

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 6月 8日現在

機関番号：82626

研究種目：基盤研究（C）

研究期間：2009～2011

課題番号：21500154

研究課題名（和文）推薦システムにおけるスタートアップ問題の転移学習による解消

研究課題名（英文）Solving a start-up problem of recommender systems by employing transfer learning

研究代表者

神島 敏弘（KAMISHIMA TOSHIHIRO）

独立行政法人産業技術総合研究所・ヒューマンライフテクノロジー研究部門・主任研究員

研究者番号：50356820

研究成果の概要（和文）：本研究では、利用者が好むであろうものを見つけて提示する推薦システムの利用開始時に予測精度が悪いというスタートアップ問題を対象とした。この問題に対し、他の利用者のデータを利用することで、この問題の改善に取り組んだ。そのために TrBagg と呼ぶアルゴリズムを開発し、ソーシャルブックマークのタグ推薦のテストデータ上で、検証を行い、予測精度が確かに向上した。その他、推薦システムの他の面の改良も行った。

研究成果の概要（英文）： In this project, we planned to address a start-up problem, which means the poor prediction accuracy of recommendation systems for new users or new items. For this purpose, we developed a TrBagg algorithm, which utilizes information of other users' data. We tested this algorithm on a tag recommendation task and succeeded to improve the prediction accuracy. We also tried to settle other problems related to recommender systems.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,200,000	360,000	1,560,000
2010年度	1,300,000	390,000	1,690,000
2011年度	900,000	270,000	1,170,000
総計	3,400,000	1,020,000	4,420,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：知識発見とデータマイニング、推薦システム、転移学習

## 1. 研究開始当初の背景

推薦システム（recommender system）は、利用者の好むであろうアイテムや情報を予測し、それを利用者に提示するシステムである。これは、必要とするアイテムや情報があっても、膨大な他のものに埋没して目的のアイテム・情報を入手できない情報過多（information overload）に対処するための手段として開発されてきた。1990年代から研究が進展し、現在では、電子商取引サイトなどで幅広く利用されている。

このように大きな成功を収めている一方で、嗜好や意図に合わないものを利用者に推薦してしまうという問題は、依然として数多く生じている。その原因は幾つもあり、それぞれについて対処が必要とされている。

## 2. 研究の目的

推薦システムの抱えるいくつかの問題のうち、以下のものの解決を試みる。

## (1) スタートアップ問題

推薦システムでは、利用者のいろいろなアイテムへの好みを提示した嗜好データを収集し、そのデータから利用者の好みのパターンを求め、そのパターンに合致したアイテムを推薦する。しかし、利用者が使い始めの時期には嗜好データがまだ十分に集積されていないため、利用者の好みのパターンを正確に把握できず、適切な推薦ができない。また、新規利用者だけでなく新規アイテムについても、利用者による評価がそれらのアイテムに対してあまりなされていないため、利用者の嗜好パターンに合致するかどうかを判断できない問題がある。これらがスタートアップ問題である。利用開始時のまだ状況が把握できていない時期や、評価を特に知りたい新製品に対しての推薦システムがより必要とされる現状において深刻な問題であるため、特に対処が求められる。

#### (2) 嗜好の収集方法の問題

推薦システムを実現するには、利用者がアイテムへの嗜好の度合いを示したデータを収集することが必要になる。そのためによく利用されているのが採点法と格付け法である。前者は、星の数などでアイテムへの嗜好を数値化して示す方法であり、後者は、優良可などの段階のあるカテゴリに分類させる方法である。これらの方法は、ある程度の成功を収めているが、自身の嗜好の強さを明示的かつ無矛盾に数値化することは現実には難しいし、限られた段階のカテゴリに分類するとどうしてもどちらにランク付けするか判断が難しい場合もある。このような問題に対処するため、他の計測方法も検討されるべきである。

