

オントロジーに基づく携帯情報端末用レコメンデーションシステムの構築

A Recommendation System for Personal Digital Assistance with Ontologies

安随 晋太郎^{*1} 福田 聡^{*1} 濱崎 雅弘^{*2*3}
Shintaro ANZUI Satoshi FUKUDA Masahiro HAMASAKI

大向 一輝^{*2*3} 武田 英明^{*2*3} 山口 高平^{*1}
Ikki OHMUKAI Hideaki TAKEDA Takahira YAMAGUCHI

^{*1} 慶應義塾大学

^{*2} 総合研究大学院大学

^{*3} 国立情報学研究所

Keio University Faculty of Science and Technology The Graduate University for Advanced Studies National Institute of Information

Here in this paper, we discuss how to build up a recommendation system for PDA, using a domain ontology and individual preference. The domain ontology comes from a commercial directory service. Individual preference comes from positioning data and categorized bookmarks. Our prototype system should support a user in finding the restaurants that the user will like better. In typical cases, when a user is strange at some area, our prototype system gives the recommendation on restaurants to the user, extending user's individual preference with a domain ontology. In order to evaluate this recommendation system, I have done the following case study; how this recommendation system works when users want to know which restaurants are in their favorites in visiting some city unknown to them. This case study has shown that the precision rate is 73% in all restaurants proposed by this system, whereas it was 51% without the system.

1. はじめに

近年、インターネットの普及により多くの情報がウェブ上に溢れ、ユーザが得る情報量は著しく増加した。インターネット上の情報には実世界の事柄やサービスと関係の深いものも多く、そうした情報はポータルサイトを通して屋外からでもアクセスすることができる。こうしたポータルサイトを利用すれば自分の位置情報を利用して最寄りの宿泊先やレストランなどの周辺情報を検索することができる。また、これらの実世界の情報とその地理的な結び付けをPOI(Point Of Interest, 関心地点)として扱うことにより、ナビゲーションなどの近接情報として利用したり、地図上へ表示したりすることも可能である。このように携帯情報端末などを用いて実世界に関する情報を容易に獲得できる環境が整いつつある。

一方、こうした周辺情報検索には、携帯情報端末の表示に適したディレクトリ型のサービスが用いられていることが多く、また、そのようなディレクトリサービスで用いられているディレクトリ構造は一般的に固定的で、巨大なものが多い。

しかし、ユーザごとの興味・関心に関するディレクトリ構造は多様であり、ディレクトリサービスが提供している汎用的かつ固定的な構造では全ての要求に対処できない。また、ディレクトリサービスの提供するディレクトリ構造は巨大であることから、期待する結果を得るまでのプロセスが長くなってしまふ。そのためディレクトリ探索には大きなコストが必要になってしまう。

こうした問題点のもと、操作性の低い携帯情報端末において検索過程や検索結果をユーザの要求や関心に合わせて適切に提供することは、検索プロセスの短縮とユーザの負担軽減につながるものと考えられる。

以上の背景より、本稿では、飲食店情報を対象として領域オントロジーと個人嗜好情報を適用させた携帯情報端末用レコメンデーションシステムを提案する。提案システムでは、エンドユ

ーザが屋外で負担なく周辺情報を得られるような検索システムを目指し、ユーザの個人嗜好情報を領域オントロジーによって拡張することにより、見知らぬ土地においてユーザの嗜好に基づいたレコメンデーションを行う。ユーザが作成する既存の世界の情報を活用した嗜好情報はユーザ個人の嗜好体系を反映していると考えられる。しかし、個別に作成されるユーザの嗜好情報は小規模であるため、プロフィールとして情報が不十分である。そこで、多くの情報を持つ領域オントロジーを併用することにより、コストをかけずにユーザの個人嗜好情報を拡張させ、新しい地域でもユーザの嗜好に基づいたレコメンデーションを行えるようにする。

以下、2章で領域オントロジーと個人嗜好情報の適用、3章では試作システムについて述べ、4章では評価実験について述べる。そして、5章では本稿のまとめと今後の課題について述べる。

2. 領域オントロジーと個人嗜好情報の適用

本稿では、領域オントロジーに個人嗜好情報を適用させることで、ユーザの嗜好体系を拡張させ、ユーザの嗜好に適した情報を取得する。また、領域オントロジーによるインスタンス群の特定と領域オントロジーの分類体系による代替検索を併用することで、携帯情報端末のような入出力に制限ある環境において検索プロセスの軽減が行える。

2.1 領域オントロジー(Domain Ontology)

