

嗜好に基づくレコメンドソリューションの実現

Retaining Consumer Attention Using Recommendation

堀井 遼太*¹ 富士谷 康*² 堀 充貴*³ 乾 隆夫*¹ 山村 博幸*¹ 芥子 育雄*¹

Ryota Horii Kou Fujitani Mitsuyoshi Hori Takao Inui Hiroyuki Yamamura Ikuo Keshi

近年、電子書籍ストアのようなオンラインストアで扱うアイテム数は急増し、新着でも直ぐロングテール化する。本論文では、ユーザ嗜好に基づく推薦において、ロングテールの課題を解決するソリューションを提案する。本提案のソリューションは次の4つの特長を備える。(1) ストアの購入履歴を元に推薦効果を最大化する推薦システムの最適化、(2) 多様な推薦を実現する嗜好の蓄積、(3) 意外性のある推薦を実現する嗜好のフィルタリング、(4) 異なるドメインから嗜好を獲得しアイテムを推薦するクロスドメインレコメンド

With the large number of items offered by online stores, even the recently launched ones can quickly lose consumer attention and fall into the long-tail.

This paper describes solutions for this problem based on recommendation, offering the following features: (1) Optimization of the recommendation system to maximize the conversion rate; (2) Preference accumulation to offer a wider variety of recommendations; (3) Preference filtering to offer “serendipity” experiences; (4) Cross-domain recommendation based on user preferences in the different domains/media.

1. はじめに

シャープは2010年12月より総合電子書籍ストア「GALAPAGOS STORE」を運営している。ユーザのニーズに応えるため、運営開始後も様々なサービスを追加してきた。その1つにレコメンドサービスがある。

レコメンドサービスは、ユーザに対しておすすめの本を推薦するサービスである。サービス開始後も、ユーザがより興味をそそられて飽きない、ひいては、ストアの売上に貢献する本の推薦を目標に課題を抽出し、様々なレコメンドソリューションを開発し解決してきた。

また、家電ビッグデータの1つであるテレビの視聴情報を用い、ユーザに本を推薦するクロスドメインレコメンドの可能性を検証した。

本稿では3つの主要なレコメンドソリューションとクロスドメインレコメンドの検証結果について述べる。

2. レコメンド手法

レコメンドサービスを実現する手法は表1に示すように分類される²⁾。本章では各手法の概要と特徴について説明する。

表1 レコメンド手法の分類

Table 1 Classification of recommendation algorithm.

大分類	中分類	小分類
協調フィルタリング	メモリベース法	類似ユーザ型
		類似アイテム型
	モデルベース法	
コンテンツベースフィルタリング		

2.1 協調フィルタリングと コンテンツベースフィルタリング

協調フィルタリングは、ユーザとアイテムの関係を嗜好と捉えて、嗜好が類似するユーザの好むアイテムを推薦する。コンテンツベースフィルタリングは、説明文やジャンルなどのメタデータから構築したアイテムの特徴とユーザの特徴を比較し、特徴が類似するアイテムを推薦する。

協調フィルタリングはユーザ自身と他のユーザの嗜好をもとに推薦する。このため、他のユーザからのクチコミ効果が期待でき、意外性がある推薦をできる傾向がある。

また、協調フィルタリングは、アイテムのメタデータや特徴を必要としないため、実現が容易である。しかしながら、新しい、不人気などの理由で、他のユーザから嗜好されていないアイテムを推薦できない。一方、コンテンツベースフィルタリングは、メタデータからアイテムの特徴を構築する手間がかかるが、他のユーザから嗜好されていないアイテムであっても、特徴が類似していれば推薦できる。

*¹ 研究開発本部 クラウド技術開発センター

*² 立命館大学 情報理工学部 情報システム学科

*³ 通信システム事業本部 マーケティングセンター

2.2 メモリベース法とモデルベース法

メモリベース法は、推薦前は嗜好データを蓄積するのみで、推薦時に嗜好データから類似ユーザを探索して推薦する。モデルベース法は、推薦前に類似ユーザを探索したモデルを構築しておき、推薦時はモデルから類似ユーザを取得して推薦する。