#### (3) 推薦を超える利用者への働きかけ

推薦システムは前述のように大きな成功を収めたが、基本的には顧客にアイテムを提示するだけで、実店舗での店員のような高度な提案はできない。例えば、服飾店でコーディネートについて提案したり、自動車販売店で車のオプションパーツについて相談したりといったことは現状ではできない。そこで、アイテムの単純な提示だけに留まらず、より高度な『おもてなし』ともいえる対応ができる枠組みを考える。

### 3. 研究の方法

#### (1) スタートアップ問題

スタートアップ問題に対処するため、帰納転移や転移学習と呼ばれる技術を導入する。これは、これから解こうとする目標タスクに専用のデータが少ないので、類似した問題のデータを転用して目標タスク予測精度を向上させるものである。推薦システムには、推薦対象の特徴を参照し、それに基づいて推薦

の可否を決める内容ベースフィルタリングと、他の利用者の嗜好パターンに基づき、いわば『ロコミ』を模して推薦を行う協調フィルタリングとがある。ここでは、内容ベースにおいて、他人のデータを類似したデータと考へて、転移学習を適用することでスタートアップ問題に対処する。

#### (2) 嗜好の収集方法の問題

従来の採点法や格付け法に対し、順位法を用いたアイテム評価の導入を行う。順位法とは、いくつか提示されたアイテムを自分がより好むものからそうでないものに順に利用者が整理することで、自身の嗜好を示す方法である。自身の嗜好の段階を任意に細かくできたり、無矛盾に数値化する困難を回避できるといった利点がある。一方で、利用者から得られる情報は相対的な嗜好の優劣のみで情報の量については減少し、その解析技術は限られているという問題がある。そこで、順位法を用いた分析技術を開発する。

#### (3) 推薦を超える利用者への働きかけ

推薦以上の利用者に対する積極的な対応の一つとして『価格個人化』を選んだ。これは、全ての利用者に同一の価格で商品を提供するのではなく、利用者の状況に応じて価格を変更するものである。価格個人化を対象に選んだ理由は、利用者の応答が購入したかどうかという明確なものであり、計算機上で扱う最初の推薦以外の行動として、よい研究対象と考えた。また、ある行動をするかどうかというフィードバックは他の行動でもよく見られるものなので、基本的には価格個人化を対象としたこの枠組みが、他の行動を分析するための基礎ともなると考えたからである。

### 4. 研究成果

#### (1) スタートアップ問題

スタートアップ問題に対処するための転移学習手法として TrBagg 法を開発した。推薦タスクとしては、ソーシャルタグ付けでのタグ推薦タスクを選び、その有効性を示した。ソーシャルタグ付けとは、Web ページなどのアイテムに対し利用者がそのアイテムを特徴付けるキーワードであるタグを付加できるようにしたものである。例えば、「東京スカイツリーのホームページ」であれば、「東京スカイツリー」の他、「東京」や「電波塔」などのタグを付加することが出来る。利用者は自身が付けたタグによって、自分が整理したアイテムを整理し、あとで検索することができる。さらに、ソーシャルタグ付けのソーシャルたる所以は、自分だけでなく他の利用者が付けたタグとアイテムも参照することができる点である。この機能により、「東京」

と「電波塔」のタグで検索すると、他の利用者が登録していた「東京タワーのホームページ」を見つけたりすることができる。

ここで、タグ推薦とは、ある Web ページが与えられたとき、そのページにその利用者が付けるであろうタグを予測し、それを提示するものである。ここで重要な点は、利用者によって適切なタグというものは変わる点である。例えば、「富士山に関するページ」の場合、日本の居住者であれば「山梨」や「静岡」といったタグが適切だろうが、海外の居住者にとっては「日本」というタグの方が適切となるだろう。よって、推薦するタグは適切に個人化されている必要がある。

このタグ推薦の問題をクラス分類の問題として定式化した。すなわち、あるページにあるタグが適切か、適切でないかの二つのクラスに分類する。特徴量には、そのページに付加した目的以外のタグを用いた。

