L_057

Weblogによる評価情報収集と推薦システムの開発

Evaluation Gathering with Weblog and Development of Recommender System

山下 晃弘† 川村 秀憲† 大内 東†
Akihiro Yamashita Hidenori Kawamura Azuma Ohuchi

1. はじめに

近年，Web上に存在する情報資源は急速に増加している．推薦システム(Recommender System)は，そのような膨大な情報の中からユーザ嗜好という感覚的側面を考慮し，効果的に提示するシステムである．推薦システムが推薦対象とする「もの」は，映画や音楽から風景やニュースまで，世の中のあらゆる事物である．書籍を対象とした代表的な推薦システムがAmazon.comである．しかし，推薦システムを構築する上で，ものに関する情報やその評価，及びユーザの嗜好データを如何に収集するかが課題となる．推薦システムは，ものに対する評価全体とユーザ嗜好の一致性から推薦情報を決定する．そのため，情報が少ない場合はコールドスタート問題[Schein 2002]により適切なお薦めができない．一方で，近年個人によるWeb上への情報発信が増加している．特にblogには，様々なものついての詳細情報やその評価が記述されている．個人が発信する情報から必要な情報を推薦システムに取り込めれば非常に効率的であるが，自由記述文章からテキストマイニングを利用して情報を自動抽出するのはまだまだ困難である．

本研究では，世の中のあらゆるものに関する情報を共有しユーザに推薦するシステムの構築を目指す．その上で，一般ユーザから情報収集を行うインタフェースとしてblogを利用する．本稿では，プロトタイプとして映画と飲食店を対象に構築したシステムについて述べる．

2. システム概要

推薦システムを実現するためには，推薦対象となる「もの」の情報とユーザの評価，及びユーザの嗜好情報が必要となる．ここで，ものの属性には，名称，位置，形状，ジャンルなどが考えられ，またそれに対するユーザ評価の集合も一つの属性と捉えることができる．本稿では以後推薦対象となるものをコンテンツと表現する．一方でユーザの嗜好には，趣味やニーズなどが関係し，正確に取得することは事実上不可能である．一般的な推薦システムでは，嗜好情報としてユーザのコンテンツに対する評価を利用する．また，コンテンツの属性やユーザの嗜好は不変であるとは限らず，日々変化していると捉えなければならない．

推薦システムは，コンテンツの属性とユーザ嗜好情報の一致性からユーザに適切な情報を提示するシステムである．類似した機能に検索が挙げられるが，検索はある情報の集合から条件にマッチした情報を抽出する機能である．本稿において推薦とは，この抽出された情報に対して，優先度に応じたランキングを付加する機能と捉える．従って，与えられた情報集合をどのようにランク付けをするかが推薦アルゴリズムである．

推薦アルゴリズムは，ユーザ嗜好を反映したランキングを行う必要がある．また，意図的な情報操作ができない仕組みが好ましい．本稿において，このような仕組みとして協調フィルタリングを利用する．協調フィルタリングは「口コミ」の原理を利用しており，ユーザ間の類似性からコンテンツの評価値を推定するアルゴリズムである．自分と類似しているユーザの影響が大きくなるため，悪意あるユーザが行う情報操作の影響は少ない．また各ユーザは，自分の嗜好を反映させるために正直な評価を行うことが得策となる．

blogは，Web上での日記風サイトであり，近年その手軽さから個人がWeb上に情報を発信するツールとして増加している．その内容は，日常の出来事から専門的な話題まで多種多様であり，日々新しい記事が投稿されている．記事には個人の嗜好や考えが強く反映されており，記事が対象としているコンテンツの情報も豊富である．そこで，blogの記事からコンテンツの属性やその評価を得られれば，推薦システムの規模拡大や精度向上に大きく貢献できると考えられる．また，blogは更新頻度が高いため，多くの情報を得られると共にコンテンツ属性やユーザ嗜好の変化に対応可能である．

blogの記事は一般に文章として記述されているため，そのまま情報を抜き出すことは困難である．そこで本稿では，コンテンツの情報やその評価を入力するフォームを用意する．このフォームはコンテンツのジャンルごとに予め決められた項目で用意されているので，ユーザは任意の項目に情報を入力する．その後blogとしての文章を書き，投稿する．ここで最も重要なことはblogとしての手軽さを失わないことである．フォームを利用する際にユーザは入力したい項目のみ記述すれば良が，従来どおり文章だけの投稿も可能とした．この場合は従来の記事と同様に扱う．また，他のユーザが登録したコンテンツ情報に対して評価や記事を書くことも可能である．本稿では飲食店と映画についてのフォームを作成し，プロトタイプシステムとして構築した．

