

A study on recommender systems based on covering-based rough sets

その他（別言語等）のタイトル	被覆に基づくラフ集合を用いた推薦システムに関する研究
著者	張 志鵬
学位名	博士（工学）
学位の種別	課程博士
報告番号	甲第407号
研究科・専攻	工学専攻
学位授与年月日	2017-09-25
URL	http://doi.org/10.15118/00009628

氏 名 張 志鵬 (チョウ シホウ)

学 位 論 文 題 目 A study on recommender systems
based on covering-based rough sets
(被覆に基づくラフ集合を用いた推薦システムに関する研究)

論 文 審 査 委 員 主査 教授 工藤 康生
准教授 岡田 吉史
准教授 渡邊 真也

論文内容の要旨

インターネットの高速な発展に従って、人類は情報社会とネットワークの時代に入った。人々はインターネットから多くの情報と便利なサービスが得られる一方で、膨大なデータ及び無駄な情報にも接することになる。推薦システム (RS) はこの問題に呼応して出現した。RS はユーザーの決定と自身によって見つけられないアイテムの発見を助けることができる。

協調フィルタリング (CF) は良好な性能のため、RS に広く応用されている。一般的に言えば、ユーザベース協調フィルタリング (UBCF) とアイテムベース協調フィルタリング (IBCF) の二つは CF において重要なアプローチであり、多くの商業 RS に適用されている。しかし、様々なデータ及び複雑なアプリケーション環境の影響で、CF は多くの問題に直面している。例えば、UBCF はアクティブユーザーに対する推薦で高い推薦精度と推薦の多様性を両立させることができない。また、UBCF は十分な情報を持っていない新しいユーザーに個人化した推薦を提供することができない。さらに、IBCF ではアイテムの重要性によって、重みに差を付けることができない。上記の問題点を考慮して、この論文では以下の研究を行った：

(1) アクティブユーザーに個人化した推薦を提供するために、被覆に基づくラフ集合を用いた新しいカバーベース協調フィルタリング (CBCF) を提案した。CBCF では被覆に基づくラフ集合における被覆縮約を UBCF に導入し、すべてのユーザーから冗長なユーザーを削除する。そして、 k -最近傍ユーザーは縮約したユーザーから選択される。実験結果によって、現実の推薦システムでよく現れる疎なデータについて、CBCF は UBCF より性能が良く、アクティブユーザーに対する推薦で高い推薦精度と推薦の多様性を両立できることが示された。

(2) 新しいユーザーに個人化した推薦を提供するために、新しいユーザーの特徴の詳細な分析を通じて、決定クラスを再定義することで、CBCF を改良した。元の CBCF と異なり、改良された CBCF の決定クラスはユーザーアイテム評価行列から容易に抽出される。さらに、改良された CBCF は特別な追加の情報を必要とせずに、個人化した推薦を提供することができる。実験結果によって、改良された CBCF は既存の関連研究より良く、新しいユーザーに対する推薦で高い推薦精度と推薦の多様性を両立できることが示された。

(3) 伝統的な IBCF はすべてのアイテムに同じ重みを与える。しかし、類似度と予測値を計算する時に、いくつかのアイテムがより重要な影響を持つ可能性がある。アイテムの分散重み付けもまた考慮されるべきである。この論文では、アイテムの重みを付け直すために、時間ベース協調度と被覆度を提案して、伝統的な IBCF に適用した。実験結果によって、提案した IBCF は伝統的な IBCF と他の関連研究に比べて、良い推薦結果を提供することが確認された。

キーワード：

推薦システム；協調フィルタリング；被覆に基づくラフ集合；カバーストリー；個人化推薦；分散重み付け

ABSTRACT

With the rapid development of Internet, the human race has entered the information society and the network era. Internet could provide people with more and more information and services; however, people have to face enormous data and useless information when they enjoy the convenience brought by Internet. Recommender system (RS) has emerged in response to this challenge, which can advise users when making decisions and help users discover items they might not find by themselves.