ここで、利用者自身が付けた他のタグは、その利用者の嗜好を正確に反映したものとなる。だが、システムを利用し始めたばかりのときは付加したタグの情報が少なく、正確な予測ができないというスタートアップ問題を生じてしまう。

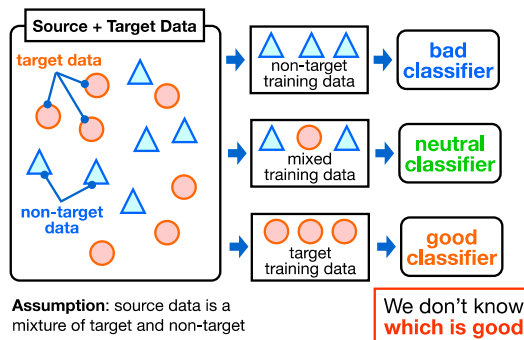
そこで、利用者が付加した他のタグから、目標タグが適切かどうかを識別する問題を目標タスクとする。それに加え、他の利用者が付加したタグから、目標タグの適切さを識別するタスクを元タスクとする。この元タスクは、適切な個人化を考慮していないため、目標タスクとは確かに異なるものである。しかし、「富士山の例」でいえば、対象とする利用者と居住地域が同様で、同様のタグが適切となる場合もあり、全く役に立たないわけではない。

このように、目標タスクに対して高い整合性を持ったデータと、完全に整合するわけではないがある程度の有用性が予測されることが混在する場合を想定し、TrBagg 法を設計した。

TrBagg 法について詳細を述べる。TrBagg 法は、訓練データからサンプリングして分類器を学習する手順を何度も繰り返し、その分類器で多数決をとって最終結果を決める Bagging と呼ばれる方法を転移学習に応用したものである。TrBagg 法は、Bagging と同様にサンプリングを繰り返し、分類器を学習する訓練段階と、その後のフィルタリング段階とに分かれる。

最初の訓練段階の概要は、次の図の通りである。最初に元データと目標データが混在したデータからサンプリングし、分類器を学習する手順を何度も繰り返す。

この段階を終えると、目標タスクにとって適切な事例から学習した良い分類器と、あまり適切でない事例から学習した悪い分類器



が混在した状況で、幾つもの分類器が作られる。

ここで問題となるのは、どの分類器がよい分類器なのかが分からない点である。そこで、次のフィルタリング段階が必要となる。この段階では、少量の目標データを分類し、そのデータをうまく分類できた分類器を良い分類器、そうでないものを悪い分類器と考え、良いものだけを残す。あとは、通常の Bagging 法と同様に、よい分類器だけで多数決をとり、タグが対象とする利用者にとって適切かどうかを識別する。

実験では、ソーシャルタグ付けサイト「delicious」と「はてなブックマーク」から取得したデータを用いた。それぞれ 20 種類の一般的なタグを選び実験を行った。対象とする利用者が多くのタグを付けている状況では、「delicious」では提案手法が良かったものが 3 個のタグ、目標データだけで bagging 法を適用した場合が 0 個のタグであり、「はてなブックマーク」では 2 対 0 と提案手法が若干よかっただけであった。しかし、スタートアップ問題が深刻になる利用者が付けたタグが少なくなるにつれ提案手法の効果は増し、最も少ない場合の実験では「delicious」データが 11 対 0、「はてなブックマーク」で 10 対 2 と、提案手法の高い有効性が示された。

この結果は、2009 年度の International Conference Data Mining にて採録された。

## (2) 嗜好の収集方法の問題

順位法を用いた推薦は、すでに継続的に研究を行ってきた。協調フィルタリングの実現方法はメモリベース法とモデルベース法とに大きく二つに分かれるが、今まではメモリベース法にて提案手法の有効性を確認してきた。今回は、もう一方のモデルベース法でも有効かどうかを検証した。

