DSIU: ネットユーザのための意思決定支援

個人ページの利用価値とプロトタイプの構築を中心に

Decision Support for Internet Users: On Availability of Personal WebPages and A Prototype Development.

藤本 和則
Kazunori Fujimoto

有限会社フジモト・リサーチパーク 人工知能研究所
Artificial Intelligence Laboratory, Fujimoto Research Park Co., LTD.
frp@palette.plala.or.jp, <http://www.frp-com.jp/ai/fujimoto.html>

島津 光伸
Mitsunobu Shimazu

NTT コムウェア (株) 関西支社
Kansai Regional Division, NTT Comware Corporation.
shimazu.mitsunobu@nttcom.co.jp

keywords: decision support systems, internet users, knowledge acquisition, personal web-pages, user interface.

Summary

This paper describes availability of personal Web-pages and a prototype development for Decision Support for Internet Users, called DSIU, which is an area of research for decision support by using information on the Internet. The availability of Web-pages concerns usage of formal pages, which are provided by companies and so on, and personal pages, which are provided by private persons. Web-pages are gathered by using an Internet search engine to determine destinations for travel and personal pages are confirmed to provide much subjective information than formal pages. The prototype development concerns a travel recommendation system, which is a kind of decision support systems. The prototype uses subjective and objective information on the Internet to select several destinations for users and to provide explanations the reason why the destinations are recommended. This paper also describes our perspective of DSIU researches.

1. は じ め に

近年、インターネットの普及により、人間は世界各国に散在する最新の情報を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。我々は、ネットユーザのための意思決定支援を DSIU と名付け、その研究を進めている [藤本 01]。ここに、DSIU は“デシュウ”と読み、*Decision Support for Internet Users* の略である。ネット上の情報を使って意思決定するには、情報を活用するための「知識」が必要となる。我々は、こうした「知識」は、それ自体、ネット上に存在するのではないかという着想から「ネット上の記述文から情報活用のための知識を自動獲得する」というアイデアのもとに研究を進めてきた。以下では、こうした研究の経過について簡単にまとめる。

- (1) ネットユーザのための意思決定支援として、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枠組を提案した [藤本 01]。この枠組では、意思決定支援に利用する事実、評価、説明の 3 種の知識をネット情報から自動獲得する。
- (2) 自動獲得の核技術として、テキストの構造のもつ意

味から知識を獲得する方式 [Kazawa 99] と単語の意味の同一性を判定する方式 (例えば“奇麗”と“美しい”は同一の意味と判定される) [Sato 00] を開発した。これらの技術により、電化製品の機種を選ぶという問題において、評価と説明の知識がある程度の精度で自動獲得できる見通しを得た。

(3) 事実、評価、説明の知識はネット上に獲得可能な形で提供されているか、について調査を行った [藤本 01]。具体的には、Web 上に提供される情報を対象に、5 つの分野 (音楽 CD、大学、薬、政治家、旅行先) について、知識がネット上に存在するかと自動獲得可能かの調査を行った。この結果、現在の Web 上には、DSIU の研究を実践するのに十分な知識が存在することを確認した。

提案の技術に基づく意思決定支援システムを近未来に実現するには、ネット上にどのような知識がどのような形で提供されているかを注意深く調べる必要がある。我々は、これまで、取得の容易さから公式ページ (企業や自治体などの組織が提供するページ) を対象とし、知識の自動獲得の可能性について検討を進めてきた。一方、個人ページには、意見や主観に基づくいわば「柔らかい情報」が多く提供されていることが予想され、公式ページにはない利用価値を期待できる。そこで、今回、客観か

ら主観まで幅広い知識が必要となる「旅行先の選定問題」を取り上げ、個人ページの利用価値について検討を行った。2章では、こうした個人ページの利用価値について述べる。また、今回、旅行先の選定を支援する旅行先推薦システムのプロトタイプを構築した。このプロトタイプは、ユーザの選択した複数のキーフレーズから適切な旅行先を選定し、その推薦理由を説明するシステムである。3章では、このプロトタイプの概要と技術的特徴について述べる。

2. 公式と個人のページの利用

本章では、ネット情報を使った意思決定支援での個人ページの利用価値について述べる。2.1節では、Webページをその提供元の種類から公式、準公式、準個人、個人の4つに分類し、それぞれのページがWeb上にどれくらいの割合で存在するかについて述べる。2.2節では、個人、公式ページ中の記述文について、事実と評価の知識がそれぞれどれくらい存在するかについて述べる。2.3節では、個人ページに存在する評価知識について詳しく調べ、その意思決定支援への利用について考察する。なお、本章の調査は、一名の人員で、2000年12月から1ヶ月程度で行った。

