

# DSIUシステム: Decision Support for Internet Users

— ネット情報を使ってホットなものをあなたに —

## DSIU Systems: Decision Support for Internet Users

— Providing Hot Choice by Using Information on the Internet —

藤本 和則  
Kazunori Fujimoto

NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories.  
fujimoto@cslab.kecl.ntt.co.jp, <http://www.kecl.ntt.co.jp/icl/about/fujimoto/index-j.html>

賀沢 秀人  
Hideto Kazawa

NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories.  
kazawa@cslab.kecl.ntt.co.jp, <http://www.kecl.ntt.co.jp/icl/about/kazawa/index-j.html>

佐藤 浩史  
Hiroshi Sato

NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories.  
hiroshi@cslab.kecl.ntt.co.jp, <http://www.kecl.ntt.co.jp/icl/about/hiroshi/index-j.html>

阿部 明典  
Akinori Abe

NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories.  
ave@cslab.kecl.ntt.co.jp, <http://www.kecl.ntt.co.jp/icl/about/ave/index-j.html>

松澤 和光  
Kazumitsu Matsuzawa

NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories.  
matuzawa@cslab.kecl.ntt.co.jp, <http://www.kecl.ntt.co.jp/icl/about/matuzawa/index-j.html>

**Keywords:** decision support systems, Internet users, knowledge acquisition, uncertain reasoning, information gathering.

### Summary

Decision Support for Internet Users, called DSIU, is an area of research for providing decision support for Internet users by using information on the Internet. DSIU aims to provide decision support with logical explanation taking account of user's preference. By using information extraction techniques, DSIU handles the names of various substances, e.g., electronic products, persons, and places, and so on, and constructs the explanations in terms of their properties. This paper describes the DSIU particularly from a viewpoint of realizing the DSIU and giving contributions to society in the near future. The information of DSIU is available at <http://www.kecl.ntt.co.jp/DSIU/>.

### 1. はじめに

近年、インターネットの普及により、世界各国に散在する最新の情報は瞬時に取得できるようになった。これにあわせて、ネットから有効な情報を集めて人間の活動に役立てる様々な研究が始められている [Knoblock 98b, Mahech 97]。こうした研究に多くの努力が払われているものの、ネット情報に基づく意思決定は依然難しいといえる。この原因としては、集められた情報が膨大であったり (ネット上の大量の情報から該当するものを集めるため)、あまり知られていなかったり (次々と新しい情報がネットに追加されるため) することがあげられる。こうしたネット情報の性質が、それらから適切な判断を下す

作業を困難にしている。こうした背景から、我々は、ネットユーザの意思決定を支援する人工知能システムの研究を進めている [藤本 98, Fujimoto 99a]。

我々の想定する意思決定支援の一例として、ユーザがある製品 (例えばデジタルカメラ) の購入を検討する場面を取り上げる (図 1)。ユーザは、自分の好み (例えば「かなり画質が良く、ある程度拡張性のある」など) を入力する。システムは、こうしたユーザに対して「どの機種がよりユーザの好みに合致しているか」を推論し、適切な機種名を提示する。また、提示した機種のおすすめ理由について論理的な説明を提供する (例えば「画質を優先するあなたには、CCD とレンズが高性能のものが良く…」など)。こうした説明をもとに、ユーザは自分のもつ知識に