領域オントロジーはある特定の領域に関するさまざまなクラスやインスタンス群を提供するものである。飲食店情報を例にとると、クラスは「和食」「洋食」などといった分類、インスタンスは各クラスに分類されている個別の飲食店情報を指す(図1)。



図1. 飲食店の領域オントロジーの例

本稿では、個人嗜好が顕著に反映され、かつ個人ごとに嗜好の概念が大きく異なる飲食店を領域オントロジーの対象とした。領域オントロジーと個人嗜好情報の適用のために、飲食店オントロジーを用意する必要があるが、適した飲食店オントロジーが現時点ではまだ提供されていないため、Yahoo!グルメ[13]が提供しているディレクトリ型データベースを対象として RDF と OWL を用いて簡易的に飲食店オントロジーを構築した。これ以降、飲食店オントロジーが提供するクラスを飲食ジャンル、インスタンスを飲食店情報と呼ぶ。

2.2 嗜好情報

本稿では、個人の嗜好体系を表現している個人嗜好情報を利用し、ユーザの嗜好体系に従ったレコメンデーションを行うために利用している。一般にユーザは獲得した情報を整理分類して知識体系を構築していくと考えられる。そのため、本稿ではユーザによって行われる分類自体がユーザの嗜好概念を表すと考える。

ユーザは各自の主観的な嗜好に基づいてユーザカテゴリを作成、分類する。飲食店情報を例にすると、分類のために作成されたユーザカテゴリには、「雰囲気がいい」「景色がいい」などが考えられる(図 2)。飲食店情報を分類するためのユーザカテゴリには、検索結果が個人の嗜好に基づいて再分類される。また、個人嗜好情報が順次追加されていくことで、ユーザの嗜好情報の特徴が拡大されていく。

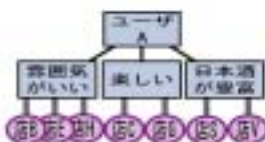


図 2. 飲食店の個人嗜好情報の例

2.3 領域オントロジーと個人嗜好情報の適用方法

提案システムでは、領域オントロジーに個人嗜好情報を適用させることで、ユーザの嗜好体系を拡張させ、ユーザの嗜好に適した情報を取得する。個人嗜好情報を拡張するために 2 段階の処理を行う(図 3)。

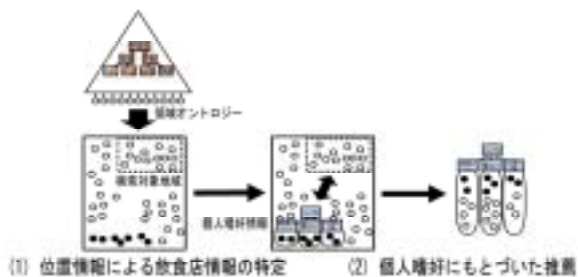


図 3. 領域オントロジーと個人嗜好情報の適用

- (1) 領域オントロジーの飲食店情報群を利用して位置情報をもとにその地域の飲食店を特定する。
- (2) 位置をもとに領域オントロジーから特定された飲食店情報と嗜好情報との類似性を求め、ユーザの嗜好を基にした分類を行う。

2.4 ユーザカテゴリへの再分類方法

飲食店とユーザカテゴリの類似度判定に用いられる飲食店の特徴ベクトルは、あらかじめ計算され飲食店コメントコーパスに格納されている(図 4)。飲食店コメントコーパスに各飲食店の特徴ベクトルが格納されるまでには、2 段階の処理手順を経る。

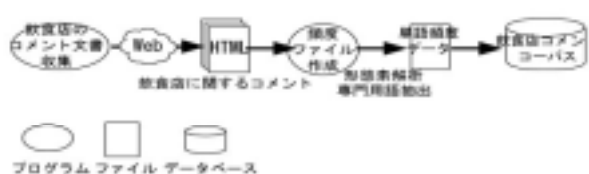


図 4. 飲食店コメントコーパスの構築

- (1) Web 上から飲食店ごとにに関するコメントを収集する。
- (2) 収集された各飲食店のコメント群から単語と出現頻度を算出して各飲食店の特徴ベクトルを作成してコメントコーパスに格納する。

(1) では、飲食店のポータルサイトが提供しているコメントを利用した。

(2) において、(1) で収集されたコメントに対して形態素解析処理[14]と専門用語抽出[15]を行う。抽出された各コメントの単語群の全てに基づいて単語の頻出情報、特徴ベクトルを作成し、それらを合わせて飲食店コメントコーパスとして用意する。文書の特徴ベクトルはベクトル空間モデル[16]に基づいて単語の頻度を用いて作成した。

