive laws, such as civil and trade law. Articles of substantive law specify legal effects which pertain to specific facts, but do not go as far as specifying who, proving what facts, would be sufficient to make a determination of whether or not there were specific legal effects. For example, in a case wherein payment for purchased goods has not been made, the party wishing the payment to be made actively presents evidence that there was a purchasing contract, but the law does not clearly specify which claims, made by which parties, are sufficient to prove facts. The theory of presupposed ultimate fact is used by judges as a guideline when trying to determine, in a case such as the above, which party, the plaintiff or the defendant, must prove the case's facts, and what conclusion the judge should arrive at. Professor Satoh says, "this presupposed ultimate fact theory process is extremely similar to the non-monotonic reasoning(*) I have been researching, so I recognized the significance of developing corresponding logic programming."

Using AI to simulate uncertain aspects of judicial cases

The logical concept of "non-monotonicity" refers to the possibility that a previous conclusion may be retracted due to the addition of new information. Much of formal logic is monotonic — that is, even if new information is added, existing information is not eliminated. Non-monotonicity can be used to establish hypotheses given incomplete information. For example, assume that there is a hypothesis: "Cars drive on streets". If nothing is added to this statement, a car would be presumed to drive on the street. However, if one later found out that the car in question were a racecar, such as an F1 car, the conclusion that it drives on streets would be retracted. In cases such as this, where there is incomplete information, non-monotonic reasoning provides a logical description how to make deductions.

In that sense, trials are, as a rule, "non-monotonic". If they were monotonic, the results of a case once won would never be overturned. However, in reality, even if evidence presented in a first trial results in victory, new appealing evidence presented at an appeal trial might result in defeat, a non-monotonic change in the trial's verdict. Professor Satoh's belief was that the aforementioned presupposed ultimate fact theory is a formulation for rational deduction given incomplete information, and that as such non-monotonic reasoning could be applied. This led to his creation of ultimate fact theory logical programming like that shown in the diagram.

The diagram shows a simulation of a trial regarding the annulment of a contract due to unauthorized subleasing. A certain apartment owner claimed to the court that "a resident with a rental contract subleased the apartment to a third party without permission, so I would like the contract to be annulled." In this case, the owner must prove that he let the apartment to the resident in accordance with the proper procedures, and that the resident sublease the apartment to a third party. He does not, however, have to prove that he did not authorize the subleasing. Instead, the resident is responsible for proving that the owner gave his approval for the subleasing. Non-monotonic reasoning-based logical programming is used to have a computer carry out the same inference process as used in presupposed ultimate fact theory. Describing essential items, such as determined facts, precedents, and exceptions, results in a system which can be used to arrive at a conclusion — that is, whether the contract can or cannot be annulled.

Working towards the creation of the new academic field of "juris-informatics"

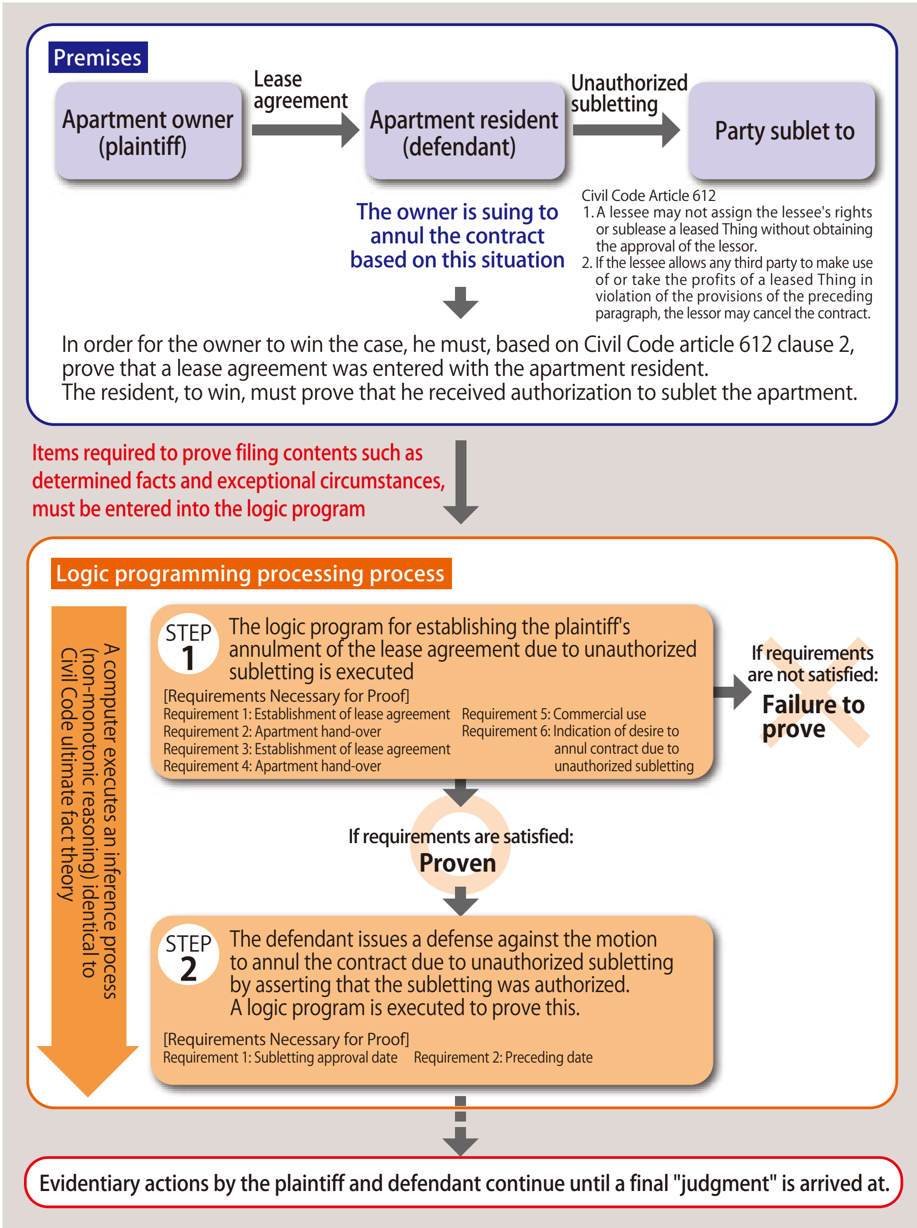
The goal of Professor Satoh's research is the creation of a new academic discipline, "juris-informatics", combining jurisprudence and informatics. However, he also points out that "logic programming cannot be applied to all aspects of laws and trials." For example, in a criminal case, when determining intent to kill, a conclusion must be reached taking into account an overall view, including how the incident occurred, how the murder weapon was used, how the defendant behaved after the incident, and the like. For this type of fact-finding, and interpretation based on common sense, logic programming cannot be used to reach optimal conclusions. Instead, human insight and judgment must be relied on.