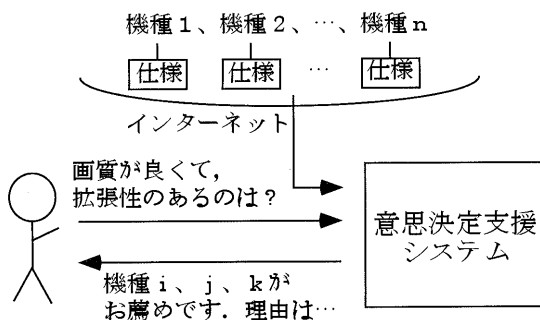


図 1 ネット情報に基づく意思決定の支援例

照らし合わせながら推薦内容を吟味し、結果として、納得のいく意思決定に到達する。

ユーザに適切なものを推薦する技術については、すでにいくつかの研究がある。情報フィルタリングの分野では、ユーザが過去に好んだものと類似したものを薦めたり、嗜好の似たグループの構成員が好むものを薦めたりする研究が行われている [Balabanović 97]。しかし、これらの方法では、ユーザから推薦の理由について尋ねられても、原理的に「似ている」以上の説明を提供できない。また、知的エージェントの分野では、ネット上からお薦めのものを探し出す研究が行われている [Doorenbos 97]。しかし、こうした方法の多くは、価格が安いなどあらかじめ用意された画一的な基準に基づくだけで、個々のユーザの嗜好を受け付けず、意思決定を支援するという立場からは、支援の内容を論理的に説明したり、ユーザの嗜好を考慮することは重要である。

我々の目指す人工知能システムは、ネット上に掲載されている“もの”をユーザの視点から評価し、その結果を論理的に説明するものである。こうした説明を前提とする評価は、人工知能での自動推論技術の得意とするところである。ネット情報を使った自動推論については様々な研究 [Craven 98, Fensel 98, Knoblock 98a] が行われているが、これらは、まだ、意思決定を支援するには至っていない。我々は、ネット情報を使った自動推論を“意思決定の支援”という視点から捉え、これに関する研究の分野を DSIU (*Decision Support for Internet Users*) と呼んでいる。本稿では、このような DSIU について、近い将来に実現して社会に貢献するという立場から述べる。まず、DSIU において重要となる研究課題について述べる。そして、DSIU への一つのアプローチとして、我々が進めている研究を紹介する。最後に、DSIU のもたらす社会貢献への期待を述べる。

## 2. DSIU という研究

ネット上に掲載されている“もの”をユーザの視点から評価し、その結果を論理的に説明するには、それを可能とする知識の獲得が必要となる。ネット上の多様で最

新な情報を扱う知識は、それ自体、多様で最新なものとして更新されなければならない。したがって、知識獲得の過程を人手に委ねるという考えは、コストの面から現実的とはいえない。そこで、DSIU では、意思決定支援に必要な以下の三つの知識をネットから自動獲得することを目指す。

**事実知識** 事実に基づく客観的な知識。例えば“…という機種がある”や“…という機種の CCD 画素数は 211 万画素である”など。Web-KB Project [Craven 98] で取り上げられている大学のコース名や職員の名前などの情報も、こうした事実知識に相当する。ネット情報から事実情報を抽出するにあたっては、情報抽出の技術 [Pazienza 97] が利用できる。また、wrapper (情報の抽出を行うツール) の共同開発も進められている [Sahuguet 99]。

**評価知識** 評価を与えた知識。例えば“…という機種は画質が良い”など。ユーザの視点に合った評価知識をネットから獲得することにより、これを用いて個々のユーザに適切な支援を提供することが可能となる。評価知識の獲得では、ネット上に存在する多様な尺度での評価を統一的な尺度での評価に統合する必要がある。例えば、評価の甘いサイトでの“かなり良い”という評価は、別のサイトでは単に“良い”と評価される場合がある。評価尺度の統一については [Grass 98] でもその重要性が指摘されている。また、評価知識の獲得では、ネット情報の信頼性の考慮が特に重要となる。

**説明知識** 事実と評価の知識を関連付けながら説明するための知識。例えば“ある機種が 211 万画素 CCD を採用する”ならば“その機種は画質が良い”など。こうした説明知識により、例えば、画質を優先するユーザに「この機種は、211 万画素 CCD なので画質が良いといえます」といった説明を提供することが可能となる。こうした説明知識をネット情報から獲得するにあたっては、テキストマイニングの技術 [Hahn 97, Hearst 99] も利用できる。

DSIU では、これら三つの知識を使って支援の内容を構成する。具体的には、ユーザの視点に合った評価知識を選び出し、これを説明知識を使って事実知識に結び付けながら説明を構成する。

以上述べたように、DSIU では、事実、評価、説明の三つの知識を自動獲得する技術、および、獲得された知識を組み合わせて支援内容を導く技術、の二つの技術の確立が重要な研究課題となる。

