

個人の嗜好情報を考慮した旅行提案システム

Travel Recommendation System Using Individual Preference Information

久保田 賢人† 芋野 美紗子‡ 土屋 誠司‡ 渡部 広一‡
 Kento Kubota Misako Imono Seiji Tsuchiya Hirokazu Watabe

1 はじめに

近年、情報技術の著しい発達により、誰もが手軽に大量の情報を取得することができるようになった。一方で大量の情報の中から自分に適したものを選択することは困難である。そこで、ユーザからの要望が漠然としたものであっても、足りない情報を補い、ユーザに適した情報を提案するシステムが必要と考えられる。

これらの問題を旅行について考える。たとえば「リフレッシュしたい」といった要望があった際、ある人は、運動をすることによって要望を満たし、一方でまたある人は、温泉に行くことにより要望を満たす。このように、同じ要望であっても、人によって目的が違ってくることがある。これには個人の嗜好が関係しているのではないかと考えられる。そこでユーザに適した情報を提案するに当たって、嗜好を考慮することで、より良い提案ができるのではないかと考えられる。本研究では、ユーザの嗜好を汲み取り要望を具体化するシステムとして「個人の嗜好情報を考慮した旅行提案システム」を作成する。このシステムはユーザの「～したい」といった漠然とした入力に対して、大量の情報が存在する Web から旅行することのできる場所を獲得する。その中から個人の嗜好に適した旅行先を出力する。

2 関連技術

2.1 概念ベースと関連度計算方式

概念ベース^[1]とは、複数の国語辞書や新聞等から機械的に構築した語（概念）と、その意味特徴を表す語（属性）の集合からなる知識ベースである。概念ベースには約 9 万語の概念が収録されている。関連度計算方式^[2]とは概念ベースに定義された語と語の関連の強さを 0.0～1.0 の実数値で算出するものであり、関連が強いほど大きな値を示す。

2.2 京大格フレーム

京大格フレーム^[3]とは、動詞とその動詞に関係する名詞を用法ごとに整理したものである。この格フレームは、Web 上の約 5 億文の日本語テキストから自動的に構築されている。また、格フレームに含まれる動詞の数は約 5 万語である。この格フレームを用いることで、動詞からその動詞に結びつく名詞、格、頻度を取得できる。

2.3 オートフィードバック

オートフィードバック^[4]とは、未定義語（概念ベースに定義されていない概念）の意味特徴を表す単語（属性）とその重要性を表す重みの組を Web を用いて自動的に構成する手法である。

2.4 拡張オートフィードバック

拡張オートフィードバック^[5]とは、入力語に対して Web

上で共に出現する単語を獲得する手法である。まず、入力語の Web 検索を行う。獲得した Web 検索結果ページ内での自立語および複合語の出現頻度をより、重み付けを行う。そして、重み順に並び替えた自立語および複合語を出力する。複合語とは「名詞」「アルファベット」「数字」の連続は複合するというルールを用いて獲得できる語である。

3 個人の嗜好情報を考慮した旅行提案システム

システムを使用するにあたって、ユーザに事前に嗜好情報として、個人情報「趣味」「好きな食べ物」「好きなスポーツ」を登録させる。入力として旅行先の都道府県と、名詞、動詞、副詞各 1 個以下の「～したい」といった要望を入力する。以下の図 1 に旅行提案システムを示す。

