

ブログによる情報収集と推薦技術を用いた 飲食店情報サイトの構築

山下 晃弘[†] 川村 秀憲[†] 山本 雅人[†] 大内 東[†]
[†] 北海道大学 大学院情報科学研究科

あらまし: 近年個人観光客の増加に伴い、インターネットは飲食店や宿泊施設などの観光情報を収集する手段になっている。しかし、そこから個人の嗜好に適合した情報を探し出すことは非常に難しい。一方でインターネット上では Web2.0 と呼ばれるユーザが主体となった新たなシステムが普及している。これらのシステムの特徴としてユーザからの情報提供、ユーザ間での情報共有、ユーザに適応した情報推薦などがあげられる。本稿では、これらの技術を応用し、ユーザ参加型の観光情報サイトの構築を目指す。本システムでは、観光情報を主観的な情報と客観的な情報に分離し、ユーザ間で共有可能なデータベースを構築する。その上で協調フィルタリングを用いた推薦システムを利用し、個人に適応したサイトを実現した。また本稿では、典型的な観光情報である飲食店情報に適応したプロトタイプシステムについて報告する。

キーワード: 個人適応型観光サイト, 推薦システム, ブログ, ユーザ参加型システム

Development of Restaurant Information Website with Weblog and Recommender System

Akihiro Yamashita[†] Hidenori Kawamura[†] Masahito Yamamoto[†] Azuma Ohuchi[†]
[†] Graduate School of Information Science and Technology, Hokkaido University

Abstract: As the number of individual tourists increases, Web browsing is becoming one of the methods to gather tourism information such as restaurants and accommodations. However, it is difficult to search for necessary and valuable tourism information. Meanwhile, the services which are called Web2.0 dramatically increase. These features are information collection from users, information sharing and user adaptive recommendation. This information distribution form will extend increasingly, and the influence on the tourism industry will be significant. This research constructs user-centred tourism information system with objective information and subjective ratings. Then it innovates tourism website using Web2.0 technologies. This study proposes to use the Weblog interface for the system to collect much tourism information. Further, to recommend some products for users, it utilizes Collaborative Filtering Algorithm to resolve which products should be suggested. This paper introduces an application to a restaurant information website.

Keyword: user-adaptive tourism website, recommender system, Weblog, participatory system

1 はじめに

近年、団体周遊型旅行に変わり個人旅行が増加し、旅行形態は一層多様化している。観光資源は、ガイドブックに掲載されるような名所から地元の人しか知らない穴場スポットまで様々である。個人が充実した旅行を実現するためには、これら訪問先の観光資源に関する多くの情報収集と趣味嗜好に適合した情報選択が必要である。

個人が情報を収集する方法として、雑誌やガイドブックなどの書籍を利用する方法と Web サイトを利用する方法が挙げられる。書籍は観光者が利用しやすいように編集され情報がまとまっている一方で、その情報は一般大衆向けであり個人観光客の趣味嗜好に適合した情報が掲載されているとは限らない。また、紙面が限られているために有名でない情報は掲載されず、ニッチな観光資源が発見されない可能性がある。

一方で、Web サイトには膨大な観光情報が存在し、住所や電話番号などの情報だけではなくユーザからの評価や感想なども観光客に必要な情報になっている [4]。特に、近年急速に普及したブログには、様々な情報が個人の主観に基づき自由に記載されている。中には、地元の飲食店や人気スポットを紹介する専門ブログも数多く開設され、これからその土地を観光しようと考えている個人観光客にとっては貴重な情報源になりつつある。しかしその反面、ユーザが自分の嗜好を考慮して大量に存在する情報の中から必要な情報を探し出すことが難しいのが現状である [3]。

本研究の目的は、ブログを始めとするユーザから提供された情報を収集してデータベースを構築し、その情報を利用した個人観光客向け情報サイトを構築することである。その際に協調フィルタリングと呼ばれる推薦技術を利用し、個人に適応した情報提供を実現することで大量に存在する情報の中から個人にとって価値のある情報を提供する仕組みを実装した。