推薦システムは，blogユーザへの情報提供のみでなく，携帯電話やカーナビへの応用も可能である．例えば，飲食店に関する情報をblogから取得している場合，外出中のユーザに対してその嗜好に合った近隣の飲食店を提示することが可能である[図 1]．ただし，携帯電話やカーナビに情報提示する場合は専用のユーザインタフェースを構築する必要がある．そこで，本稿では推薦システムをWebサービスとして構築することで，汎用的なシステムの構築を実現した．本稿ではこのサービスを推薦サービスと呼ぶことにする．blogシステムは，推薦サービスを利用して情報の登録や推薦コンテンツを取得する．

†北海道大学大学院情報科学研究科

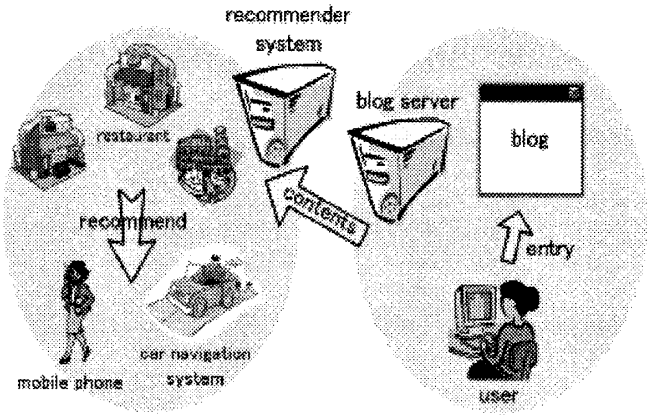

図1：推薦システムの応用

3. 推薦アルゴリズム

協調フィルタリングは，他のユーザの評価を利用して推薦するコンテンツを決定するアルゴリズムである．ユーザ間の類似度から各コンテンツについてのスコアを計算し，最もスコアの高いコンテンツをユーザに推薦する．本稿では，このスコアを用いてコンテンツのランク付けを行う．ユーザがキーワードやジャンル，又は位置情報による検索を行った際に複数のコンテンツ候補がある場合は，このランキングに従ってユーザに情報を提示する．従って，検索結果に合致し，かつ最もユーザの嗜好に合ったコンテンツが上位に表示される．

協調フィルタリングには，さらに相関係数法や逐次的2項関係学習法などいくつかの手法が提案されているが，本研究では最も広く普及し様々なコンテンツに適用されている相関係数法を採用する．ここでは，相関係数法についてそのアルゴリズムを述べる．

相関係数法は，まずユーザ間の類似度を計算し，類似度の高いユーザの評価を重視して推薦するコンテンツを計算する．ここで，ユーザ集合 $U=\{u_1,...,u_n\}$，コンテンツ集合 $C=\{c_1,...,c_m\}$，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．

$$sim(u_i,u_j)=\frac{\sum_{a\in C_i\cap C_j}(e_{ia}-\bar{e}_i)(e_{ja}-\bar{e}_j)}{\sqrt{\sum_{a\in C_i\cap C_j}(e_{ia}-\bar{e}_i)^2}\sqrt{\sum_{a\in C_i\cap C_j}(e_{ja}-\bar{e}_j)^2}} \quad (1)$$

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

ここで，C_i はユーザ u_i が評価したコンテンツの集合である．ユーザ u_i に推薦するコンテンツについて考えた場合，前述の類似度を利用してユーザ u_i にとって未知のコンテンツすべてについて，評価値の推定値を計算する．この推定値が，コンテンツのスコアとなる．今，ある未知のコンテンツ c_α の評価推定値を $\hat{e}_{i\alpha}$ とすると，

$$\hat{e}_{i\alpha}=\tilde{e}_i+\frac{\sum_{j\in U_\alpha}sim(u_i,u_j)(e_{j\alpha}-\bar{e}_j)}{\sum_{j\in U_\alpha}|sim(u_i,u_j)|} \quad (2)$$

$$\text{ただし}\quad \tilde{e}_i=\frac{\sum_{a\in C_i}e_{ia}}{|C_i|}$$

となる．この推定値が最も高いコンテンツをユーザに提示する．

一般的な相関係数法では前述の方法により推薦するコンテンツを決定するが，どのコンテンツも評価していないユーザを仮定して相関係数法を適用すると，他のユーザとの類似度の計算が出来ずコンテンツを推薦できないという結果になる．本稿においてこのような場合には，その人に特化せずとも全体的に評価の高いコンテンツを推薦することにする．式(2)において全てのユーザとの類似度を1とすると，