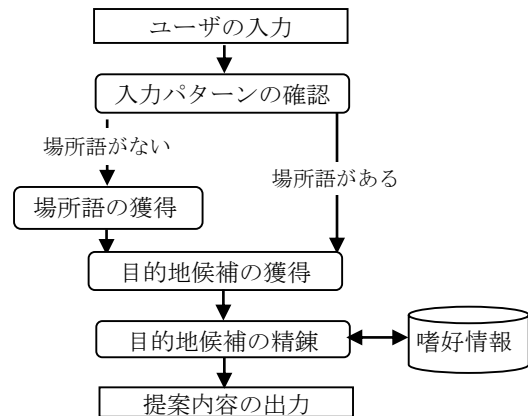


図 1. 旅行提案システム概要図

3.1 入力パターンの確認

入力パターンの確認では、入力された要望文に旅行用場所語があるかを判定する。旅行用場所語とは旅行に関係のある場所の事である。入力文に対して茶筌を用いて形態素解析を行い、すべての形態素に対して旅行用場所判断システムを使用する。旅行用場所判断システムとは、場所判断システム^[6]を基に作成したシステムであり、場所判断システムで場所と判断される単語の中から旅行用場所語を判断できるシステムである。要望文に旅行用場所語と判断される言葉がある場合、目的地候補の獲得を行う。要望文に旅行用場所語と判定される言葉がない場合、場所語の獲得を行う。

3.2 場所語の獲得

場所語の獲得ではまず、形態素解析された要望文の動詞、名詞、副詞に対して京大格フレームを使用する。これにより、要望に対して関連のある言葉を獲得する。この際、原型「する」という動詞は京大格フレームには使用しない。

次に旅行用場所判断システムを使用し、獲得された言葉に旅行用場所語があるかを判断する。旅行用場所語と判断された言葉と、要望文との関連度を算出し、関連度の高い、上位 3 件を今回の要望に対しての旅行用場所語として獲得する。入力が「ゆっくりしたい」の場合の旅行用場所判断システムを使用する例を図 2 に示す。

† 同志社大学大学院理工学研究科

Graduate School of Science and Engineering, Doshisha University

‡ 同志社大学 理工学部

Faculty of Science and Engineering, Doshisha University

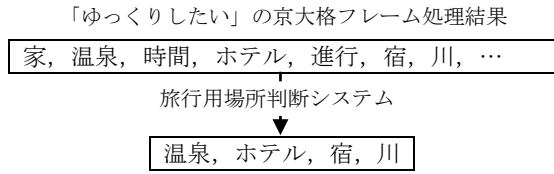


図 2. 場所語の獲得 1

まず要望「ゆっくりしたい」を京大格フレームに使用する。その後、旅行用場所判断システムを使用し、旅行用場所語を獲得する。

次に、関連度を算出し、要望に対しての場所語の獲得をする例を図 3 に示す。

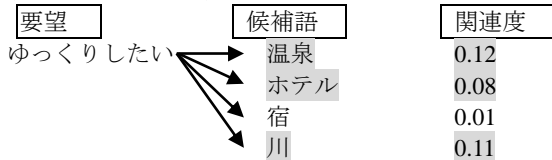


図 3. 場所語の獲得 2

このようにして、関連度の高い旅行用場所語「温泉」「ホテル」「川」の 3 つを獲得できる。

3.3 目的地候補の獲得

目的地候補の獲得では、拡張オートフィードバックで 3.1 節もしくは 3.2 節で獲得した旅行用場所語+要望+目的地域+「旅行」を検索し、具体的な旅行先を獲得する。獲得できた出力を目的地候補とする。語の中に助詞を含むもの、記号を含むもの、「県」「市」「区」「町」「村」などの住所を含むものは旅行先として不適切と判断し、目的地候補から削除する。削除したのち、獲得できた目的地候補すべてにオートフィードバックを使用し、属性と重みを獲得する。それぞれの目的地候補の重み上位 3 件に 3.1 節もしくは 3.2 節で獲得した旅行用場所語があれば、目的地候補として残す。要望「ゆっくりしたい」目的地域「兵庫」の例を図 4 に示す。

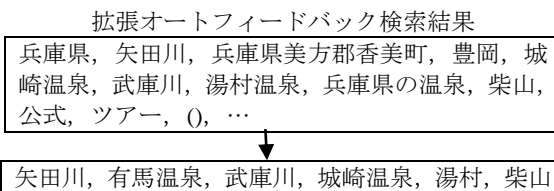


図 4. 目的地候補の獲得

3.4 目的地候補の精練

3.3 節で獲得した目的地候補と 3 種類の嗜好情報でそれぞれ関連度を算出し、関連度の平均上位 3 件の目的地候補を提案内容として出力する。3.3 節までの処理結果と嗜好情報、趣味「温泉めぐり」好きな食べ物「海鮮」好きなスポーツ「ゴルフ」の場合の例を図 5 に示す。