## 3. 我々のアプローチ

本章では、2 章で述べた DSIU への一つのアプローチとして、我々が進めている研究を紹介する。

### 3.1 テキストからの知識獲得技術

我々は、テキスト構造のもつ意味から知識の獲得を行う知識構造変換法 [Kazawa 99a, 賀沢 99b] の研究を進めている。自然言語の意味を理解する方法を利用した場合、表現の多様さの扱いが困難となる。そこで、文の列挙構造や単語の記述位置などの比較的扱いやすい情報に着目した。知識構造変換法は、テキスト構造に意味を与えるにあたって、人間の記述慣習(テキストを記述するときの慣習)を利用する。例えば、人間が宣伝文を記述するとき、より人目を引くようにするため、優れた仕様をテキストの初めの方に書くことが多い。こうした記述慣習を利用すると、テキスト中での単語の記述位置から、その単語の示す仕様の“良さ”を知ることが可能となる。我々は、これまでに、記述慣習の考えに基づいて、次の二つの種類の知識を獲得する方法を示した。

**依存関係** 「画質は CCD には依存するがバッテリーには依存しない」などの依存の関係についての知識。こうした知識は、説明知識の一つとして利用できる。

**影響順序** 「補色より原色フィルタの方が良い」などの順序関係についての知識。こうした知識は、評価知識の一つとして利用できる。

そして、ネット上の製品記事(デジタルカメラに関する 62 記事)について実験を行い、その有効性を確認した。今後、製品記事に限らず、様々な種類のテキストを対象に記述慣習に基づく知識獲得の体系化を進める。

### 3.2 不確実で不完全な知識のための推論技術

我々は、不確実で不完全な知識から得た結論について、その妥当性を数量的に評価する確率超空間推論法 [藤本 98, Fujimoto 99b] の研究を進めている。テキストから自動獲得された知識は、確実なものとはいえ、また、不足や矛盾がある場合が多い。したがって、不確実で不完全な知識を扱う推論法が必要となる。確率超空間推論法は「問題領域の確率としてあり得る全体」という分布超空間を構成し、知識を分布超空間についての制約式として統一的に扱う。これによりテキストから得られた様々な形式の知識を統一的に扱うことが可能となる。さらに、この推論法では、集められた知識の分布超空間上での特徴(例えば、充足領域の広さ)に基づいて、得られた結論の妥当性を数量化する。これにより、最も妥当な結論を選び出すことが可能となる。しかし、確率超空間推論法は分布超空間という大きい空間を扱うため、計算時間が大きいという問題点がある。したがって、今後は、計算の高速化が必須となる。

### 3.3 日常単語の扱いのための意味照合技術

ネット上のテキストでは表現の仕方が統一されておらず、同じような意味に多様な表現が用いられる。例えば“奇麗”と“美しい”は異なる表記であるが同じような意味で使われる。こうしたテキストからの知識獲得、及び、

獲得された知識のもとでの推論では、多様な表現を意味に基づいて照合する技術が必要がある。我々の研究グループでは、これまでに、単語の意味の近さを測る概念ベース [笠原 97] を構築した。現在、こうした概念ベースをはじめ、その他の類似性判別手法(シソーラスベース、コーパスベースのもの)を組み合わせて、高精度の意味照合を実現する研究を進めている [Fujimoto 99c]。さらに、より柔軟な照合を可能とするため、因果についての知識 [佐藤 99] を用いた照合の検討や、照合判定を行う箇所を導出する補間仮説推論 [阿部 98] の適用も進めている。

## 4. DSIU の社会貢献

現代の社会は、高度情報流通社会という言葉に代表されるように“情報”や“もの”のあふれた社会であるといえる。情報の入手はインターネットの普及により格段に容易になった。商品は個人の嗜好の重視のもとに多種多様化が進められている。こうした社会において、人間がより豊かな生活をおくるには、流通する情報やものをうまく活用する意思決定の過程が重要・不可欠であろう。DSIU は、こうした高度情報流通社会での人間の意思決定を支援するための研究である。