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

となり，この値が大きなコンテンツほど全体的に人気があると考えられる．個人の嗜好を反映したコンテンツを推薦するか，全体に人気のあるコンテンツを推薦するかは，推薦対象となるユーザの評価数の他に，コンテンツ数や全評価数に影響される．本稿では単純化のため，推薦対象となるユーザの評価数のみを考慮し次の式によって推薦するコンテンツを決定する．

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

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

μ は，個人の嗜好を反映したコンテンツによる推薦と，全体に人気のあるコンテンツの推薦のどちらを重視するかという重みを示す．また k は，重みを調整するための定数である．

4. 実装

4.1 全体構成

図2は本稿で構築したシステムの概略図である．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blogシステム及びWebサービスとして提供される推薦サービスによって構成される．

本稿では，blogシステムの構築を行う基盤としてXOOPSを用いた．XOOPSはPHPにより構築されたフリーのCMS(Contents Management System)である．blog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blogシステムもXOOPSモジュールとして構築している．

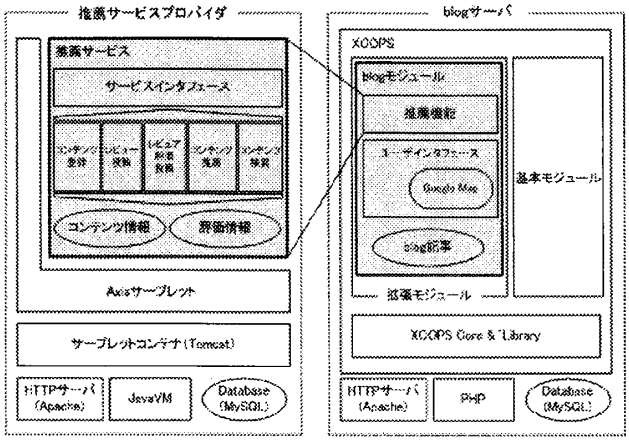

図2：システム概略図

4.2 推薦サービスプロバイダ

推薦サービスプロバイダは，コンテンツ情報及びその評価情報の管理と，ユーザの嗜好を考慮した情報提示機能を Web サービスとして提供する．全ての情報はここに集められ，推薦を行うためのデータとして利用される．本稿では Web サービスの実装として SOAP を利用し，通信ライブラリの Axis を使用した．

推薦サービスとして管理する情報は，コンテンツ情報とその評価情報である．本システムでは様々なコンテンツに対応する為に，コンテンツのジャンル毎に属性の項目を管理し，後から新たなジャンルの追加を可能にした．また，あるコンテンツに対してどのような属性が存在するかを事前に正確に判断することは困難である．そこで，後から項目の追加・削除が行えるようデータは全て項目名と値のペアで管理している．ただし，評価情報に関しては推薦アルゴリズムとの整合性を確保する為に項目名を予め定めている．

4.3 blog モジュール

blog モジュールは，XOOPS 上で記事の投稿や表示のユーザインタフェースを提供する．XOOPS の機能を利用することでユーザを特定し，過去に投稿した記事の一覧や，推薦コンテンツの表示機能を提供する．また，コンテンツの詳細情報は，ユーザによる情報の追加や修正が可能である．その他の表示機能として，ログインユーザ以外の記事表示や，コンテンツごとの記事表示を提供する．一方，記事の投稿時には，blog 情報，コンテンツ情報，及びその評価情報に分割し，コンテンツ情報と評価情報を推薦サービスに登録する．その他全てのユーザインタフェースは XOOPS が提供するテンプレートエンジンによって HTML 化するため，レイアウトや配色は容易に変更可能である．

システムを利用する際，ユーザはまず XOOPS にアクセスし，ユーザ名とパスワードを用いてログインする．その後は一般的な blog と同様に利用する．ただし，新しく記事を投稿する際に，音楽や飲食店などの情報を専用のフォームに記述することで，独立したコンテンツ情報として登録することができる．この時，コンテンツの名称以外の全ての項目は任意であるため，入力は自由である．また，あるコンテンツに対しての記事には，その評価を入力することができる．コンテンツ情報やコンテンツに対する評価は，Web サービスを通して推薦サービスプロバイダに蓄積され，ユーザに提示するコンテンツのランキングを決定する際に利用される．ただし，blog としての利便性や自由度を損なわないようにするため，コンテンツを特定しない記事の投稿も可能とした．この場合は，従来の blog と同じ手順となる．

blog モジュールは，推薦サービスプロバイダと連携してコンテンツの登録や推薦を行う．推薦サービスを利用するための SOAP 通信には，PHP5 から標準でサポートされている SOAP 関数を利用する．また，Google Map が提供する地図表示機能を利用することで，飲食店などの位置属性を持つ情報を地図上に表示することができる．