本研究で構築したシステムの特徴は次の通りである。

- Weblog 型インタフェースによる情報収集
- wikipedia に代表される集合知を利用した情報共有
- Amazon.com に代表される個人適応型の情報推薦

また、本稿では代表的な観光情報として飲食店情報を採用し、実際に運用を見据えたプロトタイプシステムの構築について報告する。

2 システム概要

2.1 観光情報の定義

Web 上には様々な観光情報が豊富に存在するが、それらは二つに大別できる。一つ目は、住所や電話番号、飲食店のメニューや価格などコンテンツに対して一意に定まる情報である。これらの情報はユーザ間で共有可能であり、本稿では「客観的な観光情報」と表現する。二つ目は、各個人が主観的に記述したコンテンツに対する感想や評価である。この情報は個人に依存するため共有は不可能である。本稿では「主観的な観光情報」と表現する。本システム

ではこれら二種類の情報をユーザから収集及び提供する。

2.2 システム全体像

本システムは、観光情報データベースを中心とした二つの機能から成り立つ [図 1]。一つは観光情報データベースを利用したブログシステムである。利用者は、一般的なブログと同様 Web 上から記事を投稿し自分のブログサイトを構築する。記事を書くのと同時に関連する観光情報を専用インタフェースに入力することでその情報が観光情報データベースに蓄積されユーザ間で共有可能になる。また、既に蓄積された情報を参照しての投稿や誤った情報を編集することも可能である。

もう一つは、観光客への観光情報提供機能である。システムは、ブログユーザにより蓄積された観光情報を活用した Web サイトを構築し、観光客に提供する。その際にブログ記事への参照やその情報に対するユーザの主観的な感想や評価も表示することで観光客はより多くの情報を得ることが可能である。また推薦機能を実装することで膨大な情報の中からユーザの嗜好を考慮した情報を提示する。

本システムでは、ブログ記事を投稿する際に入力した主観的な評価や、観光情報サイト上で入力した評価を利用し情報を推薦する。本稿では、そのような情報推薦アルゴリズムとして、協調フィルタリングを利用した。協調フィルタリングは推薦対象となるユーザの過去の評価とからそのユーザと嗜好が似ていると思われる別のユーザを検索し、そのユーザの評価を元に推薦を行うアルゴリズムである。これにより、システムを利用し続けることでより個人に適応した情報が提示される。



図 1: システムの利用イメージ

2.3 ブログによる情報の収集

Web 上のブログには近所の飲食店や旅行記などを紹介した記事が豊富に存在し、日々新たな記事が投稿されている。またブログの記事は、その目的が日記であるという特徴から、記事投稿者の主観に基づいた率直な意見であると考えられる。このような膨大な情報を収集するアプローチとしてテキストマイニングによる収集が考えられる。しかし、本稿では収集した情報を元に Web サイトを構築するため、より正確な情報が求められる。また、ブログ記事内に表現された主観的な評価や感想をテキストマイニングで抽出することは非常に難しい。そこで、本稿ではユーザから多くの情報を収集するためにブログインタフェースを用いる。専用インタフェースを実装することにより、ユーザから正確な情報を収集ことができ、また複数のユーザが新規投稿、編集、修正を繰り返すことでより洗練されたデータベースの構築を目指す。

図 2 にブログユーザが記事を投稿する流れを示す。ユーザは一般的なブログと同様にシステムに記事を入力すると同時に、専用インタフェースに観光情報を入力する。既に観光情報データベースに情報が存在する場合は、そのまま引用、または編集して引用することができる。新規に観光情報を入力する際、システムは観光情報を共有データベースに蓄積した後、データベースからユーザの記事に情報を引用する形でブログを更新する。



