

氏 名 Padipat SITKRONGWONG

学位(専攻分野) 博士(情報学)

学位記番号 総研大甲第 2194 号

学位授与の日付 2020 年 9 月 28 日

学位授与の要件 複合科学研究科 情報学専攻
学位規則第6条第1項該当

学位論文題目 Unsupervised Context Extraction for Review-Based
Recommendations

論文審査委員 主 査 教授 高須 淳宏
教授 相澤 彰子
教授 山田 誠二
准教授 相原 健郎
准教授 Saranya Maneeroj
チュラロンコン大学

(Form 3)

Summary of Doctoral Thesis

Name in full Padipat SITKRONGWONG

Title Unsupervised Context Extraction for Review-Based Recommendations

Recommender system (RS) was invented to provide a personalized recommendation on the suitable items that would be interested by an individual user. A standard RS approach such as collaborative filtering utilizes past interactions from users (e.g. their numeric ratings on items) to build a predictive model for future recommendations on unseen items. However, considering only rating data might not be sufficient to overcome two common challenges in RS that need to be addressed in order to produce effective recommendations: how to improve the prediction accuracy, and how to alleviate the rating sparsity. First, the success of the recommendation engines are often measured by their prediction accuracy. The highly accurate recommendations would increase user satisfactions on the systems, and consequently increases their revenues. On the other hand, the rating sparsity occurs when most users only rate few portion of all available items. Having insufficient amount of ratings, the recommendation engines would not be able to learn high-quality preference patterns from users, and therefore, result in low recommendation accuracy.

In order to improve recommendation accuracy, a context-aware recommendation approach has been introduced. The main idea is that a contextual information or simply "context" such as location, weather or season could have a major influence on users' decisions when they are choosing items. For example, most users tend to travel to the beaches in summer, whereas the ski-resorts are more preferred in winter. Incorporating contexts could help suggesting more relevant items to users, which can be a crucial factor for producing more accurate recommendations than the standard rating-only recommendation methods. However, there are two main concerns that could have significant impact on the performances of context-aware recommendations. First, obtaining contexts is not a trivial task. Many works collected them by first predefining a list of contexts, and then ask users to select from those values as their contexts at the time they consumed items. Predefining optimal values of contexts, however, could potentially be very expensive since their values are varied across different recommendation domains. After obtaining contexts, the next challenge is to identify which contexts are relevant to a specific recommendation task. Incorporating too many contexts not only degrades the quality of recommendations, but also increases dimension of data and thus triggering a sparsity problem.

On the other hand, a review-based recommendation approach has been proposed with the main goal to alleviate rating sparsity problem occurred in standard RS by utilizing

user-generated reviews. In many systems, users have options to write text reviews on products or services they have purchased, in addition to the ratings. In reviews, users can provide comments explaining reasons behind their decisions on items, which offer more meaningful and useful information than numeric ratings. Many review-based recommendation methods take this opportunity to effectively extract the user personal preferences and item unique features from reviews in the form of numeric user and item representation vectors. Recent works in review-based recommendations proved that, learning such representations from reviews and using them for rating prediction, is more robust to rating sparsity than learning them from ratings alone.

Furthermore, in reviews users can express opinions describing their experiences, situations, and satisfaction on their selected items, which can be a valuable source of contexts. Extracting contexts from reviews could be the key to overcome the challenge of obtaining contexts, and provides recommendations with high accuracy as well as robustness to rating sparsity. Moreover, since contexts are information that affects users' decisions on items, they might be useful in constructing high-quality user and item representations for review-based recommendations.

In addition to a contextual information and user-generated reviews, this thesis also studies another type of data, the multi-criteria ratings. Unlike a standard RS approach that considers only single overall ratings from users, the multi-criteria recommendation approach lets users express their preferences toward items in multiple aspects, such as service or cleanliness of hotels, with multi-criteria ratings. Effectively utilizing such multi-criteria ratings could help analyzing user preferences and item features in more details, and yielding even more accurate recommendations.

In summary, this thesis studies extensively the challenges on extracting contexts from reviews, and how to utilize such extracted contexts for personalizing recommendations that would satisfy both accuracy and sparsity. In addition to contexts and reviews, this thesis also studies on how to improve the prediction accuracy with multi-criteria ratings. These studies result in three main proposed methods as follows.