飲食店との類似度に用いられるユーザカテゴリの特徴ベクトルは、個人嗜好情報として作成されたユーザカテゴリ内に分類された飲食店の特徴ベクトルの和とする。ユーザカテゴリ内の各飲食店の特徴ベクトルは、飲食店コメントコーパスに格納されているものを用いる。また、ユーザカテゴリの特徴ベクトルは、嗜好情報を適用される都度、再計算される。

2.5 類似度計算

検索結果の飲食店情報をユーザカテゴリに再分類するための類似度計算は、飲食店の特徴ベクトルとユーザカテゴリの特徴ベクトルとの類似度を TF-IDF 法を拡張した SMART システム[17]による類似度算出方法を用いて計算を行う。SMART においてユーザカテゴリの特徴ベクトル c と飲食店の特徴ベクトル d との類似度を $\sin(c,d)$ とおくと、 c と d の類似度は、 $\sin(c,d) = \text{SMART}(c,d)$ である。なお、SMART のアルゴリズムによる類似度の計算には GETA[18]を用いている。

3. 試作システム

試作システムでは、エンドユーザが屋外で負担なく周辺の情報を得られるような検索システムを目指し、ユーザの個人嗜好情報を領域オントロジーによって拡張することにより、見知らぬ土地における個人の嗜好に基づいた飲食店のレコメンデーションを行う。本システムは、処理する内容から以下の 3 つの部分から構成されている(図 5)。

- ・ 飲食店検索部
- ・ ユーザ適応部
- ・ 個人嗜好情報編集部

3.1 飲食店検索部

飲食店検索部ではユーザの位置情報をもとに最寄駅を特定して、ユーザの周辺にある飲食店を検索する。位置検索部と飲食店ジャンル選定部の 2 つの処理部により構成されている。

(1) 位置検索部

位置検索部では、ユーザの現在地情報をもとに最寄駅を検索する。ユーザは、GPS 付携帯情報端末から飲食店検索部に対して緯度・経度を位置情報として送信する。緯度・経度の位置情報を受信した位置検索部は、最寄駅情報として利用するために必要な情報が格納されている駅データベースにある駅の緯度・経度情報を順番に参照し、現在地と対象としている駅の

距離を求める。全ての駅との距離を算出し、距離の近い上位 3 件をユーザに提示する。ユーザにより最寄駅を 1 件特定された後、駅データベースにアクセスして該当駅情報を取得する。最寄駅情報を取得後は、該当駅をクエリーとして飲食店情報として利用するために必要な情報が格納されている飲食店データベースから該当駅付近の飲食店数を調べる。最寄駅付近に飲食店が存在しない場合、位置検索部はユーザにその旨を伝え隣接する駅を代替駅として提示、ユーザは代替駅を選択する。新たに駅が選択された後は、その駅の情報を取得し、同様の処理を繰り返す。代替駅の検索は、駅データベースから最寄駅の路線を特定し、その路線上で隣接する駅を代替駅としてユーザに提示することで行っている。

(2) 飲食店選定部

周辺検索を行う最寄駅が確定した後、駅名をクエリーとして最寄駅周辺の飲食店情報を取得する。飲食店情報を取得した後は、ユーザに飲食店ジャンルによる絞り込み検索を行う選択肢を提示する。もし、ユーザが飲食店ジャンルによる絞り込み検索を行う場合、飲食店オントロジーファイルから飲食店ジャンル一覧を表示する。ユーザは提示された飲食店ジャンルの中から絞り込みを行いたい飲食ジャンルを選択して、最寄駅の駅名と飲食店ジャンルの 2 つの条件で再度飲食店データベースから飲食店情報を取得する。最寄駅の駅名と飲食店ジャンルの条件から飲食店情報が見つからなかった場合、飲食店オントロジーの飲食店ジャンルから代替ジャンルをユーザに提示する。ユーザは提示された飲食店ジャンルをもとに、再度飲食店ジャンル選定部で同様の処理を行う。なお、ユーザにより飲食店ジャンルによる絞り込みが行われない場合は、最寄駅周辺の飲食店情報を全て、ユーザ適応部へ送る。

