

携帯電話によるレコメンデーションシステムの一般ユーザによる評価実験 Subjective Evaluation of Recommender System for Mobile Phone

2C5-1

小野 智弘^{*1} 本村 陽一^{*2} 麻生 英樹^{*2}
Chihiro Ono Youichi Motomura Hideki Asoh

^{*1} KDDI 研究所
KDDI R&D Laboratories, Inc.

^{*2} 産業技術総合研究所
AIST

Movie preferences change according to not only the users' personality, but also to the situation such as mood, location, accompanying person, and so forth, and the mobile user may access the recommender system under various situations. We are developing a movie recommender system for mobile phone users that provides context-aware personalized recommendations. In this paper we describe the result of subjective evaluation from various viewpoints such as interface, usefulness, and satisfaction.

1. はじめに

近年、携帯電話経由のインターネットアクセスが増加していることにより、サービスの多様化とユーザの価値観の多様化とともに、ユーザがサービスを利用する状況も家の中や街中など多様化している。このような現状を踏まえ、筆者らは、携帯電話等のユーザが持ち歩く端末からの利用に適したコンテンツ推薦サービスを検討している。ユーザの嗜好は例えば映画鑑賞の場合、家族と観るか恋人と観るか、元気であるか沈んでいるか等、状況や気分に応じて変化するため、ユーザ毎の差異に加えて同一ユーザ内でも状況に応じた差異を考慮してコンテンツを推薦する必要があり、モバイルユーザにとっては特にこの要望が強い。本論文では、個性と状況の両方を考慮したモバイル端末のための映画レコメンデーションシステムの被験者によるニーズ、使い勝手、推薦リストへの納得度に関する主観的な評価実験を行った結果を報告する。

2. 携帯電話によるレコメンデーションシステム

これまでに推薦を行う枠組みとして、ユーザのコンテンツに対する購買や評価履歴の類似性に基づきコンテンツを推薦する協調フィルタリング方式や[Resnic 1997]、コンテンツそのものや属性群からユーザの嗜好を反映させる特徴を抽出し、ユーザの好むコンテンツと類似した特徴を持つコンテンツを推薦するコンテンツベース(属性ベース)方式[Resnic 1997]、両者を組み合わせた方式 [Popescul 2001][小野 2005]などが検討されている。ところがこれらの推薦方式はユーザがどんな状況でも同じ結果になってしまう問題点がある。筆者らはこれらの問題を解決するためにユーザの属性、評価履歴、コンテンツの特徴情報に加えて、気分や場所等の状況も含めた変数間の複雑な依存関係をページャネット[本村 2006]により表現し、個々のユーザのそれぞれの状況に最適な映画コンテンツを推薦する方式を実装した映画レコメンデーションシステムを開発した[Ono 2007, Asoh 2006]。システムの処理の流れを図1に示す。

ユーザはモバイル端末上のインタフェース画面を通じてログインし、簡単なプロフィールや映画の嗜好、選択に関する質問に答える。その後、の画面上で、プルダウンメニューを用いて、「誰と観ますか? (一人で、家族と、恋人と)」「どこで観ますか? (映画館で、自宅で)」「今の気分(感動したい、笑いたい、泣きたい、怖がりしたい、手に汗握りたい、癒されたい、スカッとしたい)」を入力し、システムに推薦要求を送る。

推薦システムは状況 S_k にあるユーザ U_i からコンテンツ推薦要求を受け付けると、候補となる各コンテンツ C_j について、その属性群および S_k, U_i に関する属性情報を入力とし、ページャネットモデルによりユーザ U_i が状況 S_k でコンテンツ C_j を V_{ij} と評価する確率 $P(V_{ij} | U_i, C_j, S_k)$ を求め、 V_{ij} = 良いとなる確率の高い順にコンテンツを推薦する。ユーザへの応答時間を短縮するため、システムの実装にあたっては上記の確率の代わりに、あらかじめ計算してある映画の印象確率ベクトル(その映画が観た人にどんな印象を与えるか)とユーザ、状況から推論される印象確率ベクトル(どんな印象の映画を観たいと思っているか)のマッチングを計算し、近似値として利用した。

同時確率分布 $P(V_{ij}, U_i, C_j, S_k)$ を表すページャネットモデルの構築には、2006年3月に実施した、映画の嗜好に関する大規模な Web アンケート(回答者数 2153人)から構築した。まず、アンケート項目変数を、1)ユーザ属性変数、2)状況属性変数、3)コンテンツ属性変数、4)印象属性変数、5)総合評価変数の5つの変数群に分類し、図2に示す変数群間の依存関係に関する仮説に基づいてネットワークのおおまかな構造を決定した。この仮説は、ユーザのコンテンツ評価が、コンテンツに対する感じ方によって決まり、感じ方はユーザ属性、状況属性、コンテンツ属性によって決まることを表してしている。