メモリベース法は、推薦の度に類似ユーザを探索するため、推薦に時間がかかる。一方、嗜好データが追加や削除などで変更された場合、モデルベース法はモデルを再構築するまで最新の嗜好をもとに推薦できないが、メモリベース法は同様の問題は生じない。

2.3 類似ユーザ型と類似アイテム型

類似ユーザ型は、ユーザと他のユーザを比較し、類似度の高いユーザが嗜好するアイテムを推薦する。類似アイテム型は、ユーザが嗜好するアイテムと他のアイテムを比較し、類似度の高いアイテムを推薦する。

一般に、類似ユーザ型はユーザに意外性があるアイテムを推薦できる傾向にある。一方、類似アイテム型はユーザの嗜好が十分に蓄積されていない場合でもある程度の確なアイテムを推薦できる。

3. レコメンドサービス

3.1 概要

図1に GALAPAGOS STORE 向けレコメンドサービスのシステム構成を示す。レコメンド手法にはメモリベース協調フィルタリングを採用している。これは、ユーザに意外性がある書籍を推薦し、書籍購入後すぐに新たな書籍を推薦するためである。メモリベース協調フィルタリングは、前述の通り、推薦に時間がかかる問題があるが、蓄積した嗜好データをメモリ上に配置することで高速化して対応している。

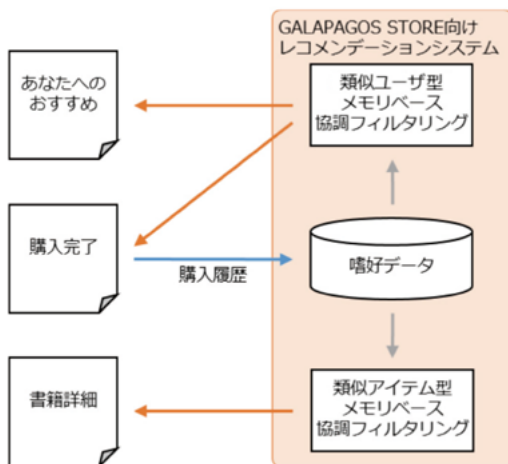


図1 システム構成
Fig. 1 System configuration.

推薦前は書籍の購入履歴を嗜好データとして蓄積する。推薦時は、あなたへのおすすめページと購入完了ページに類似ユーザ型、書籍詳細ページに類似アイテム型の推薦を提示する。

3.2 課題

図2は、GALAPAGOS STOREにおける書籍売上のパレート図を示している。一般的な店舗と同様に一部の書籍に売上が偏っている。推薦に期待される効果は、ロングテールに含まれるアイテムを適切なユーザに提示することで、売上に貢献することである。

そこで、類似ユーザ型メモリベース協調フィルタリングを用いた推薦における課題は以下が考えられる。

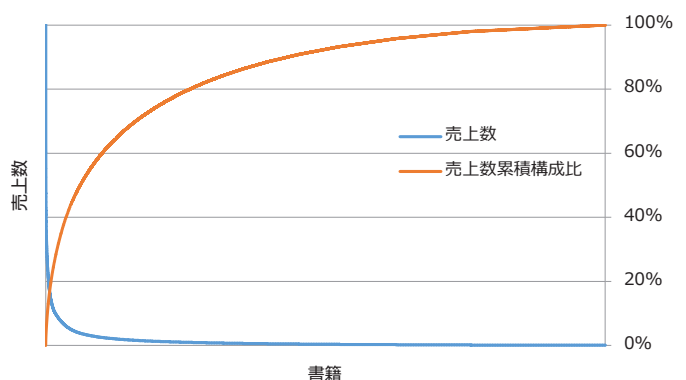


図2 書籍売上のパレート図
Fig. 2 Pareto chart of book sales.