2.1 公式ページと個人ページ

調査にあたっては、まず、提供元の種類からWebページを次の4つに分類した。

公式ページ 自治体など政府機関の提供するページ

準公式ページ 民間企業の提供するページ

準個人ページ 個人の提供するページで、公共の情報を提供するもの(観光地に関する総合情報など)

個人ページ 個人の提供するページで、個人的な情報を提供するもの(日記、旅行記など)

そして、Webから観光スポットに関するページ群を収集したとき、これら4つのページ種別がどのような割合で現われるかを調べた。観光スポットについては、以下のように京都と奈良からそれぞれ5スポットずつランダムに選んで調査対象とした。

京都府 京都ジョイボリス(京都市)、青谷梅林(城陽市)、湯の花温泉(亀岡市)、京都フラワーセンター(精華町)、琴引浜(網野町)

奈良県 谷瀬の吊り橋(十津川村)、奈良ドリームランド(奈良市)、檀原神宮(檀原市)、竹取公園(広陵町)、室生寺(室生村)

ページの収集は検索エンジン goo[go] を用いて行った。まず、各スポット名をキーワードとしてページを検索し、そして、最新のページからページ数30を超える日付を確定し、その日以前のページを全て調査対象とした。この結果を表1に示す(ミラーやURL先のデータがないページは“その他”とし表からは除いた)。表1から、個

表1 個人ページの割合(単位: %)

観光スポット	公式	準公式	準個人	個人
京都ジョイボリス	0.0	48.6	5.7	45.7
青谷梅林	33.3	36.7	20.0	10.0
湯の花温泉 亀岡	6.3	37.5	18.8	34.4
京都フラワセンタ	33.3	50.0	6.7	6.7
琴引浜 網野	2.9	41.2	2.9	52.9
谷瀬の吊り橋	0.0	25.0	6.3	68.8
奈良ドリームランド	0.0	56.3	3.1	40.6
檀原神宮	11.8	58.8	8.8	20.6
竹取公園 広陵	27.3	21.2	21.2	27.3
室生寺	0.0	37.8	13.5	45.9
平均	11.4	41.3	10.7	35.3

表2 抽出された知識数(単位: フレーズ/ページ)

スポット名	個人ページ		公式ページ	
	事実	評価	事実	評価
京都ジョイボリス	0.1	0.0	7.0	0.0
奈良ドリームランド	0.0	1.0	2.0	2.0
琴引浜	0.1	1.0	10.0	2.0
谷瀬の吊り橋	1.8	1.7	6.0	2.0
室生寺	1.8	1.0	6.0	2.0
平均	0.8	0.9	6.2	1.6

人ページは、全体の35%程度を占めることがわかる。このように3割を超えるページを占める個人ページを利用するという考えは、Web上の情報を有効活用するという視点からは重要であるといえる。

2.2 個人ページ中の評価知識

本節では、個人ページ中に意見や主観に基づく評価知識がどの程度存在するかについて述べる。調査にあたっては、個人ページの多い5つのスポット(京都ジョイボリス、奈良ドリームランド、琴引浜、谷瀬の吊り橋、室生寺)について、ページ中の記述文から観光スポットを説明するフレーズを取り出し、それが事実知識か評価知識かの判定を人手で行った。例えば「高さは54メートルです」というフレーズは事実知識と判定され、「めっちゃくちゃ熱かったです」というフレーズは評価知識と判定される。そして、5スポットのそれぞれについて、個人、公式の1ページあたりの「知識を含むフレーズの数」を事実、評価のそれぞれについて調べた。この結果を表2に示す。なお、個人ページはgooで検索された最新の10ページ(ミラーサイトは除いた)、公式ページは市区町村の提供する1ページを対象とした。記述文からのフレーズの抽出規則については、[島津 01] に詳しいので参照されたい。

表2から、個人ページでは、事実と評価の知識がほぼ同程度提供されているのに対し、公式ページでは評価より事実知識の方が4倍程度多く提供されていることがわかる。また、公式ページ中の知識と個人ページ中の知識の重複については、谷瀬の吊り橋の形状(長さ297m、高