その後、学習用データを用いて、それぞれの群の変数をクラスタリングし、代表的な変数を選択し、さらに、選択された変数間のリンクの有無と、条件付確率を推定した。最終的に得られたモデルはノード数 75、リンク数 115 であった。システムおよびページャネットモデルに関する詳細は[Ono 2007, Asoh 2006]を参照されたい。

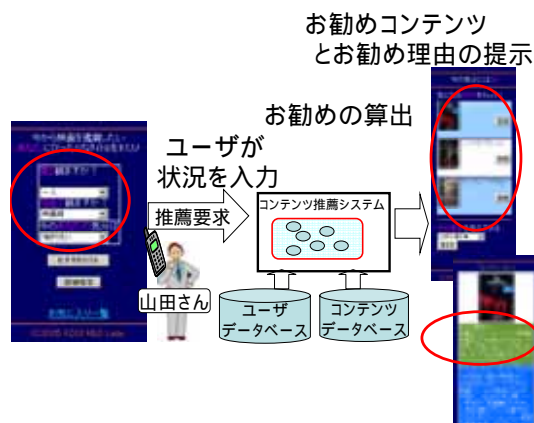


図1. 映画推薦システムの処理の流れ

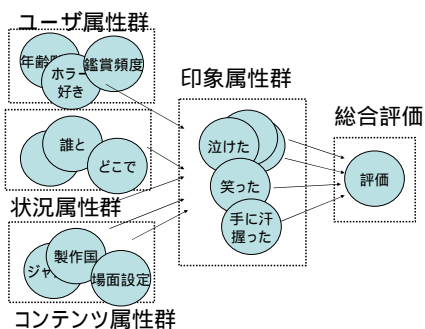


図2. 映画評価構造の仮説

3. 主観評価実験

推薦システムをどう評価すべきかは難しい問題であり、これまでも様々な考察が行われてきている。たとえば、Herlockerらは、最も普及している協調フィルタリング方式の推薦システムの評価方法について詳しく検討している[Herlocker 2004]。

推薦システムの問題は、ユーザとコンテンツと評価値の関係を学習する機械学習の問題として捉えることができるため、オフラインで、正解のある評価用データに対する推薦精度を評価することがよく行われている[Breese 1998]。この評価は有用ではあるが、Herlockerらも述べているように、推薦システムのユーザ満足度は、推薦精度以外にも、ユーザインタフェースや、推薦結果の意外性(似たものばかりが推薦されると情報量が少ない)、などといった多くの要因に左右されることが知られている。

筆者らも、これまでに、モデルの推薦精度に関する評価を行ってきたが[Ono 2007]、今回は、システム全体としてのユーザ満足度を調査するため、被験者に映画推薦システムを実際に使用してもらい、アンケートによる主観評価実験を行った。実験の目的としては、1)携帯電話による映画推薦システムへのニーズ、2)実装したシステムの使い勝手、3)推薦リストに対する満足度、の三つを設定した。以下に調査の概要を示す。

実施時期:2007年1月

調査場所:新宿および千歳烏山

被験者数:171名(10代から50代までの男女)

被験者徴集方法:ストリートキャッチ方式(アンケート会場周辺の通行人であらかじめ設定された条件を満たす人を選ぶ)

被験者選定条件:映画鑑賞頻度が比較的高く、携帯電話によるメールやウェブ閲覧の経験がある人

推薦対象:調査日時時点で公開中の映画48作品

各被験者は、説明員のガイドの下、以下の手順でシステムを操作し、アンケートに回答した。

1. 簡単なプロフィールや映画選択態度(映画選択時のこだわり度、参考にする情報など)に関する質問に回答する
2. 評価用の携帯電話を用いて映画推薦システムにログインし、ユーザ属性とその日の気分を入力して推薦結果を得て、推薦リストの評定と推薦映画の個別評定を行う。なお、今回の調査では、条件を揃えるために、ユーザの状況のうち、「誰と」は一人で、「どこで」は映画館で、に固定し、「今の気分」だけを選択するようにした。
3. 上記の操作を、気分などを変えて数回行った後に、システムへのニーズと使い勝手に関する質問に回答する。
4. 1)興行収入ランキング上位リスト、2)ユーザ属性のみを入力した場合の推薦リスト、3)ユーザ属性と気分の両方を入

力した場合の推薦リスト、4)興行収入上位20作品の中でユーザ属性と気分の両方を入力した場合の推薦リスト(人気重視)、5)興行収入21位以降でユーザ属性と気分の両方を入力した場合の推薦リスト(意外性重視)、の5パターンの推薦リスト(いずれも上位5位までを紙上で提示)を、最初に入力した気分に適した推薦と思われる順に順位付ける。