(1) 推薦精度

類似ユーザ型は複数の類似ユーザを抽出して推薦する。しかしながら、どれだけの類似ユーザを抽出すれば、精度が高い推薦ができるか分からない。

(2) 嗜好蓄積

GALAPAGOS STOREでは、購入履歴を嗜好データとして蓄積している。しかしながら、サービスに登録していても書籍を購入していないユーザがいる。これらのユーザには嗜好が分からないため推薦できない。また、購入している場合であっても、短期間に何冊も購入するユーザは少ない。したがって、嗜好があまり変化せず、推薦もほとんど変化しないため、飽きられてしまう課題もある。

(3) 意外性がない推薦

協調フィルタリングは、類似するユーザが好む書籍を推薦する。このため、多く購入される人気書籍が推薦されやすい。しかしながら、ユーザはすでに宣伝や口コミなどで人気書籍を認知している可能性が高い。また、不人気な書籍の購入履歴はユーザの嗜好を強く反映するが、人気書籍の購入履歴に埋もれてしまう。これらのため、意外性がある推薦をできない。

4. レコメンドソリューション

本章では、前章で取り上げた類似ユーザ型メモリベース協調フィルタリングの課題を改善するソリューションについて述べる。

4.1 推薦精度の改善ソリューション

推薦精度を改善するため、ユーザの過去の購入履歴を利用し、精度が高い推薦ができる類似ユーザ数を実験的に算出するソリューションを適用している。具体的には、まず、平均的な購入頻度のユーザを抽出する。つぎに、抽出したユーザの購入履歴を利用して、推薦に利用する類似ユーザ数別に、複数回の交差検証を行う。評価尺度には、適合率、再現率、F値を用いる。交差検証の結果、精度が最も高い類似ユーザ数を推薦時に利用する。

図3は、GALAPAGOS STOREの購入履歴に対して、前述のソリューションを適用し得られた類似ユーザ数別推薦精度を示している。類似ユーザ数が少ない場合、ユーザと類似ユーザの購入履歴の差が小さく、推薦できる数が少ないため、適合率が高く、再現率が低くなる。類似ユーザ数を増やすとともに精度が向上し、類似ユーザ数63人でF値が最大となる。以後、徐々に精度は低下していく。これは、ユーザと類似ユーザの嗜好が乖離していくためだと考えられる。以上から、GALAPAGOS STOREでは、類似ユーザ数を63人に設定して推薦をすれば、精度が高く、コンバージョン率を最大化する推薦ができることが分かる。

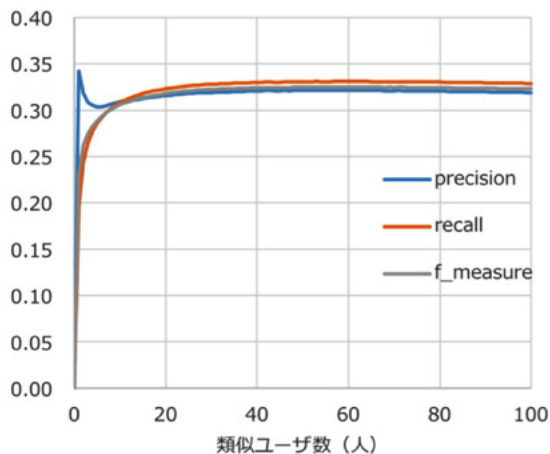


図3 類似ユーザ数別推薦精度

Fig. 3 Recommendation accuracy according to the number of similar users.