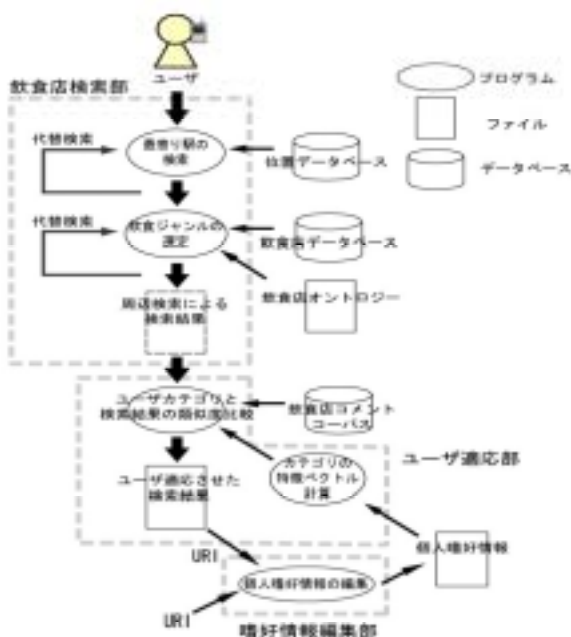


図 5. 試作システムの構成、処理の流れ

3.2 ユーザ適応部

ユーザ適応部は飲食店検索部によって得られた検索結果に対して個人嗜好情報を適用させて、ユーザカテゴリとの類似性を比較する。

飲食ジャンル選定部から送られてきた検索結果はユーザ適応部に送られる。ユーザ適応部は送られてきた検索結果の飲食店情報の特徴ベクトルを飲食店コメントコーパスから参照する。

そして、その特徴ベクトルをユーザカテゴリごとに作成した特徴ベクトルを比較する。比較した結果、類似度が高かったものから順に結果として表示される。ユーザは検索結果を参照するとともに、興味のある飲食店を新たな個人嗜好情報として追加できる。

3.3 個人嗜好情報編集部

個人嗜好情報編集部は、URI ブックマークを登録・削除・ユーザカテゴリ作成など個人嗜好情報を編集する機能を持つ。ユーザは領域オントロジーの飲食店情報の中から興味のあるもの、自分の気に入ったものを登録しておく。登録方法はあらかじめ携帯情報端末や PC などから直接 URI を登録しておく方法と、ユーザ適応部によって出力された検索結果を用いて登録する方法がある。登録された飲食店情報はユーザの好みに従って自由に分類することができ、分類するためのユーザカテゴリも自由にラベル名を付与することができる。ユーザカテゴリと登録された飲食店情報は各個人のプロフィールとして保存される。

4. 提案システムの評価実験

本稿における提案システムの有効性を評価するために、アンケートを用いた被験者による評価実験を行った。

4.1 評価指標

本システムの有効性を測る指標として、ユーザの正答率を定義した。正答率は、提案システムがユーザの嗜好に近いと判断して出力した結果に対して、ユーザが提案システムの出力が適切と答えた割合である。ユーザによる正答集合を J 、評価される対象の出力集合を E とすると、嗜好正答率 (prefer) は以下のよう計算される。

$$prefer = \frac{J}{E}$$

嗜好正答率は、システムが出力したうちでユーザの嗜好を抽出できたものの割合であり、個人嗜好に沿っているかを表している。

4.2 実験準備

被験者には、実験準備として嗜好情報の登録を以下の手順で行ってもらった。

- (1) 被験者にはあらかじめ準備された 20 個のユーザカテゴリから 10 個を選択してもらう。
- (2) 選択した 10 個のブックマークカテゴリに対して飲食店の URL をブックマークとして追加していく。
- (3) インターネットブラウザにより作成されたブックマークによる嗜好情報を HTML ファイルとして出力する。出力された HTML ファイルは、嗜好情報変換モジュールにより嗜好情報データファイルとして変換される。

4.3 実験概要

被験者に銀座を対象として提案システムを用いて飲食店の検索を行ってもらった。実験では、学生と社会人による計 16 名を対象として行った。

飲食店検索は、被験者が全員銀座にいと仮定して銀座の飲食店検索を行ってもらい、提案システムが被験者の嗜好情報として持っているユーザカテゴリに再分類した飲食店を 2 段階で評価してもらった。評価実験では、スコアの高い上位 5 件と、残りの飲食店から適当に抽出した 5 件の計 10 件をランダムに並び替えて表示した。ユーザに出力された飲食店の URL をひとつずつ確認してもらい、その飲食店が被験者が選択したカテ

ゴリに適切に分類されているのかを Yes-No の 2 段階評価で判断してもらい、4.1 で述べたように正答率を求めた。その作業を 10 個それぞれにある 10 件の飲食店 URL、100 個を 2 段階で繰り返し判断してもらった。なお、ここで言うユーザにとって適切とは、ユーザが選択したカテゴリに対して、そこに分類されている飲食店が主観的に適切と感じられたかとした。