5. 5パターンの推薦リストに現れたすべての映画(重複があるため、10~20作品程度)に対して、個別評定を行う。一人あたりの回答所要時間は、携帯電話の操作に対する習熟度によって異なり、20分から40分程度であった。

4. 評価結果

4.1 基本集計

被験者の男女、年齢分布は表1に示すとおりである。10代から50代までの男女を幅広くカバーしていることがわかる。映画選択態度についてみると、「こだわりを持って映画を選ぶほうである」という問いに、64%が「そう思う」あるいは「ややそう思う」と回答している。また、「人気話題作品以外でも興味のあるものを探してみる」とについても、68%が「あてはまる」あるいは「ややあてはまる」と答えている。一方、「テレビや雑誌で話題になっているものを見る」という問いに対しては、73%が「あてはまる」あるいは「ややあてはまる」と答えている。この結果からは、多くの人が、自分は映画の選択に関してある程度のこだわりを持っていると考えているが、しかし、実際に見る映画は人気話題作品が多い、ということが伺える。

表1 被験者の性別・年齢分布(人)

	10代	20代	30代	40代	50以上
男	7	23	18	25	7
女	1	10	25	29	20

4.2 使い勝手に関する評価

使い勝手に関しては、「操作の仕方や画面の見やすさなどを総合的に考えた場合、使いやすいと思いますか?」という質問に、76%が「使いやすい」あるいは「やや使いやすい」と回答した。また、「プロフィールの入力を手間と感じましたか?」には48%が「手間と感じない」「あまり手間と感じない」と答え、気分の入力については、62%が同様に答えている。以上から、今回実装したシステムの使い勝手は良く、また、プルダウンメニューによる状況の入力も手間とは感じない人が多いことがわかる。

4.3 ニーズに関する評価

映画推薦システムへのニーズに関する質問「あなたは、今後『おすすめ映画サービス』を携帯電話サービスの一つとして利用したいですか?」に対しては、81%が利用を希望した。また、「プロフィールと気分の両方を入力して自分専用気分にあつたおすすめをして欲しい」という問いに対しては、システムを利用したいと答えた人の約60%が、「そう思う」あるいは「ややそう思う」と答えており、その比率は、「プロフィールだけを入力して自分専用におすすめして欲しい」「気分だけを入力しておすすめして欲しい」と思う人よりも多かった。以上から、携帯電話による状況に依存した映画推薦サービスへのニーズは高いと言える。

4.4 推薦リストに関する評価

推薦リストの評価に関しては、そのリストが 5 種類中最も現在の気分に適している、と答えた被験者が多い順に、1)=36%、4)=24%、3)=15%、5)=14%、2)=11%であった。ここで、興行収入ランキング上位リスト 1) を最適とした被験者が約 4 割を占めているのは、今回の実験での推薦対象が劇場公開中の映画であったためと考えられる。たとえば、2006 年度の公開作品の統計で、年間 821 本公開のうち上位 10 本(本数比 1.2%) が全興行収入の 36 % を占めていることからわかるように、劇場公開映画については、ユーザは宣伝規模が大きなメジャー作品を非常に好んで観ることがよく知られている。

一方、その 4 割を除いた残りの約 6 割の被験者は、なんらかの形で自分専用に推薦されたリストを好んでいることがわかる。そこで、1) 以外の推薦リストの優劣について詳細に検討する。まず、ユーザ属性と気分の両方を利用した推薦リスト 3) が気分を利用せずにユーザ属性のみを利用した 2) に対して約 1.8 倍のユーザから支持を得ていることは、気分を利用した推薦の有効性を示していると言える。

また、気分に加えて人気も考慮した推薦リスト 4) が 3) の 1.6 倍のユーザから支持されていることから、公開中の映画を対象とする場合には、人気も合わせて考慮した推薦を行うことによりユーザ全体の平均的な満足度は高くなるのがわかる。

さらに、映画の選択する際に「人気ランキングを見て決める」という質問に対して「あまりあてはまらない」「あてはまらない」と答えた被験者(59 名)のみを抽出して集計したところ、5)=24%、1)=22%、4)=20%、3)=20%、2)=14%となり、興行収入 21 位以降でユーザ属性と気分の両方を利用したリスト 5) を最適とする被験者数が、1) を最適とする被験者数を上回り最大となった。このことは、人気ランキングを見て決めるユーザには人気重視の推薦リストを提示し、そうでないユーザには意外性重視の推薦リストを提示する、というように、映画選択に対する態度属性によって推薦リストを変えることで、より効果的な推薦ができる可能性を示している。

4.5 個別映画評価による推薦リスト評価

今回のアンケートでは、推薦リストの順位づけに加えて、リストに含まれる個々の映画について、それが現在の気分に適しているか否かを評定してもらっている。そこで、その評価値(5 段階)を用いて、推薦リストの評価を算出してみた。