4.4 飲食店への適用

本稿で構築した blog システムを利用し，飲食店情報を取得する例について述べる．まず，ユーザが記事を投稿する際に，飲食店専用のフォームから飲食店情報を入力する．フォームには次に示す項目が入力可能である．

- 店名（必須）
- 紹介文
- 住所
- アクセス
- 電話番号
- Web
- E-Mail
- 営業時間
- 定休日
- 予算
- クレジットカード利用
- 予約
- 座席数
- 貸切
- 個室
- 座敷
- 駐車場
- 写真
- 地図上での座標

飲食店に関する属性は，全てが普遍であるとは限らない．例えば，予算や貸し切りの可不可などは，年月が経過するにつれて変わる可能性があり，そのためユーザ嗜好とのマッチングに影響を及ぼす可能性がある．このような属性を持つコンテンツについての情報を blog から収集可能であることは，非常に効率的であると考えられる．また，地図上での座標は，GoogleMap を使用して入力することが可能である．登録する飲食店が存在する位置を GoogleMap にてポイントすることで緯度，経度情報が登録される．登録された飲食店の詳細情報画面を図3に示す．

飲食店情報を入力後，ユーザはその飲食店について blog のコメントや評価を記入する．本稿において，評価は総合的な観点からの5段階評価とした．この評価値は推薦サービスにおいて利用される．一般的には，評価値をベクトルとして扱うことで複数の項目で評価することも可能であるが，今後の課題とし本稿ではプロトタイプとして一つの項目とした．登録した飲食店情報を元に投稿した blog の記事を図4に示す．blog 記事では自動的に飲食店の情報が挿入される．

4.5 映画への適用

2 つ目のプロトタイプシステムとして，映画についての登録フォームを作成した．映画は，飲食店とは全く異なる属性を持ち，またその属性は時間の経過による変化が

図3：飲食店詳細情報

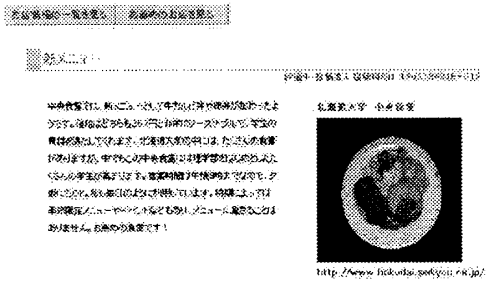

図4：blog記事

ないと考えられる．映画についての項目は次のとおりである．

- 映画名
- 監督
- 主演（4名登録可）
- 公開年
- ジャンル
- キーワード

映画についての評価も飲食店同様総合的な5段階評価とした．図5は登録されている映画の一覧表示とログインユーザへの推薦映画を表示した画面である．

5. まとめ

本稿では，世の中に存在するコンテンツについての情報収集の方法として blog を使用する仕組みを提案した．これにより，コンテンツについての詳細な情報とユーザ評価，及びユーザの嗜好情報を収集し，コンテンツの属性とユーザ嗜好との一致性から情報の適切な提示が可能であることを示した．また，そのような情報提示をすることで blog 記事を投稿するインセンティブ向上や，不正な評価の影響を抑えることが可能であり，よりコミュニティの発展が期待できる．さらに，収集した情報は携帯電話やカーナビに応用することでその利用範囲を拡大することができる．本稿ではそのような仕組みを実現するために飲食店と映画についてのプロトタイプシステムを構築した．その際に情報のランキングを行う推薦機能を Web サービス化することで，より汎用的なシステムを構築できた．今後はより多くのコンテンツに対応し，実際の運用を通じてコンテンツを収集する．ただし，協調フィルタリングを用いたコンテンツのランク付けアルゴリズムは，情報が少ない場合の対応などまだ検討の余地があると考えている．

図5：映画一覧と推薦

参考文献

[1] 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)

[2] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. (1994). 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.

[3] 中島伸介，田中克己 "信用度に基づく blog 情報フィルタリング" 日本データベース学会 Letters Vol.3, No.2, pp.105-108

[4] 小原恭介，山田剛一，絹川博之，中川裕志 "Blogger の嗜好を利用した協調フィルタリングによる Web 情報推薦システム" The 19th Annual Conference of the Japanese Society for Artificial Intelligence, 2005

[5] Eliseo Reategui, John A. Campbell, Roberto Torres "Using Item Descriptors in Recommender Systems" AAAI Workshop on Semantic Web Personalization, San Jose, USA, 2004.

[6] 古川忠延，松澤智史，松尾豊，内山幸樹，武田正之 "Weblog ネットワークの特徴とユーザの行動に関する分析" The 19th Annual Conference of the Japanese Society for Artificial Intelligence, 2005