

M-092 モバイル環境におけるユーザの嗜好と位置を利用した情報推薦システムの検討

A Study on Recommender System using User Preference and Location for Mobile Application

金田 瑞規† 亀山 渉† 渡辺 裕†
Mizuki KANADA Wataru KAMEYAMA Hiroshi WATANABE

1. はじめに

近年高機能なモバイル端末が普及し、モバイル端末からインターネットに接続して情報を得る環境が整備されている。また、GPS等の位置測定デバイスを搭載した端末が登場し、位置情報を用いたモバイル端末向けの情報配信サービスが展開されている。

一般に、モバイル端末はPC等と比較するとディスプレイ能力が劣っている。そのため限られた画面により効果的なコンテンツを表示する技術が必要となる。また、コンテンツ数が増加するにつれて通常の検索システムではユーザが有用な情報にたどり着くのが困難となりつつある。

このような背景から、ユーザにより効果的な情報を提示するために、適応型の情報推薦システムが注目されている。

従来の適応型情報推薦システムには、ユーザの利用履歴等を用いた協調的手法が用いられている。しかし、それらの手法には計算コストやシステムのスケーラビリティ等の問題点が存在する。

そこで本稿では、地理情報に付加されているメタ情報を用いてユーザのクラスタリングを行うことで上記の問題点を解決するシステムの提案を行い、その基礎実験を行う。

2. パーソナライズ技術

ユーザの状況や環境に応じて提供するコンテンツを選択する技術を一般的に「パーソナライズ技術」と呼ぶ。このパーソナライズ技術には大きく分けて「静的マッチング方式」、「ルールベース方式」、「協調フィルタリング方式」の3種類が存在する [1]。

最初の二つは人手を介するシステムである為、現在は自動的にコンテンツを推薦してくれる協調フィルタリング方式が多く用いられている。

この協調フィルタリング方式はユーザの利用履歴を用いて自動的にコンテンツを推薦する方式である。この協調フィルタリング方式は今日、非常に多くのWebサイトで用いられている。以下、この協調フィルタリング方式の概要と問題点を簡単に述べる。

2.1 協調フィルタリング方式

協調フィルタリング方式は、「似ている人の考えは、また似ているに違いない」という経験則に基づいた手法である。具体的には、個々のユーザのアクセス履歴等を分析し、そのユーザと似た因子を持つユーザを探し出し、そのユーザの持つ利用履歴に基づいてコンテンツの推薦基準を定める手法である。

協調フィルタリング方式は、基本的に以下の数式で表現されるユーザ a に対するコンテンツ j の推薦値に基づいて情報推薦を行う [2]。

$$P_{a,j} = \bar{v}_a + \kappa \sum_{i=1}^n w(a,i)(v_{i,j} - \bar{v}_i) \quad (1)$$

このうち、 n はユーザ数、 κ は正規化用の変数、 $w(a,i)$ はユーザ a とユーザ i の関係を表す重み値、 $v_{i,j}$ はユーザ i のコンテンツ j に対する評価値 (ページの閲覧回数等) である。重み値 $w(a,i)$ は利用履歴が近いほど大きな値を取り、逆に利用履歴が異なるほど小さな値を取る。

2.2 協調フィルタリング方式の問題点

協調フィルタリング方式には下記に示す問題点が報告されている [3]。

● システムのスケーラビリティ

協調フィルタリング方式は基本的にユーザのコンテンツに対する評価を基に推薦しているため、新規に加わったコンテンツの様に誰も評価していない、あるいはほんの少ししか評価していないといったコンテンツを推薦できないという問題点が存在する。

● 処理コスト

コンテンツの推薦順位を決定するために、(1) 式を全ユーザ、全コンテンツに対して処理する。そのため、ユーザ数やコンテンツが増加すると処理時間が膨大になってしまうという問題点がある。

これらの問題点は、システム応答の高速性や情報の新規性が強く求められるモバイル環境においては深刻な問題点である。

スケーラビリティに関しては、ロボットがコンテンツを自動採点するシステムが研究されている [4]。しかしこの手法は、あくまで文法等の客観的な評価しかできないため、良好な結果は得られていない。

そこで本稿では、上記の問題点の改善を目的としてモバイル環境に適応したシステムを提案し、その基礎実験と検討を行う。

3. 提案システム

モバイル環境での情報推薦や情報検索のアプリケーションは、特に応答時間 (ユーザが要求を出してからシステムがユーザの端末に情報を提示する時間) が重要である。

したがって、システムとしては以下の3点を考慮する必要がある。

- 効率的なデータ管理
- 効率的なユーザ管理
- 効率的な推薦・検索方式

本研究ではこのうち、効率的なユーザ管理と推薦・検索方式の2点に着目してシステムの検討を行う。

3.1 ユーザクラスタリングを用いた方式

本研究では、「ユーザ1人毎に傾向を把握してコンテンツを推薦するよりも、グループ単位で行うほうが処理コストを軽減させることができるだろう」という仮定に基づき、ユーザを利用履歴を用いてクラスタリングして管理し、クラスタ単位で処理を行うという手法を提案する。提案システムの概念図を図1に示す。

ユーザクラスタリングは、利用履歴を特徴量として特徴空間にマッピングし、その距離に基づいてクラスタリングを行う。すなわち、特徴空間上での距離が近いユーザは利用履歴が似ている事を表し、利用履歴が似ている集団でクラスタを構成することになる。ユーザクラスタリングの概

† 早稲田大学大学院 国際情報通信研究科
Graduate School of Global Information and
Telecommunication Studies, Waseda University