さ 54m) に関する記述で同一、または、表記は異なるが同じ意味のものが多数あったものの、それ以外の記述については重複があまりみられなかった。したがって、公式ページに加え、個人ページを利用することは、多くの知識を獲得するという点から重要であるといえる。

2.3 個人ページの利用価値

個人ページについては、公式ページにない知識を得るという目的以外の利用も考えられる。例えば、公式ページにすでに記述があっても、それと同一の記述を個人ページに見つけることにより、その記述の信頼性が増したりすることがある。こうした場合、必ずしも公式ページ中の知識と異なるものが得られたとは言えないが、信頼性の高い知識を得ることができたといえる。また、同じ対象について記述されている場合でも、公式ページより詳しい内容が提供されていることもある。以上の考えから、本章では、今回調査したページの内容のうち、複数の独立なページに同一の対象について記載がある谷瀬の吊り橋と室生寺を取り上げ、個人ページの利用価値を論じる。

谷瀬の吊り橋 公式ページでは「長さ 297 m」との記載があるのに対し、個人ページには「長さ 297.7 m」、「長さ 297 m」、「全長 2997 m」という記述がみられた。「長さ 297.7 m」や「長さ 297 m」という記述に関しては、それぞれ公式ページのものより詳細であったり、公式ページのものと同じのものである。したがって、それぞれ、公式ページより詳しい知識として利用したり、公式ページの知識を支持するものとして利用することができる。これに対して、「全長 2997 m」という記述に関しては、小数点が落ちた「誤った情報」なので、これを利用することはできない。

室生寺 公式ページでは五重塔の被害に関する記述はなかったのに対し、個人ページには「甚大な被害をうけた」、「倒れてしまった」、「台風で壊れた」、「先年の台風にて倒れた」、「破損した」という記述がみられた。「被害をうけた」や「破損した」という記述に関しては、似たような内容を提供するものであり、お互いの信頼性を向上させるのに利用できる。これに対して、「倒れてしまった」や「先年の台風にて倒れた」は、内容を誇張した「誤った情報」であり、適切に排除されなければならない。

以上から、個人ページは、公式ページにはない知識を補ったり、知識の信頼性を向上させたりするという2つの利用価値があるといえる。また、個人ページでは、同じような内容に多様な言い回しが使われていたり、誤った内容が提供されていたりする。したがって、個人ページを利用する枠組みには、(1) 表記ではなく意味に基づいて文のマッチングを行う技術、(2) 誤った知識を他の記述と照らし合わせながら適切に排除する技術、の2つの技術の確立が必要となる。

3. 旅行先推薦システムの構築

本章では、旅行先の選定を支援する旅行先推薦システムのプロトタイプ「わがまま KIDS」について、その動作概要と技術的特徴を説明する。わがまま KIDS では、具体的な願望を持たないユーザに適切な旅行先を薦めるため、ユーザの願望となりえるフレーズをシステム側から提示し、それを選択させる方法を採用した。これにより、ユーザの潜在的な願望を引き出すことができ、選ばれたフレーズに基づいて検索/推論が可能となる。なお、潜在的な願望を引き出す効果については、[三井 01] にて詳しい考察がされているので参照されたい。

わがまま KIDS のフレーズ選択画面を図 1 に示す。図 1 では、システム側から「黒部区の氏神・深田部神社の秋の祭礼に行われる」など多くのフレーズが提示されている。ユーザは、こうして提示されるフレーズをクリックし、気に入ったものを5つまで選択できる。選択されたフレーズは画面下部の車両に記載される。フレーズを選択が終わるとユーザは検索ボタンを押す。検索ボタンが押されると、システムはユーザの選択したフレーズと意味的に関連の強い観光スポットを検索する。そして、検索されたスポットを多く有する市区町村を推薦し、そのお薦めの理由を解説する。この解説にあたっては、観光案内ページ中の内容のうち、ユーザが選択したフレーズと意味的に関連の深いものだけが取り出される。このように、解説文を短く端的に提示することにより、ユーザは短い時間で容易に内容を確認できるようになる。わがまま KIDS はプログラミング言語 perl により実装され、京都府、奈良県の全市区町村(各 44, 47)を対象に動作中である(各市区町村のスポット数は 30 を上限とした)。