図 2: ブログを利用した情報収集

2.4 観光サイトの構築と情報共有

蓄積された情報は観光サイトとして観光客に提供される。観光サイトに表示される情報とブログの記事に引用される情報はリンクしており、主観的な情報はユーザ全体が追加、編集することができる [図 3]。これにより、ブログ記事を投稿した時点で詳細な情報が入力できない場合も、後からユーザ全体で修正することができる。

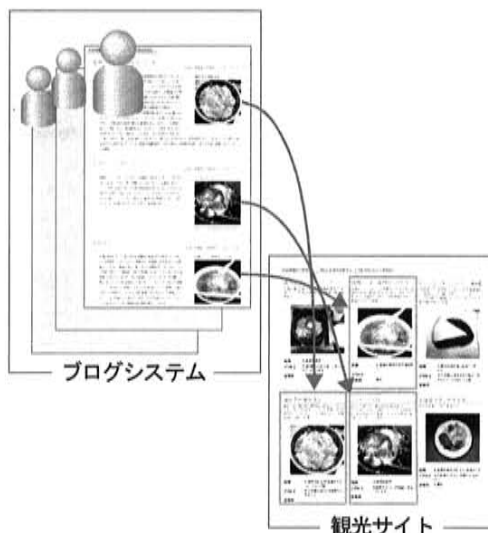


図 3: 観光情報サイトの構築

2.5 飲食店情報の収集

本稿では個人の嗜好が大きく影響する典型的な観光情報として飲食店情報を採用し、実運用に向けて収集する情報の項目を調査した。次に示す大手飲食店情報サイト 5 つを調査対象とし、その結果表 1 に示す項目を収集対象項目とした。

- ぐるなび
- livedoor グルメ
- Yahoo グルメ
- グルメぴあ
- グルメ Walker

表 1: 本システムで収集する飲食店情報

基本情報	料理の種類 (複数選択)
飲食店名	和食 (10 種類)
飲食店名読み	洋食 (11 種類)
飲食店名 (英語)	エスニック (6 種類)
本店または支店名	中華 (6 種類)
チェーン名	ノンジャンル (6 種類)
紹介文	喫茶・甘味 (3 種類)
郵便番号	設備・目的
住所	朝食
アクセス	昼食
電話番号	夕食
FAX	軽食
ホームページ URL	宴会
メールアドレス	合コン
価格帯	デート
営業時間	家族
OPEN 時間	子供入店可
ラストオーダー時間	ペット入店可
CLOSE 時間	駐車場有
ランチタイム	喫煙可
定休日	予約可
	貸切可
	個室あり
	クレジットカード可
	食べ放題あり
	飲み放題あり
	テイクアウト
	夜景がきれい
	海が見える
	景色がきれい
	スポーツ観戦

3 推薦方法

膨大な情報源から必要な情報を見つけ出す一つの方法に情報推薦が挙げられる。協調フィルタリングは「ロコミ」の原理を応用した推薦アルゴリズムであり、被推薦者の評価と他のユーザの評価の類似度から推薦するコンテンツを決定する。協調フィルタリングの基本アルゴリズムは、相関係数法 [5] であり実際のシステムにも多く利用されている。本稿においても相関係数法を採用することにする。次に相関係数法のアルゴリズムについて述べる。

まず、ユーザ集合 $U = \{u_1, \dots, u_n\}$ コンテンツ集

合 $C = \{c_1, \dots, c_m\}$ ユーザ u_i によるコンテンツ c_j の評価を e_{ij} とするとユーザ間の類似度は次の式で計算される。

$$\text{sim}(u_i, u_j) = \frac{\sum_{a \in C_i \cap C_j} (e_{ia} - \bar{e}_i)(e_{ja} - \bar{e}_j)}{\sqrt{\sum_{a \in C_i \cap C_j} (e_{ia} - \bar{e}_i)^2} \sqrt{\sum_{a \in C_i \cap C_j} (e_{ja} - \bar{e}_j)^2}}$$