算出方法はいろいろ考えられるが、ここでは、まず、5 段階の映画の評価値を、中央の 3 を 1 とし、4 を $\frac{1}{2}$ 、5 を $\frac{1}{4}$ の 2 乗、2 を $\frac{1}{2}$ 、1 を $\frac{1}{4}$ の 2 乗、というように実数値に変換した。さらに、推薦リスト内での順位によって、1 位は $\frac{1}{2}$ 、2 位は $\frac{1}{4}$ 、3 位は $\frac{1}{8}$ の 2 乗、... というように重みをつけて加算した。ここで、 $\frac{1}{2}$ は 1 より大きい定数、 $\frac{1}{4}$ は 1 より小さい定数である。この算出方法で、 $\frac{1}{2}$ 、 $\frac{1}{4}$ とした場合の 5 つの推薦リストの評価値の平均と標準偏差を表 2 に示す。

表 2 推薦リストの評価値の平均と標準偏差

推薦リスト	1)	2)	3)	4)	5)
平均	4.08	3.61	3.77	3.88	3.67
標準偏差	1.09	1.05	1.04	1.10	1.03

評価の良い順に 1),4),3),5),2) となっており、推薦リストを直接順位づけた場合と同様、公開映画に対する興行収入ランキングによる推薦 1) および、興行収入を考慮した推薦 4) の優位性が確認された。しかし、ユーザ間の評価値のばらつきを示す標準

偏差と比較すると、評価の平均値の差はそれほど大きくはない。また、この結果は、 $\frac{1}{2}$ や $\frac{1}{4}$ の値にあまり依存しなかった。

さらに、上記の評価値によって被験者ごとに、推薦リストの順位を求めたところ、それぞれのリストを 1 位とした被験者の割合は、多い順に、1)=39%、5)=20%、4)=18%、3)=15%、2)=8%、となり、リスト全体を直接的に順位づけした場合よりも 5) の評価が高い結果となったが、おおよその傾向は変わらなかった。このことは、今回の被験者が比較的一貫した評価を行っており、その評価結果の信頼性が高いことを示していると思われる。

最後に、個別映画評価による各被験者ごとの推薦リストの順位づけと、被験者による直接的な順位づけに対して、スピアマンの順位相関係数を計算してみたところ、全ユーザの相関係数の平均値は 0.46 であった。スピアマンの順位相関は、全く同じ順位の場合に 1、完全に逆の順位の場合に -1 を取るため、上の値は、二つの順位が同じ傾向を有していることを示している。

5. まとめ

本稿では、筆者らが提案している、個性と状況の両方を考慮したモバイル端末のための映画推薦システムの概要について述べ、システムの主観評価実験の結果について報告した。171 人の男女に実際にシステムを使ってもらい、推薦システムへのニーズ、使い勝手、推薦リストの満足度、等についてアンケートを行った結果、モバイル端末による状況に依存した推薦システムへのニーズが高いこと、および、実装したシステムの使い勝手がよいこと、が明らかになった。さらに、推薦リストの満足度に関しても、状況を考慮することによって、ユーザ満足度が向上することなどを示すことができた。

今後は、調査結果をより詳細に分析し、その結果に基づいてシステムを改良し、よりユーザの満足度が高いシステムを実現してゆきたい。

謝辞 日頃ご指導頂く KDDI 研究所秋葉重幸代表取締役所長に深く感謝いたします。

参考文献

- [Resnic 1997] P. Resnic and H.R. Varian, "Recommender systems," Communications of the ACM, 40, No.3, 56-58, 1997
- [Popescul 2001] A. Popescul, L. H. Ungar, D. M. Pennock, and S. Lawrence, "Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments," Proc. 17th Conference on Uncertainty in Artificial Intelligence, 2001
- [本村 2006] 本村, 岩崎, "ページアンネットワーク技術", 東京電機大学出版局, 2006.
- [Ono 2007] C. Ono, M. Kurokawa, Y. Motomura, H. Asoh, "A Context-Aware Movie Preference Model Using a Bayesian Network for Recommendation and Promotion", Proc. Of 11th Conf. of User Modeling (UM2007), 2007 (To be published)
- [Asoh 2006] H. Asoh, C. Ono, and Y. Motomura, "A movie recommendation method considering both user's personality and situation," Proc. ECAI2006 Workshop on Recommender Systems, 49-53, 2006.
- [Herlocker 2004] J. Herlocker, et al. "Evaluating collaborative filtering recommendation systems," ACM Transactions on Information Systems, Vol.22, pp.5-53, 2004.
- [Breese 1998] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering", Proc. of 14th Conf. on Uncertainty in Artificial Intelligence, pp.43-52, 2004.