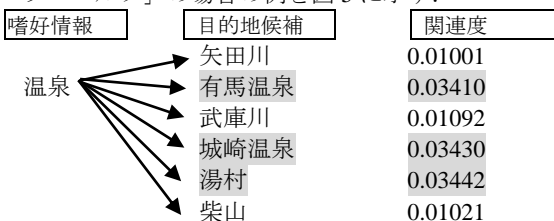


図 5. 目的地候補の精練

このようにして提案内容が「有馬温泉」「城崎温泉」「湯村」と決定する。

4 評価方法・評価結果

本稿で構築したシステムの精度を評価する。ユーザ 12

名に実際にシステムを使用してもらい、評価を行った。出力される提案内容を以下に示す○・△・×の 3 段階で評価させた。

- 行きたいと思える旅行先である。
- △ 場所であり、旅行先としてはおかしくない。
- × 行きたくない。もしくは場所ではない。

ユーザには要望と目的地域を入力してもらい、評価を行った。嗜好を考慮する有用性を見るために、3.3 節で獲得できる目的地候補全と、嗜好を考慮した目的地 3 つの 2 種類を評価した。

4.1 評価結果

評価結果を表 1 に示す。左側が 3.3 節で獲得された目的地候補の評価結果で、右側が本システムの評価結果である。

表 1. 評価結果

	旅行提案システム	本システム
○	53%	57%
△	5%	10%
×	42%	33%

5 考察

評価を取った結果、嗜好を考慮した旅行提案システムは、○と△の合計 67%の精度となった。システムを使用した時 3 つの旅行先が提案される中で、それぞれのユーザへの出力に○が含まれないものはなかった。3.3 節で獲得できる目的地候補の評価の結果は 58%の精度となった。嗜好を考慮した旅行提案システムの評価結果は、それより 9%精度がよかったことを考えると、嗜好を考慮することで、より良い提案ができたと考えられる。図 4 と図 5 を見てみると、個人に適した旅行先を提案できていることがわかる。

6 おわりに

本稿では、ユーザの要望から旅行用場所語を獲得し、獲得された旅行用場所語と要望とで Web 検索を行い個人の要望に関係のある言葉を獲得する。獲得できた言葉をユーザの嗜好情報と関連度を取り、関連度平均上位を提案することにより、ユーザに適した旅行先を提案するシステムを作成した。個人の嗜好情報を考慮した旅行提案システムの精度は 67%である。旅行提案システムの精度が 58%となっているため、嗜好を考慮することにより、より良い提案ができるようになった。

謝辞

本研究の一部は、科学研究費補助金(若手研究(B)24700215)の補助を受けて行った。

参考文献

- [1] 奥村紀之, 土屋誠司, 渡部広一, 河岡司, “概念間の関連度計算のための大規模概念ベースの構築”, 自然言語処理, Vol.14, No.5, pp.41-64, 2007.
- [2] 渡部広一, 奥村紀之, 河岡司: “概念の意味属性と共起情報を用いた関連度計算方式”, 自然言語処理, Vol.13, No.1, pp.53-74, 2006.
- [3] 河原大輔, 黒橋禎夫, “高機能計算環境を用いた Web からの大規模格フレーム構築”, 情報処理学会自然言語処理研究会資料, 2006-NL-171-12, pp.67-73, 2006.
- [4] 辻泰希, 渡部広一, 河岡司: “www を用いた概念ベースにない新概念およびその属性獲得手法”, 第 18 回人工知能学会全国大会論文集, 2D1-01, 2003.
- [5] 梅田司: “Web を用いた未定義概念の属性獲得手法”, 同志社大学工学部知識工学科卒業論文, 2005.
- [6] 手原信太郎, 渡部広一, 河岡司: “共起情報を用いた場所語未知語処理の精度向上”, 電子情報通信学会 2008 総合大会講演論文集, D-5-4, 2008.