従来の研究は、支援の内容を論理的に説明したり個々のユーザの嗜好を考慮するという点で不十分であった。DSIU では、意思決定の支援にあたって、ユーザの嗜好を考慮した論理的な説明を提供することを目標とする。こうした説明を提供することにより、はじめて、ユーザは決定の良し悪しを自分自身で検討できるようになり、結果として、納得のいく意思決定に至ることができる。こうした DSIU は、単に、自分に合ったものを手軽に取得できるようにしたり、最新のものの中から選択できるようにするだけではなく、高度情報流通社会に納得のいく意思決定を提供する研究であると我々は考える。

一方、XML などの言語により、あらかじめネット情報を構造化して提供しようとする動きがある [Bray 98]。こうした研究は、計算機によるネット情報の扱いを容易にするものとして期待できる。しかし「仕様の画一化が難しい領域(例えば、演劇、アートの展覧会など独自性の強いもの)」や「短期間のみ提供の価値があるような領域(例えば一時的なキャンペーン商品など)」では、常に、記述に手間のかかる構造化言語が使われるとは限らない。また、ネット上の多くの情報を総合して知識を得るという過程は、たとえ情報が構造化されて提供されたとしても、情報の信頼性を考慮するという観点から依然不可欠といえる。したがって、構造化言語が普及したとしても DSIU が重要であることに変わりはない。なお、DSIU の情報は <http://www.kecl.ntt.co.jp/DSIU/> にて提供されている。

## 謝 辞

本研究に関して有益なコメントを頂いた査読者に感謝の意を表します。

## ◇ 参 考 文 献 ◇

- [阿部 98] 阿部明典: 欠如節を生成する推論法, 信学会論文誌, Vol. J81-D-II, No. 6, pp. 1285-1292 (1998).
- [Balabanović 97] Balabanović, M. and Shoham, Y.: Content-Based, Collaborative Recommendation, *Communication of the ACM*, Vol. 40, No. 3, pp. 66-72 (1997).
- [Bray 98] Bray, T., Paoli, J., and Sperberg-McQueen, C. M.: Extensible Markup Language (XML) 1.0, *W3C Recommendation*, Available from <http://www.w3.org/TR/REC-xml/> (1998).
- [Craven 98] Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K., and Slattery, S.: Learning to Extract Symbolic Knowledge from World Wide Web, in *AAAI-98*, pp. 509-516 (1998).
- [Doorenbos 97] Doorenbos, R. B., Etzioni, O., and Weld, D. S.: A Scalable Comparison-Shopping Agent for the World Wide Web, in *Proceedings of the First International Conference on Autonomous Agents*, pp. 39-48 (1997).
- [Fensel 98] Fensel, D., Decker, S., Erdmann, M., and Studer, R.: Ontobroker: The Very High Idea, in *Proceedings of the Eleventh International Flairs Conference (FLAIRS-98)*, pp. 131-135 (1998).
- [藤本 98] 藤本, 松澤: インターネット上の記述文から確率知識を構成する一手法: 構成の基本原則を中心に, 情報学シンポジウム, pp. 129-136 (1998).
- [Fujimoto 99a] Fujimoto, K. and Matsuzawa, K.: Intelligent Systems Using Web-pages As Knowledge Base for Statistical Decision Making, *New Generation Computing*, Vol. 17, No. 4, pp. 349-358 (1999).
- [Fujimoto 99b] Fujimoto, K., Matsuzawa, K., and Kazawa, H.: An Elicitation Principle of Subjective Probabilities from Statements on the Internet, in *Proceedings of the Third International Conference on Knowledge-Based Intelligent Information Engineering Systems (KES-99)*, pp. 459-463 (1999).
- [Fujimoto 99c] Fujimoto, K. and Sato, H.: Semantic Word-matching for Knowledge Acquisition from Text Containing Daily-used Words: A Multiagent-based Approach, in *The Eighth Workshop on Multiagents and Cooperative Computing (MACC-99)* (1999).
- [Grass 98] Grass, J. and Zilberstein, S.: From HTML to usable data: Problems in meaning and credibility in the WWW, *AAAI Workshop on AI and Information Integration* (1998).
- [Hahn 97] Hahn, U. and Schnattinger, K.: Knowledge Mining from Textual Sources, in *Proceedings of the Sixth International Conference on Information and Knowledge Management (CIKM-97)*, pp. 83-90 (1997).
- [Hearst 99] Hearst, M. A.: Untangling Text Data Mining, in *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL-99)* (1999).
- [笠原 97] 笠原, 松澤, 石川: 国語辞書を利用した日常語の類似性判別, 情報処理学会論文誌, Vol. 38, No. 7, pp. 1272-1283 (1997).
- [Kazawa 99a] Kazawa, H., Fujimoto, K., and Matsuzawa, K.: Attribute Dependency Acquisition from Formatted Text, in *Proceedings of the Third International Conference on Knowledge-Based Intelligent Information Engineering Systems (KES-99)*, pp. 464-468 (1999).
- [賀沢 99b] 賀沢, 藤本, 松澤: Web テキストを知識ベースとして用いる推論システムの提案: テキストからの知識獲得方式を中心に, 人工知能学会研究会資料 SIG-KBS-9803-9, pp. 49-54 (1999).
- [Knoblock 98a] Knoblock, C. A. and Minton, S.: The Ari-