今までの提案手法は、順位法で並べたアイテム  $x$  の順位を  $i$  とし、同時に並べたアイテムの数を  $k$  とすると、アイテム  $x$  にスコア  $(i + 1) / (k + 1)$  を与えて、通常の採点法で集めたデータを処理するアルゴリズムと同様のアルゴリズムを適用するという

ものである。これは非常に単純には見えないが、ある仮定のもと、並べていないものも含めた全てのアイテムの中で、アイテム  $x$  がどれくらい好みの順位であるかという期待値を表すという理論的背景をも持ち合わせている。

このように変換したスコアを用いて、モデルベースの中で SVD++ 法と pLSI 法と呼ばれる二つの方法で、推薦の予測精度が向上したかどうかを測った。予測精度を、従来の採点法で集めたデータによる結果と比較したところ、同時に並べたアイテムの数が3個以上であれば、順位法による予測の方が高い予測精度を示した。これにより、提案手法がモデルベース型の協調フィルタリングにおいても有効であることが確認された。

この結果は International Conference on Recommender Systems 2010 にて採録された。

### (3) 推薦を超える利用者への働きかけ

推薦より踏み込んだ利用者への働きかけとして価格個人化を行った。これは、定価では購入しないが、割引を提示すると購入する顧客にのみ割引価格を提示することで、定価で購入する顧客から得られる従来の収入に加え、追加で利益を得られるようにするものである。

この価格個人化が推薦システムの運用に果たす重要性について少し述べておく。推薦システムが店舗に導入される理由は、顧客が好むであろうものを推薦されることにより、購入数が増加し、追加の売上が得られるというものである。しかし、推薦されたからといって顧客は簡単には購入しないため、システムの運用コストをまかなうほど十分な追加売上が得られるとは限らない。そのため、システムを運用できなくなってしまうたり、必要以上に高価なものを推薦するといった不誠実な推薦をしてしまうことが考えられる。

そこで、価格個人化を導入して、十分な追加収入が確保できるようにし、不誠実な推薦をするインセンティブを取り除くようにする。また、顧客も場合によっては割引価格で購入できるメリットも生じる。

価格個人化を実現するためには転売について考慮する必要がある。最終消費者としてではなく、転売目的の顧客に割引価格を提示すると転売目的で購入される場合がありうる。そのため、各顧客が、定価で購入する定価顧客、割引があった場合のみ購入する割引顧客、最終消費者として購入する意思のない不買顧客の三つのいずれかであるかを予測し適切な価格を提示する必要がある。

ここで、このタスクを実現するには三つの技術課題がある。一つは、顧客の行動から顧客のタイプを完全には予測できない観測の曖昧性である。二つ目は、顧客データの収集

のため、現在の予測結果では最適といえない価格を提示する必要が生じる活用-探索のトレードオフである。そして最後は、不買顧客は他のタイプの顧客に対して圧倒的に多いというクラス不均衡問題である。

最初の、観測の曖昧性の問題について述べる。割引価格を提示すると割引顧客はもちろん、定価より低い価格を提示された定価顧客も購入する。よって、このときには顧客が定価顧客か割引顧客であったかのどちらかの区別がつかない。そこで、定価を提示したときの反応で、定価顧客とそれ以外を分類する段階と、割引価格を提示した場合での反応で残りの不買顧客と割引顧客を分類する多段階分類のアイデアを導入した。

二つ目の、活用-探索トレードオフの問題については、多腕バンディットと呼ばれるトレードオフの調整方法を採用した。今回は、その中でも単純な方法である  $\epsilon$ -Greedy と呼ばれる方法を採用した。

最後の、クラス不均衡問題に対しては、再現率と精度のバランスを見ながら、分類の決定しきい値を調整する対策をとった。

推薦システムとの関連は、推薦システムのモデルを学習し、そのモデルのパラメータを、顧客のタイプの分類の特徴量として用いた。これは、これらのモデルパラメータには顧客のアイテムへの嗜好の情報が凝縮されているためである。