ただし、 $C_i \cap C_j$ は、ユーザ u_i とユーザ u_j が共に評価しているコンテンツ集合を現し、 \bar{e}_i は、その集合におけるユーザ u_i の評価の平均値、すなわち、

$$\bar{e}_i = \sum_{a \in C_i \cap C_j} e_{ia}$$

である。

このとき、ユーザ u_i にとって未知のコンテンツ c_α の評価推定値 $\hat{e}_{i\alpha}$ は

$$\hat{e}_{i\alpha} = \bar{e}_i + \frac{\sum_{j \in U_\alpha} \text{sim}(u_i, u_j)(e_{j\alpha} - \bar{e}_j)}{\sum_{j \in U_\alpha} |\text{sim}(u_i, u_j)|} \quad (1)$$

ただし、 \bar{e}_i はユーザ u_i が評価した全てのコンテンツ C_i に対する評価値の平均値、すなわち

$$\bar{e}_i = \frac{\sum_{a \in C_i} e_{ia}}{|C_i|}$$

であり、また U_α はコンテンツ c_α を評価したユーザの集合である。従って、ユーザ u_i に推薦するコンテンツは、そのユーザが未知のコンテンツの中で、評価推定値 $\hat{e}_{i\alpha}$ が最大となるコンテンツである。

一般的な相関係数法はこのような計算によって推薦コンテンツを決定する。しかし、このアルゴリズムをどのコンテンツも評価していないユーザに適用すると、他のユーザとの類似度が計算できずに推薦できない。本稿では被推薦対象者の評価数が少ない場合に限りその人に特化せずとも全体に人気のあるコンテンツを推薦することにする。式 (1) において、全てのユーザとの類似度を 1 とすると

$$\hat{e}'_{i\alpha} = \bar{e}_i + \sum_{j \in U_\alpha} (e_{j\alpha} - \bar{e}_j)$$

となり、この値が大きいくほど全体的な人気が高い。本稿では $\hat{e}_{i\alpha}$ と $\hat{e}'_{i\alpha}$ から

$$\arg \max_{\alpha} \{\mu \hat{e}_{i\alpha} + (1 - \mu) \hat{e}'_{i\alpha}\}$$

により推薦コンテンツ α を決定する。ここで μ は個人に特化した推薦と全体に人気のあるコンテンツ

推薦の重みを表しており、 $\mu = 1$ で個人に特化し全体の意見は無視される。本稿では被推薦者の評価数から次の式により μ を計算する。

$$\mu = k \times \left(\frac{2}{1 + e^{-|C_i|}} - 1 \right)$$

ただし、 k は調整用の定数である。

4 実装

4.1 全体構成

図4は本稿で構築したシステムの構成図である。灰色で示した部分が実際に構築した部分である。システムは、データサーバ及び表示サーバの二つのサーバから成り、サーバ間はWebサービスで通信する。データサーバでは、観光情報を蓄積する機能と推薦機能を実装している。また、表示サーバはデータサーバから取得した情報を元にブログを表示したり、観光情報Webサイトを構築する。本稿では、表示サーバの構築を行う基盤としてXOOPSを利用した。XOOPSはPHPにより構築されたフリーのCMS(Content Management System)である。ブログシステムを実現するためには、ユーザを特定し管理する必要があるが、XOOPSでは多くのユーザ管理機能を提供している。また全ての機能はモジュールとして提供されるため機能拡張が容易であり、インストールするだけで利用することができる。本稿で示すブログシステムもXOOPSモジュールとして構築した。

