

称号及び氏名	博士（経済学） 中原 孝信
学位授与の日付	平成21年3月31日
論文名	グラフ分割問題を利用した推薦システムに関する研究
論文審査委員	主査 森田 裕之 副査 竹安 数博 副査 石垣 智徳

論文要旨

本論文では、グラフ分割手法を応用した新たな推薦システムを提案している。推薦システムは、E-commerceサイトを運営する企業、そして、その利用者数の増加に伴い、近年、積極的に実用化が行われているシステムである。推薦システムの実現方法には、検索対象の内容を考慮して推薦を行う内容ベースフィルタリングと、他人の好みに基づいて推薦を行う協調フィルタリングがあり、本論文は、協調フィルタリングを用いた推薦システムを対象にしている。

推薦システムでは、嗜好データが利用されるが、そのデータは、大きく2種類に分類される。1つは、評価値データで、利用者に好き嫌いの程度や関心の有無を質問し、利用者に評価値として、星の数などで回答してもらう方法である。もう一方は、2値（購買/非購買など）データであり、利用者の示す行動から、嗜好や関心を推察し利用する。これには、購買履歴データやWebのクリックログデータなどが該当する。これまでに、協調フィルタリングが対象としてきたデータは、主に評価値データで、購買履歴データなど評価値の無い2値データに対しても適切な推薦が行える方法を実現することが重要である。

提案手法は、両方のタイプのデータに適用することができる。まず、対象データをアイテム間の関係性を考慮したグラフ構造で表現する。そして、それにグラフ分割手法を適用して少数のクラスにアイテム进行分类する。次に、アイテム間の接続関係の強さを考慮することで、計算対象となるアイテムを限定し、最終的に推薦すべきアイテムを決定する。推薦アイテムを決定する方法は、ターゲットユーザがすでに評価したアイテムと接続関係の強いアイテムから順に推薦する方法と、平均最

高方式の1つであるドント方式により、すでに評価したアイテムの傾向を反映させて、各クラスタから推薦すべきアイテム数を決定し、その数に従って推薦する方法を用いている。

提案手法の有効性を確認するために、計算実験として、映画の評価値データと、2値データである音楽CDに関するID付きPOSデータに提案手法をそれぞれ適用した。提案手法の比較対象は、伝統的なメモリーベース法の協調フィルタリングを利用した。この方法は、他の手法に比べ、計算コストの観点からは劣る部分もあるが、推薦アイテムの質に関しては優れている。計算実験の結果、提案手法によって推薦したアイテムは、既存のメモリーベース法の協調フィルタリングを利用して推薦したアイテムよりも、精度の観点で優れていたことから、提案手法は有効であると考えられる。

本論文の構成は次のとおりである。2章では、データマイニングと企業活動との関係を述べており、3章では、推薦システムの既存研究について概観している。そして4章では、提案手法について詳細に記述し、5章と6章では、それぞれ異なる種類のデータに対して提案手法を適用し計算実験を行っている。

2章は、データマイニングを概観している。近年、情報通信技術の発展により、膨大なデータが容易に蓄積できるようになり、それを分析することで、有益な情報や知識を獲得し、他社との差別化による競争優位性を確立しようという試みから、データマイニングは注目を集めている。そして、多くの企業で導入が進んでいる。ここでは、それらの導入事例を概観し、考察することで、データマイニングが、企業活動にとって有用であることを明らかにしている。

3章では、推薦システムに関する研究を概観している。推薦システムは、企業が運営するE-commerceサイトで実用化され、多くの企業で導入されている。協調フィルタリングは、購入したアイテムと関連するアイテムを薦めて追加購入を促したり、ユーザの望むアイテムを薦めたりすることにより、顧客とサイトの間価値を創造させ、顧客ロイヤルティを向上させる。そしてサイトの売り上げ増加に貢献している。

協調フィルタリングの実現方法は大きく、メモリーベース法、モデルベース法、そして、ハイブリッド法に分類される。メモリーベース法では、データをあらかじめデータベースに格納し、推薦が行われる度に嗜好に関する予測値の計算が行われ、推薦するアイテムが決定される。一方モデルベース法では、推薦システムが利用される以前に、嗜好を表すモデルを構築しておき、推薦を行う際には、そのモデルを利用し推薦するアイテムが決定される。そしてハイブリッド法は、メモリーベース法とモデルベース法を両方兼ね備えた方法であり、本稿で提案している方法は、このハイブリッド法に分類される。

メモリーベースは更にユーザ間型とアイテム間型に分けられる。どちらも類似度

を計算する点では共通しているが、ユーザ間型は、ターゲットユーザと好みの類似するユーザをまず発見し、類似ユーザの好みを利用する方法である。一方、アイテム間型は、ターゲットユーザが好むアイテムと類似するアイテムを推薦する方法である。類似度の計算は主にピアソンの相関係数などが利用される。計算実験ではこれらの方法と提案手法の比較を行っている。