我々は、わがまま KIDS の構築にあたって新しく2つの技術を開発した。以下では、それぞれについて説明する。

(1) 意味解析に基づく知識獲得技術

ユーザの願望となりえるフレーズをシステム側から提示するには、そうしたフレーズを多数用意する必要がある。そこで、観光案内ページ中の記述文の係り受け関係を解析してフレーズを自動獲得する方法を開発した。これは日英機械翻訳システム ALT-J/E[Ikehara 91] の係り受け解析機能を用い、その係り受け種別に基づいてスポットを説明する部分を抽出する方法である[島津 01]。

(2) フレーズの意味的な照合技術

ユーザの選択したフレーズを使って検索/推論を行うためには、フレーズ間の意味的な関連性を判定する必要がある(例えば「荘厳な美しさの」と「金色に輝く」とは関連があるとする)。そこで、フレーズ間の意味的関連性を自動判定するためのデータベース構築法を開発した。これは、観光案内ページをコーパスとして、スポットを修飾するフレーズを集め、そのスポットの属するカテゴリ上に定義された類似度からフレーズ間の関連の度合を数量化するものである[前川 01]。

図 1 わがまま KIDS フレーズ選択画面

4. 今後の展望

今後、提案の意思決定支援の枠組を広く社会に普及させるには、次の2つの研究が重要となると考えている。一つは、ネット情報を使った意思決定支援について、その支援性能をあらかじめ知ることができる数理モデルの研究である。こうした数理モデルにより、システム設計者は、システムの構築前に、要求性能がどの程度満たされるかを知ることが可能となる。もう一つは、システムとユーザが図表や音声によって情報をやりとりするインタラクションの研究である。こうしたインタラクションにより、システムは、ユーザの嗜好や意図を適切に取り出し、支援の内容を効率良く伝えることが可能となる。我々は、これらの研究を進めるとともに、ビジネスの枠組で AI システムを普及されるための方法論についても検討を進める。

謝 辞

本研究は、NTT コミュニケーション科学基礎研究所で行われた。2章での調査は、岩本 一弘氏 (NTT コムウェア (株)) に協力頂いた。3章でのシステム実装は、前川 宗保氏、藤野 耕献氏 (NTT コムウェア (株)) に協力頂いた。また、松下 光範氏 (NTT) からは、本稿について貴重なコメントを頂いた。ここに記して感謝する。

◇ 参 考 文 献 ◇

- [goo] goo, : <http://www.goo.ne.jp/>.
 [Ikehara 91] Ikehara, S., Shirai, S., Yokoo, A., and Nakaiwa, H.: Toward an MT System without Pre-Editing Effects of New Methods in ALT-J/E, in *MT Summit 91*, pp. 101–106 (1991).
 [Kazawa 99] 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).

- [Sato 00] Sato, H. and Fujimoto, K.: A New Approach to Semantic Word-matching for Knowledge Acquisition from Text Containing Daily-used Words, *Advances in Intelligent Systems: Theory and Applications*, pp. 135–140, IOS PRESS (2000).
 [三井 01] 三井, 坂田, 丸山, 国領, 前川, 藤野, 岩本, 小林: DSIU 技術を用いた新しい旅行先検索サイトの考察, 人工知能学会第 15 回全国大会, 1C1-05 (2001).
 [前川 01] 前川, 藤本, 佐藤: DSIU における説明文の意味照合法, 人工知能学会第 15 回全国大会, 1C1-03 (2001).
 [島津 01] 島津, 藤本, 畑山, 山田: 観光案内文書からの観光スポット説明文の自動獲得, 人工知能学会第 15 回全国大会, 1C1-02 (2001).
 [藤本 01] 藤本, 賀沢, 佐藤, 島津, 北: ネット情報を使った意思決定支援 DSIU における知識獲得技術, 人工知能学会論文誌, Vol. 16, No. 1, pp. 120–129 (2001).

〔担当委員: 大澤幸生〕

2001 年 10 月 23 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) に入社し、人工知能の研究に従事。2001 年同社を退社し、有限会社フジモト・リサーチパークを設立。現在、同社代表取締役として人工知能の研究を推進。情報処理学会、日本ファジィ学会、行動計量学会各会員。

島津 光伸 (正会員)

1985 年大阪私立柏原高等学校卒業。同年、日本電信電話 (株) 入社、1999 年より NTT コミュニケーション科学基礎研究所にて人工知能、DSIU の研究に従事。2001 年から NTT コムウェア (株) へ出向、現在、ヒューマンセンタードesignの研究に従事。ヒューマンインタフェース学会会員。