4.2 嗜好蓄積の改善ソリューション

嗜好蓄積の課題を解決するため、書籍詳細の閲覧履歴も嗜好データとして蓄積するソリューションを適用している。書籍詳細の閲覧履歴は、ユーザが興味を持って閲覧したと考えられるため、嗜好として捉えられる。また

購入と比較して、ユーザは気軽に利用できるため、嗜好データを蓄積しやすく、変化しやすい。

GALAPAGOS STOREでは、購入履歴に加えて閲覧履歴による推薦を導入したことで、登録ユーザのうち約90%に対して推薦できるようになった。残り約10%のユーザは購入履歴、閲覧履歴がともないため、サービスの登録のみ行った非アクティブなユーザと考えられる。また、セッション情報などを利用して閲覧履歴を蓄積すれば、サービスに登録していないユーザに対しても推薦できるようになる。

また、購入履歴と閲覧履歴による推薦が互いに重複する割合は約6%しかないことが分かっている。これは、購入と閲覧により捉えられる嗜好が互いに異なることを示唆している。したがって、両履歴による推薦を交互に、もしくは、混ぜあわせて推薦することで、ユーザが飽きない推薦をできる。

4.3 意外性がない推薦の改善ソリューション

意外性がある推薦をするため、推薦時に嗜好データをフィルタリングするソリューションを適用している。具体的には、ユーザ同士を比較する際と推薦書籍を抽出する際にフィルタリングしている。

図4は、書籍A, B, Eを人気書籍と仮定し、これらの書籍を、フィルタリングしない場合とフィルタリングした場合についての推薦の流れを示している。類似ユーザとして、フィルタリングしない場合はユーザA, Bが抽出されるが、フィルタリングした場合にはユーザBのみ抽出される。このように、人気書籍以外の購入履歴で比較することにより、ユーザの強い嗜好を用いて類似ユーザを抽出でき、意外性がある推薦をできる。また、類似ユー

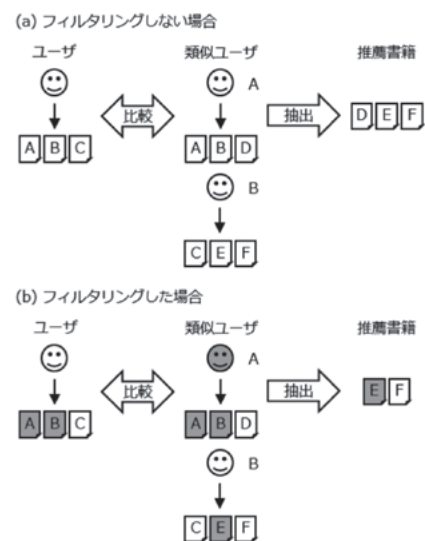


図4 推薦のながれ

Fig. 4 Flow of recommendation.

ザから推薦書籍を抽出する際に、人気書籍をフィルタリングすることでも、意外性がある書籍の推薦につながる。

また、特定期間に購入された書籍をフィルタリングすることで、トレンドを反映した推薦をすることもできる。さらに、推薦の度にフィルタリングする条件を変更することで、ユーザが飽きない推薦をすることも考えられる。

5. クロスドメインレコメンド

協調フィルタリングでは、書籍との関係がないユーザに対して推薦できない。つまり、推薦するにはアイテムと同じドメイン（領域）において、ユーザが嗜好を持つ必要がある。一方、アイテムと異なるドメインからユーザの特徴を獲得することで推薦する、クロスドメインレコメンドに関する研究が行われている¹⁾。本章では、テレビの番組情報と視聴履歴からユーザの特徴を獲得し、書籍を推薦するクロスドメインレコメンドの検証結果を述べる。

図5にクロスドメインレコメンドシステムの概要を示す。システムは大きく「書籍プロフィールの構築」「ユーザプロフィールの構築」「マッチング（比較）」の3つに分けられる。書籍プロフィールの構築では、TF-IDF手法を用いて、書籍情報から特徴ベクトルを構築する。ユーザプロフィールの構築では、番組情報と視聴履歴から、書籍と同様に特徴ベクトルを構築する。株式会社ワイヤーアクションが提供する番組情報を用い、番組タイトル、トピック説明文、トピック名、キーワード、出演者を利用した。なお、ユーザプロフィールは、以下の3種類を構築する。番組別プロフィールは、視聴番組情報ごとに、視聴番組数分の特徴ベクトルを構築する。ジャンル別プロフィールは、ジャンルごとにまとめた視聴番組情報ごとに、ジャンル数分の特徴ベクトルを構築する。要約プロフィールは、視聴番組情報をまとめて、一つの特徴ベクトルを構築する。マッチングでは、書籍プロフィールとユーザプロフィールとの距離を算出し、距離が近い順に書籍を提示する。