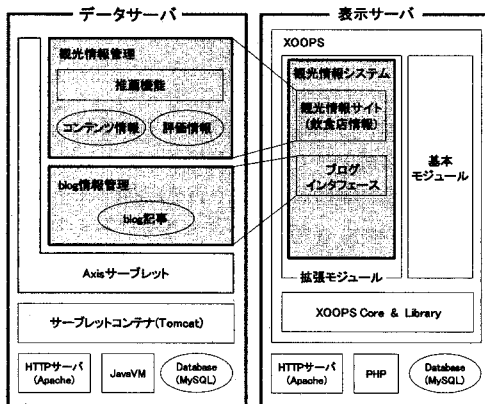


図4: システム構成図

4.2 テーブル構造

データサーバ上でのテーブル構造の概略図を図5に示す。データサーバ上にはブログ記事を保存するブログ情報データベースと、観光情報データベースが存在する。観光情報データベースには、住所や電話番号などを保存する共有情報テーブルと、主観的な評価やコメントを保存する評価情報テーブルが存在する。表示サーバはこれらのデータベースから情報を取得し、ブログサイトや観光情報サイトを構築する。

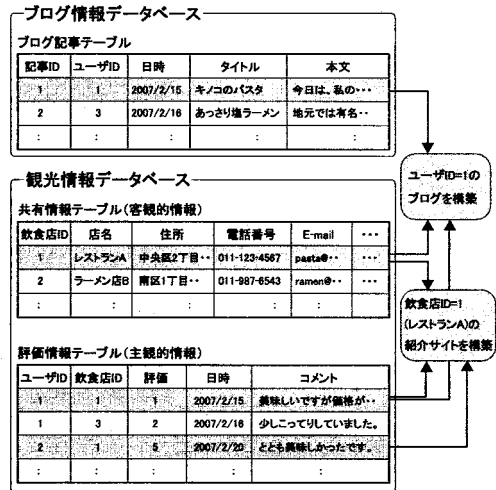


図5: テーブル構造概略図

4.3 プロトタイプシステムの実装

実際に構築したプロトタイプシステムの動作画面を、図6、図7に示す。図6はある飲食店についての詳細情報画面である。上側に飲食店に関する基本情報や地図情報が掲載されており、一番下にその飲食店に対して評価を行ったユーザの評価一覧が表示されている。また、図7はブログ記事投稿画面における観光情報投稿フォームである。ここでは、既に観光情報データベースに存在する情報を検索し、引用する場合のインタフェースを示している。



図 6: システム動作画面 1 (飲食店詳細情報)



図 7: システム動作画面 2 (ブログ記事投稿フォーム中の観光情報引用部分)

5 まとめ

本稿では、Web2.0と呼ばれる技術を応用し、個人観光客向けの観光情報サイトを構築した。情報収集には近年急速に普及しているブログのインタフェースを利用しユーザーから提供される情報の利用を試み

た。観光情報は、観光資源の住所や電話番号などの客観的情報と、主観に基づいた評価やコメントなどの主観的情報に大別され、本システムではこれら独立に蓄積するデータベースを構築している。これにより、推薦システムの利用が可能となりユーザーの主観に適合した情報を提供する観光サイトを構築した。また、代表的な観光情報である飲食店情報に適合し、プロトタイプシステムを構築した。

参考文献

- [1] Gediminas Adomavicius and Alexander Tuzhilin "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions" IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.6, JUNE 2005 pp. 734-749.
- [2] Hung-Wen Tung, Von-Wun Soo "A Personalized Restaurant Recommender Agent for Mobile E-Service", Proceedings of the 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'04)
- [3] N.Good, J. B. Konstan, J. A. Borchers, B. Sarwar, J. Herlocker and J. Riedl "Combining Collaborative Filtering with Personal Agents for Better Recommendations" Proceedings of the 1999 Conference of the American Association for Artificial Intelligence (AAAI-99), pp. 439-446. (1999)
- [4] F. Ricci and René T.A. Wietsma. "Product Reviews in Travel Decision Making" Information and Communication Technologies in Tourism, 296-307. ENTER'06, Proceedings of the International Conference in Lausanne, Switzerland 2006, Springer: Wien-New York(2006)
- [5] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. "GroupLens: An Open Architecture for Collaborative Filtering of Netnews" In CSCW '94: Conference on Computer Supported Cooperative Work (Chapel Hill, 1994), ACM, pp. 175-186.