In other words, there are elements of trials which require that conclusions be drawn only by people. As such, the goal of present research into juris-informatics is not "creation of artificial intelligence which will make all case judgments", but "application to logical work which is difficult using natural language". By making a clear delineation between human and computer roles, logic programming and artificial intelligence in the law field can achieve well-defined research goals.

(Written by Junichi Morimoto)

*Non-monotonic reasoning: Logic in which conclusions and results may change when additional inferences are added. For example, the verdicts may differ between first and second trials if additional evidence is discovered. In monotonic reasoning, conclusions and results do not change, even if new logical formulas are added.

Figure Logic Programming Application Example "Trial Regarding Contract Annulment Due to Unauthorized Subletting"



NII ESSAY

Sharing experiences—sharing privacy as well?

Isao Echizen
Associate Professor , Digital Content and Media Sciences Research Division, NII

For the first time in a while, I took vacation with my family. The hotel has an amazing ocean view, and I wanted to share what I'm feeling with others. Ok, I'll tweet about it. I'll give the name of the hotel, too.

Web travelogues are a sign that one is not at home

With the spread of cellular phones and networking, our society has become a convenient one in which we can share experiences like this in real-time. However, this is being accompanied by a growing societal problem where private information, such as where Twitter or SNS users are, and where they are travelling, is being discovered and maliciously used by third parties, to the detriment of those users. There was a case in America where someone tweeted that they had taken some time off and were travelling, and several thousand dollars were stolen from their home during their absence. There are also reports that position information can be analyzed to guess private information. A group of Rutgers University, led by Prof. Gruteser, observed 65 test subjects with car GPS systems for a week, and were able to identify where 85% of the test subjects lived. The low level of privacy awareness among Twitter and SNS users is exacerbating this problem. According to a USA TODAY article, most Americans who take long trips worry about their homes being burglarized in their absence, and intentionally turn off their answering machines and stop newspaper and mail delivery. However, experts point out, their vigilance does not extend to SNS use during their travels, and they freely disclose information about their location, uploading position information and photos in real-time from places hundreds of miles from home, or writing things like "I'll be staying here for a week".

What technologies make it possible to safely share ones moving experiences?

The problems described above have resulted in a greater interest in privacy protection technologies. In privacy protection research, it is important to ensure that third parties are not able to access excessive amounts of private information pertaining to those providing information. Spurred by society's needs, since the late 1980s researchers in a variety of fields, such as information security and data engineering, have taken multidisciplinary approaches to tackling this problem. A great deal of privacy protection technology research has been dedicated to obfuscating information contents, and adding false data, in order to make it difficult for third parties to acquire private data. However, little research has gone into how the quality of this obfuscated information degrades, and what affects it has on its recipients.

In today's age, where it is commonplace to communicate and share moving experiences and topics of conversation with the general public through Twitter and other social networks, focusing on privacy to the extent that one cannot share these experiences and conversations would be putting the cart before the horse. There is a need for a system that maintains the quality of information while offering a stable level of privacy protection.

This month's cover illustration: A small boy and an artificially intelligent robot play with an educational toy. The robot also learns and grows by interacting with the real world (society) around it. When will society be composed of humans and independent robots like this living together?

Weaving Information into Knowledge
NII

NII Today No.35 October 2010 (This English language edition of NII Today corresponds to No.49 of the Japanese edition)

Published by National Institute of Informatics, Research Organization of Information and Systems http://www.nii.ac.jp/
Address: National Center of Sciences 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430
Chief editor: Yoh'ichi Tohkura Cover illustration: Makoto Komori Photography: Hiroyuki Taniguchi Production: Commercial Design Center Inc.
Contact: Publicity and Dissemination Team, Planning and Promotion Strategy Department TEL:+81-3-4212-2131 FAX:+81-3-4212-2150 e-mail: kouhou@nii.ac.jp