Collaborative filtering (CF) approach is popularly used in RSs owing to its satisfactory performance. Generally speaking, user-based collaborative filtering (UBCF) and item-based collaborative filtering (IBCF) are two significant approaches in CF, they have been successfully applied to many commercial RSs. However, with various kinds of data and complicated application environment, CF approaches are facing many challenges. For instance, UBCF cannot provide recommendations for an active user with satisfactory accuracy and diversity simultaneously. Personalized recommendations cannot be provided by UBCF for a new user which often has insufficient information. In

addition, items that make a more significant contribution cannot have high weighting in IBCF. In view of the above key issues, this dissertation launched a study of the following aspects:

(1) Aiming to provide personalized recommendations for an active user, we apply covering-based rough sets to improve UBCF, and propose a new covering-based collaborative filtering (CBCF) approach. CBCF inserts a user reduction procedure into UBCF, covering reduction in covering-based rough sets is utilized to remove redundant users from all users. Then, k-nearest neighbors are selected from candidate neighbors comprised by the reduct-users. Our experiment results suggest that, for the sparse datasets that often occur in real RSs, CBCF outperforms than the UBCF, and can provide satisfactory accuracy and coverage for an active user at the same time.

(2) In order to provide personalized recommendations for a new user, through a detailed analysis of the characteristic of new users, we reconstruct a decision class to improve the previous CBCF. Unlike the previous CBCF, the decision class in improved CBCF can be extracted easily from the user-item rating matrix. Furthermore, the improved CBCF could provide personalized recommendations without needing special additional information. Our experiment results suggest that the improved CBCF significantly outperforms those of existing work and can provide personalized recommendations for a new user with satisfactory accuracy and diversity simultaneously.

(3) The traditional IBCF approach treats all items as the same weighting; however, because some items may have more important impact when computing the similarity and predictions, item-variance weighting should also be considered. In this paper, we present the time-based correlation degree and covering degree, and apply them to the traditional IBCF approach to rearrange the item weighting. Our experimental results suggest that, our proposed approach can produce recommendations superior to the traditional IBCF and other existing work.

Keywords:

Recommender systems; Collaborative filtering; Covering-based rough sets; Covering reduction; Personalized recommendations; Item-variance weighting

論文審査結果の要旨

近年、インターネット上での情報の爆発的増加に伴い、ユーザが関心を持つと予想される情報を抽出し提示する推薦システムの必要性が高まっている。協調フィルタリングは推薦システムで一般的に用いられる手法であり、主にユーザ間の類似性に基づくユーザベース協調フィルタリング(UBCF)とアイテム間の類似性に基づくアイテムベース協調フィルタリング(IBCF)に大別される。しかし、UBCFおよびIBCFは、データの疎性および推薦内容への高い要求などの要因から、様々な問題を抱えている。そこで本論文では、ラフ集合理論の一種の拡張である被覆に基づくラフ集合(Covering-based rough set)を用いて、UBCFおよびIBCFの性能を向上させることを目的とし、以下の成果を得ている。

(1) UBCFでは、推薦精度(Accuracy)と推薦の多様性(Diversity)の両指標で同時に高評価を得る推薦を行うことが困難であった。これに対し、本論文では推薦を受けるユーザによるアイテム評価スコアに基づき推薦対象アイテムを絞り込み、更に被覆に基づくラフ集合における被覆縮約(Covering reduction)を用いることで、ユーザへの推薦を行う際に冗長とみなされるユーザを除去して推薦を行う手法(CBCF)を提案した。計算機実験により、疎性の高いデータセットでは、CBCFはUBCFと同程度の高いAccuracyとUBCFを上回るDiversityを同時に達成する推薦を行えることが確認された。

(2) (1)で提案したCBCFは、推薦を受けるユーザが十分多くのアイテムについて評価スコアを持つことを前提としているため、新規ユーザに対して使用することは困難であった。そのため、各アイテムに対する全ユーザからの評価件数を用いて推薦対象アイテムを絞り込むようにCBCFを改良した。計算機実験により改良版CBCFを従来のUBCFおよびその関連手法と比較した結果、疎性の高いデータセットでは、改良版CBCFは新規ユーザに対して、UBCFおよびその関連手法より高いAccuracyと高いDiversityを同時に達成する推薦を行えることが確認された。

(3) 従来のIBCFでは、アイテム間の類似性を評価する際にアイテムの重要度を表す重みづけが考慮されていないため、これまでに様々な重みづけ手法が提案されている。本研究では、時間ベース協調度と被覆度の2種類の指標を新たに導入し、IBCFにおけるアイテムの重みづけ手法を提案した。計算機実験により、提案した重みづけIBCFは、従来のIBCFおよび重みづけを行う従来手法との比較で、高いAccuracyと高いDiversityを同時に達成する推薦を行えることが確認された。

以上、本論文は被覆に基づくラフ集合の手法をUBCFおよびIBCFに用いることで、推薦システムによる推薦の精度と多様性の両立を実現している。これらの成果は推薦システムの分野の発展に寄与するところが大きく、本論文は博士(工学)の学位に値すると認められる。