adne approach to Web-based information integration, *IEEE Intelligent Systems*, Vol. 13, No. 5, pp. 17-20 (1998).

- [Knoblock 98b] Knoblock, C. and Levy, A. eds.: *AAAI-98 Workshop on AI and Information Integration*, The AAAI Press (1998).
- [Mahech 97] Mahech, K. ed.: *Natural Language Processing for the World Wide Web, AAAI Spring Symposium Series Technical Reports*, The AAAI Press (1997).
- [Pazienza 97] Pazienza, M. T. ed.: *Information Extraction*, Springer-Verlag (1997).
- [Sahuguet 99] Sahuguet, A. and Azavant, F.: Wysiwyg Web Wrapper Factory (W4F), Available from <http://db.cis.upenn.edu/W4F/> (1999).
- [佐藤 99] 佐藤, 笠原, 松澤: テキスト上の表層的因果知識の獲得とその応用, 電子情報通信学会技報 TL98-23, pp. 27-34 (1999).

[担当委員: 大澤幸生]

1999 年 6 月 18 日 受理

## — 著 者 紹 介 —



藤本 和則(正会員)

1989 年同志社大学工学部電気工学科卒業。1992 年京都大学大学院工学研究科修士課程修了。同年、日本電信電話(株)入社。以来、人工知能、特に不確実性を扱う推論技術の研究に従事。現在、NTT コミュニケーション科学基礎研究所研究主任、情報処理学会、日本ファジィ学会、行動計量学会各会員。



賀沢 秀人

1995 年東京大学理学部物理学科卒業。1997 年同大学院理学系研究科修士課程修了。同年、日本電信電話(株)入社。現在、NTT コミュニケーション科学基礎研究所勤務。人工知能、特にテキスト情報からの知識獲得の研究に従事。情報処理学会会員。



佐藤 浩史

1994 年東北大学理学部数学科卒業。1996 年同大学院理学研究科数学専攻修士課程修了。同年、日本電信電話(株)入社。現在、NTT コミュニケーション科学基礎研究所勤務。知識処理技術、主に大規模知識ベース研究に従事。情報処理学会会員。



阿部 明典(正会員)

1986 年東京大学工学部電子工学科卒業。1991 年同大学院工学系研究科電子工学専攻修士課程修了、工学博士。同年、NTT 入社、現在 NTT コミュニケーション科学基礎研究所勤務。Abduction などの人工知能の研究を行っており、最近では、ことば工学という新しい分野の研究に興味を持っている。電子情報通信学会、社会言語科学会各会員。



松澤 和光(正会員)

1975 年東京工業大学工学部電子工学科卒業。1977 年同大学院修士課程修了。同年、電電公社電気通信研究所入所。現在、NTT サービスインテグレーション基盤研究所主幹研究員、コミュニケーション科学の研究に従事。情報処理学会等、各会員。