こうして作成した実験システムを、推薦システム用のベンチマークデータに、人工的に付加した購買データを付加し、半人工のデータを生成し実験を行った。実験の結果、常に定価もしくは、割引価格のみを提示する戦略に対し、提案手法で追加利益が得られるとの結果を得た。この成果は、RecSys2011 併設ワークショップ HetRec2011 に採録された。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計4件)

① T. Kamishima and S. Akaho “Personalized Pricing Recommender System - Multi-Stage Epsilon-Greedy Approach” Proc. of the HetRec2011 Workshop, pp. 57-64 (2011)  
DOI: 10.1145/2039320.2039329

② T. Kamishima and S. Akaho “Nantonac Collaborative Filtering: A Model-Based Approach” Proc. of the RecSys2010, 査読有, pp. 273-276 (2010)  
DOI: 10.1145/1864708.1864765

③ 神嶌 敏弘 “解説『転移学習』人工知能学会誌, 査読無, vol.25, no.4, pp.572-580 (2010)  
DOI: なし

④ T. Kamishima and M. Hamasaki and S. Akaho “TrBagg: A Simple Transfer Learning Method and Its Application to Personalization in Collaborative Tagging” Proc. of the ICDM2009, 査読有, pp. 219-228 (2009)  
DOI: 10.1109/ICDM.2009.9

[学会発表] (計 12 件)

① 神嶌 敏弘, 赤穂 昭太郎 “価格個人化推薦システム” 電子情報通信学会WI2研究会, 大阪 (2011. 11. 8)

② 神嶌 敏弘, 赤穂 昭太郎 “公正配慮型学習 - 正則化によるアプローチ” 人工知能学会全国大会, 盛岡 (2011. 6. 1)

③ 神嶌 敏弘, 赤穂 昭太郎, 佐久間 淳 “カスタム価格設定推薦システム - 簡単な実装と予備実験 -” 電子情報通信学会IBISML研究会, 東京 (2010. 6. 14)

④ 神嶌 敏弘, 赤穂 昭太郎, 佐久間 淳 “カスタム価格設定推薦システム” 人工知能学会全国大会, 長崎 (2010. 6. 9)

⑤ T. Kamishima “Invited Talk: Object Ranking” ECMLPKDD 2009 Workshop: Preference Learning, Bled, Slovenia (2009. 9. 11)

⑥ 神嶌 敏弘, 赤穂 昭太郎 “転移学習を利用した集団協調フィルタリング” 人工知能学会全国大会, 高松 (2009. 6. 17)

[図書] (計 3 件)

① T. Kamishima and H. Kazawa and S. Akaho “A Survey and Empirical Comparison of Object Ranking Methods” In Preference Learning, Springer, pp.181-201 (2010)  
DOI: 10.1007/978-3-642-14125-6\_9

② T. Kamishima and S. Akaho “Dimension Reduction for Object Ranking” In Preference Learning, Springer, pp. 203-215 (2010)  
DOI: 10.1007/978-3-642-14125-6\_10

③ T. Kamishima and S. Akaho “Efficient Clustering for Orders” In Mining Complex Data, Springer, pp.261-280 (2009)  
DOI: 10.1007/978-3-540-88067-7\_15

[その他]  
ホームページ等  
<http://www.kamishima.net/>

## 6. 研究組織

### (1) 研究代表者

神嶌 敏弘 (Toshihiro Kmaishima)  
独立行政法人産業技術総合研究所・ヒューマンライフテクノロジー研究部門・主任研究員  
研究者番号: 50356820

### (2) 研究分担者

赤穂 昭太郎 (Shotaro Akaho)  
独立行政法人産業技術総合研究所・ヒューマンライフテクノロジー研究部門・研究グループ長  
研究者番号: 40356340

### (3) 連携研究者

なし