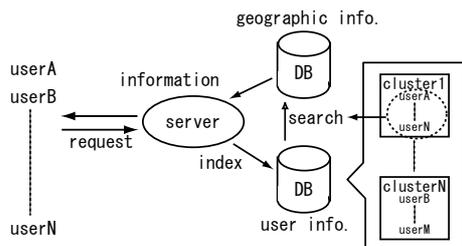


図1 提案システム

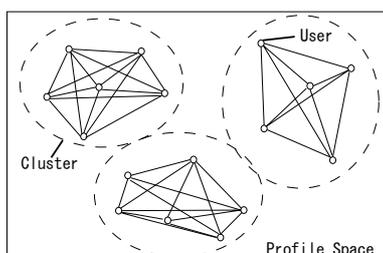


図2 処理の概念図

念図を図2に示す。

コンテンツに対する利用履歴を特徴量として用いると次元数が多く、クラスタリングにかかる演算量が膨大になってしまう。そのため、本研究ではクラスタリングに用いる特徴量として、コンテンツに付加されているメタ情報(「ジャンル」等)を用いる方式を提案する。

クラスタ単位での処理手順は以下の通りである。

1. ユーザの利用履歴を基にクラスタリングを行う
2. 作成されたクラスタ毎にユーザデータを管理する
3. ユーザからの要求があった場合は、クラスタの内部で協調フィルタリングを行う

以上の事を踏まえて、次章で基礎実験を行う。

4. 基礎実験

実際にクラスタリングによってどの程度処理時間が減少するのかを調べるために基礎実験を行う。クラスタリング手法には、現在最も一般的と言われている k-means 法を用い、クラスタ数を変化させてシステムの処理時間を測定する。

4.1 実験条件

実験に用いたパラメータを表1に示す。

また、実験に用いたPCのスペックは、CPU:1GHz, RAM:512M, OS:Windows2000である。

これらの条件で、ユーザが一人一回システムを利用した場合の処理時間を比較する。なお、今回はクラスタリングにかかる演算量は考慮しない。

また、ユーザの利用履歴や条件の中で重要だと思われる、時系列データと場所のデータは本実験では用いない。

4.2 実験結果

実験結果を図3に示す。

結果より、通常の協調フィルタリング手法に対してクラスタリングを適用することによって、大幅に演算コストが減少することが確認できた。

4.3 考察

ユーザ数を n とすると、 n 人が協調フィルタリングを用いた場合の総処理時間は $O(n^2)$ となる。したがってユーザ数が増えるほど処理時間がかかってしまう。一方提案システムの場合は前処理としてクラスタリングを施してユー

表1 実験パラメータ

ユーザ数	100,500,1000,2000,3000,4000,5000
コンテンツ数	50
ジャンル数	10
クラスタ数	5,10,20,30,40,50

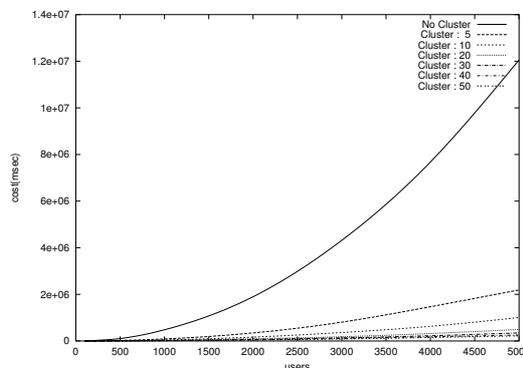


図3 処理時間の比較

ザ数を減らしているため、従来手法に比べ処理時間を大幅に減少させる事ができる。

ここで問題となるのがクラスタ再構築のタイミングである。ユーザの嗜好の変化をより反映するためには、頻繁にクラスタを再構築する必要がある。しかし、クラスタリング処理自体にも多くの処理時間がかかってしまうために、頻繁なクラスタ再構築は処理時間の増加を招いてしまう。したがって、的確なクラスタ再構築のタイミングと効率的な再構築の手法を検討する必要がある。

また、推薦精度にも注意する必要がある。いかに処理性能が向上しても、推薦精度が落ちてしまえば、システムとしての性能は向上したとは言えない。したがって、より推薦精度の劣化が少ないクラスタリング手法と、推薦手法を検討する必要がある。

5. まとめ

本稿では、適応型情報推薦システムに現在用いられている協調フィルタリング方式の概要と問題点を述べた。また、その問題点を改善し、モバイル環境に適したシステムの検討と基礎実験を行い、それを報告した。

検討課題としては以下の点が挙げられる。

- 提案手法における検索精度の調査実験
- 有効なユーザクラスタリング方法の検討
- クラスタ再構成のタイミングの検討
- 場所の人気度 [5] や時間の利用

今後は上記を考慮したシステム設計を行う予定である。

参考文献

- [1] 新井 範子, 北川 和裕, "ユーザの好きと嫌いを聞き分けるパーソナライゼーション技術," 日経エレクトロニクス, pp.127-138, 2003
- [2] John S.Breese 他, "Empirical Analysis of Predictive Algorithm for Collaborative Filtering," 14th Conference on Uncertainty Artificial Intelligence, 1998
- [3] Bamshad Mobasher 他, "Discovery of Aggregate Usage Profiles for Web Personalization," Web KDD Workshop, 2000
- [4] Badrul M.Sarwar 他, "Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System," ACM conference on Computer supported co-operation work, 1998
- [5] 吉沢 進, 亀山 渉, "モバイル環境における行動履歴を利用した情報配信システムの実装と評価," 信学総大 B-15-19, 2003