推薦システムを評価する尺度として、様々な評価指標が提案されている。特に予測の正確性を判断する尺度として、予測評価値と実際の評価値との近さを表す *MAE* や、推薦リストに含まれるアイテムを、ユーザが実際に好むかどうかを評価する精度、再現率、そしてその2つの調和平均を用いた *F-measure* などの評価指標が広く用いられている。5章と6章の計算実験の結果は、これらの評価指標を利用して提案手法の有効性を評価している。

4章では、グラフ分割手法を用いた新たな手法を提案している。推薦システムを構築するに際し、まず、アイテムを複数のグループに分類するために、アイテム間の接続関係の強さを反映したグラフにグラフ分割手法を適用する。そして、分類したクラスタに属するアイテムを利用して、ターゲットユーザが未購入のアイテムを対象に、推薦すべきアイテムの効率的な選択を実現する。

グラフ分割問題は、分割後の各グラフの頂点に関する重みの和が均衡するという制約を満たし、異なる集合を横断する枝のコスト和が最小になるように、グラフの頂点を分割する問題である。頂点に単位コスト、そして枝にアイテム間の関係の強さを表す値を与えることで、アイテム間の強い関係は保存されたまま、分割後の各クラスタに属するアイテム数が、各クラスタで均等になるように分割される。提案手法では、枝の重みにサポート値を利用している。

そして次に分割された各クラスタを利用して、ターゲットユーザが未評価のアイテムに対して、評価値の予測と推薦すべきアイテムを決定する。まず、評価値の予測については、ターゲットユーザが普段どの程度の評価値を与えており、対象アイテムが他のユーザからどの程度評価されているのか、そして、各アイテムに対する評価は、対象データ全体の評価の傾向とどの程度異なるのかという点を考慮した予測を行っている。

次に、推薦対象となるアイテムを選択する方法として、まず、アイテム間の接続関係を利用し、ターゲットユーザがこれまでに評価したアイテムと関連の強いアイテムが推薦の対象になるように、推薦の候補として選択する。そして、候補となったアイテム集合を対象に、ターゲットユーザがすでに評価したアイテムと接続関係の強いアイテムから順に推薦する方法と、ドント方式を用いた方法を提案している。ドント方式は、平均最高方式の1つであり、政党名簿比例代表制度などで用いられている方法である。この方法によって、すでに評価したアイテムの傾向を反映させて、各クラスタから推薦すべきアイテム数を決定し、その数に従ってアイテムを推

薦することが可能である。5章では、映画の評価値データ（★印が1から5）に対して、上述の提案手法を適用し、評価値データに対する提案手法の有効性を確認するために、メモリーベースの協調フィルタリングを2種類と、最も単純な方法として、平均評価値だけを利用した方法で比較を行った。その結果、提案手法は、予測評価値と実際の評価値との近さを表すMAEが0.748と最も小さく、予測精度は最も良かった。また、推薦の質を示す評価指標である、精度、再現率、そして*F-measure*の値は、5アイテムと10アイテムを薦めた場合の両方において、提案手法の値が高く、特にドント方式による推薦が最も高い値を示していた。ドント方式を利用することで、特定のジャンルだけに偏った推薦ではなく、ユーザの好みを反映させたバランスの良い推薦が行えるため、このことが良い結果に結びついていると考えられる。

6章では、2値データである音楽CDに関するID付きPOSデータに提案手法を適用した。ここでも比較対象として、メモリーベースの協調フィルタリングを2種類と、ランダムサンプリングによる方法を利用した。ここでは0, 1からなる2値データのため、MAEによる評価は行わず、推薦の質を示す評価指標である、精度、再現率、そして*F-measure*を利用した。また、ここではランダムサンプリングの精度に比べ、各手法が何倍の精度を得たかを示す改善比も評価指標として利用している。結果として、2値データの場合も、推薦の質を示す評価指標の値は、5アイテムと10アイテムを薦めた場合の両方において、提案手法の値が高く、特にドント方式による推薦が最も高い値を示していた。また、ドント方式を利用した提案手法は、ランダムに比べ、精度の差が約14倍あることが確認できた。

以上の2種類の異なるデータに対して、提案手法を適用した結果、メモリーベース法よりも有効な結果が得られることを示した。今後の課題は、グラフの分割数に関する基準を明確にすること、本研究で得られた示唆を基に、推薦アイテムを選択する方法について更なる改善を検討すること、そして、より多くのデータに対して提案手法を適用し、より実用化に向けた改善を行いたいと考えている。