検証は、社内の20代から50代の被験者10名が1週間

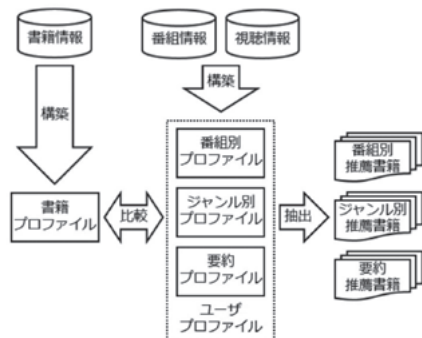


図5 クロスドメインレコメンドシステムの概要
Fig. 5 Cross-domain recommender system.

分の視聴履歴を手入力して行った。ユーザプロフィールごとに上位5件の書籍を提示し、興味があったか、視聴番組と関連があったかの2点について、5段階評価を得た。評価結果を図6に示す。評価尺度には、評価値の平均を算出した評価平均、評価値が4以上を正解として5冊中1冊以上正解があった場合の正解率を算出したTop5精度を用いた。評価では、番組別プロフィールで抽出された書籍を距離でソートし上位5冊を抽出した番組Top5プロフィールを加えた³⁾。番組Top5プロフィールは精度が高く、距離によるソートが有効であるといえる。

ユーザプロフィールの構築方法によって差はあるが、クロスドメインレコメンドを用いて、ある程度の精度で書籍を推薦できることを確認した。

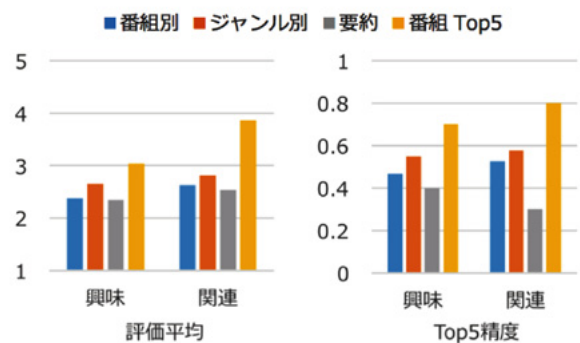


図6 評価結果
Fig. 6 Evaluation results.

6. 最後に

これまで述べてきたように、書籍を推薦する課題を解決するレコメンドソリューションを提供することにより、GALAPAGOS STOREの発展を支えてきた。今後も継続してソリューションを提供してさらなる発展を支えるとともに、ソリューションを含むレコメンドサービスの応用展開を検討していく。

クロスドメインレコメンドの検証は、シャープのインターンシップの一環として実施した。クロスドメインレコメンドの研究にあたり、ご討論いただいた立命館大学西尾信彦教授に深く感謝いたします。

参考文献

- Ignacio Fernández-Tobías, Iván Cantador, Marius Kaminskis, Francesco Ricci. "Cross-domain recommender systems: A survey of the state of the art." In Proceedings of the 2nd Spanish Conference on Information Retrieval, CERI, 2012.
- 神島敏弘. 推薦システムのアルゴリズム (1). 人工知能学会誌, Vol. 22, No. 6, pp. 826-837, 2007.
- 富士谷康, 堀井遼太, 芥子育雄, 西尾信彦. テレビ番組視聴履歴を用いた電子書籍推薦, ユビキタス・ウェアラブルワークショップ2014論文集, pp. 3, 2014.