図6. 検索結果画面

4.4 実験結果

本システムを用いた場合と汎用飲食店検索エンジンを利用した場合の正答数の平均を求めた。正答数の平均値と標準偏差の一覧を表1に示す。

	平均値	標準偏差
本システム	3.63	0.517
汎用エンジン	2.54	0.955

表1. 実験結果

本システムを利用した場合の正答率の平均値が 73%であったのに対して、汎用飲食店検索エンジンを利用した場合の正答率の平均値は 51%と、平均値に関しては本システムを利用した場合の方が高い結果を得ることができた。また、検定においても統計的にシステムの有意性を示すことができた。

5. おわりに

本稿では、ユーザに適應させた情報提供を目的として、ユーザの嗜好情報拡張による概念検索を提案し、嗜好情報の拡張による概念検索のためにオントロジーに基づく携帯情報端末用レコメンデーションシステムを開発した。

評価実験の結果、本システムの利用によりユーザに対して、個々人の嗜好に基づいた情報を提供することができることを示すことができ、入出力環境に制限ある携帯情報端末に適した情報提供方法を示せた。また、実験対象地を被験者がほぼ行くことのない地域を選択して実験を行ったが、被験者に対して適切な情報提供をすることができたため、見知らぬ土地における利用を目標として開発した本システムの有用性を示すことができた。

今後は、フィルドワークによる有用性の評価、特徴ベクトルによる類似度判定の精度向上、個人の嗜好概念の動的変化、他人の嗜好の利用について考察していく予定である。

参考文献

- [1]Berners-Lee, T., Hendler, J. and Lassila, O.: The Semantic Web, Scientific American, vol. 284, pp.34-43, 2001
- [2]World Wide Web consortium: Resource Description Framework(RDF), <http://www.w3.org/RDF>
- [3]World Wide Web consortium: Web Ontology Language(OWL), <http://www.w3.org/2001/sw/WebOnt>
- [4]Gruber, T.: A translation approach to portable ontology specifications, Knowledge Acquisition, Vol.5, No.2, pp199-220, 1993.
- [5]DARPA Agent Markup Language Program: DAML Ontology Library, <http://www.daml.org/ontologies>.
- [6]Swoogle, <http://pear.cs.umbc.edu/swoogle>
- [7]伊藤 英毅: “オントロジーを利用した知識の共有/再利用”, UNISYS TECHNOLOGY REVIEW 第 64 号, February, 2000.
- [8]佐保田 圭介, 波多野 賢治, 宮崎 純, 吉川 正俊, 植村 俊亮: “ブックマークの階層構造情報を組み込んだ協調フィルタリングによる Web ページの推薦手法”, 電子情報通信学会第 15 回データ工学ワークショップ(DEWS2004), Mar. 2004.
- [9]濱崎 雅弘, 武田 英明, 松塚 健, 谷口 雄一郎, 河野 恭之, 木戸 出 政継: “Bookmark からの共通話題ネットワークの発見手法の提案とその評価”, 人工知能学会論文誌 17 巻 3 号 SP-D, pp.276-284, 2002.
- [10]宮上 大輔, 河合由紀子, 田中克己: “A3:オントロジーの共有によるユーザ適應のためのフレームワークの提案”, 電子情報通信学会第 15 回データ工学ワークショップ(DEWS2004), 2004.
- [11]大島 裕明, 田中 克己: “個人オントロジーを基にした Web 情報検索に関する研究”, 電子情報通信学会第 15 回データ工学ワークショップ(DEWS2004), 2004.
- [12]仲川 ころこ, 高田 喜朗, 関 浩之: “検索目的を反映したカテゴリ構造に基づく WWW 検索支援”, 情報処理学会研究報告, 2000-DBS-120, pp.1-8, Jan. 2000.
- [13]Yahoo!グルメ, <http://gourmet.yahoo.co.jp>.
- [14]松本 裕治, 北内 啓, 山下 達雄, 平野 善隆, 松田 寛, 浅原 正幸: 日本語形態素解析システム「茶筌」version2.0 仕様書 第 2 版, NAIST Technical Report, NAIST-IS=TR99012, 1999.
- [15]中川 裕志, 森 辰則, 前田 朗, 島 浩之: “専門用語自動抽出システム”, <http://gensen.dl.itc.u-tokyo.ac.jp>.
- [16]Salton, g., singhal, A., Buckley, C. and Mitra, M.: “automatic text Decomposition Using Text segments and Text Themes”, Hypertext '96 Proc., pp.53-65, April. 1996.
- [17]Gerard Salton, Michael J. McGill: Introduction to Modern Information Retrieval, McGraw-Hill, 1983.
- [18]高野 明彦, 丹羽 芳樹, 西岡 真吾, 今一 修, 久光 徹: 汎用連想計算エンジン“GETA”, <http://geta.ex.nii.ac.jp>, 2002.