First, the context-aware region embedding (CARE), a novel unsupervised method for defining and extracting contexts from reviews, is proposed. CARE deals with the problems of obtaining and identifying relevant contexts in a standard context-aware recommendations by applying region embedding techniques to extract and represent contexts from reviews. Such relevant contexts are represented as text regions that influence the distributions of ratings. The experiments demonstrated that CARE has a flexibility to extract contexts from review data in any recommendation domains. Moreover, the extracted contexts effectively captured the polarity of reviews' ratings, which can be useful for the rating prediction task.

Next, the attentional interaction model for context-aware region embedding (CARE-AI), an extended model of CARE for rating prediction is proposed. CARE-AI aims to overcome the challenges of learning user and item representations from reviews of the

recent review-based recommendations. This model introduces the interaction and attention modules for constructing the user and item representations from the extracted contexts, which are derived from CARE. Such representations are generated specifically for each particular review, based on the different level of relevance among contexts in that review. Extensive experiments show that, not only achieving better prediction accuracy compared to the state-of-the-art review-based recommendations, CARE-AI was also more robust in generating recommendations in the rating sparsity situations.

Finally, a novel method for multi-criteria rating conversion is proposed. This method aims to overcome the limitation of applying a standard rating conversion techniques to multi-criteria recommendations, which might cause a loss in implicit relation among multi-criteria ratings. The proposed method applies the principle component analysis to extract the multi-criteria rating patterns from users, which are then used to convert all multi-criteria ratings simultaneously to maintain their implicit relation. The experiments demonstrated that the proposed method outperforms both single and multi-criteria rating conversion techniques with higher accuracy and prediction coverage.

博士論文審査結果

Name in Full
氏名 Padipat SITKRONGWONG

Title
論文題目 Unsupervised Context Extraction for Review-Based Recommendations

本学位論文は、レビューテキストからのコンテキストの抽出法とその情報推薦への適用法を提案している。映画やホテルなどの各種サービスの利用者が書くレビューテキストにはコンテキストに関わる情報が多く含まれていることに着目し、コンテキストに関連するフレーズを抽出する埋め込み表現 Context-Aware Region Embedding (CARE)を提案している。提案手法はアイテムへの評価値と相関の大きなフレーズはコンテキストを表している可能性が高いという仮定のもとコンテキストを表すレビュー部分テキストを抽出する。次に、CAREを用いてレビューテキストから利用者およびアイテムの特徴ベクトルを求め、その類似度に基づいてアイテムを推薦する方法を提案している。

本学位論文は7章で構成され英語で書かれている。

第1章では、本研究が扱うコンテキストウェア推薦法の概要と研究の意義、本研究で取り組んだ課題および主な貢献を示している。

第2章では、推薦システム一般および、コンテキストウェア推薦法、レビューに基づく推薦法の概要と技術的課題について述べている。

第3章では、本研究の関連研究であるレビューに基づく推薦法およびコンテキストの抽出法についての関連研究をまとめている。続く3つの章で本論文の主たる貢献を述べている。

第4章ではレビューテキストからコンテキストを抽出する手法 CARE を提案している。まず、レビューに含まれる語でアイテムの評価値と相関の高い語をコンテキスト語として抽出する。続いて、その語を含むテキスト部分領域から評価値を生成するニューラルネットワークを用いてその特徴ベクトルを求めコンテキストに対する埋め込み表現とする。従来の研究では、単一の単語でコンテキストを表すことが多かったが、本研究ではテキスト領域を用いることでより正確にその特徴を捉えられるところに特徴がある。ソフトウェア、旅行、映画の3つの異なる領域のレビューテキストを用いて比較評価実験を行い、テキスト領域を使用する提案手法が既存手法より正確に評価値を推定できることを実験的に示している。

第5章では、CARE用いたアイテムの評価値の推定法を提案している。CAREを用いて抽出したコンテキストの特徴ベクトルを用いて、利用者およびアイテムの特徴ベクトルを補正し、回帰分析用のニューラルネットワークで評価値を推定する。4章と同様に3つの異なる領域のデータを用いて評価実験を行い、CAREによって抽出されたコンテキストが評価値推定に効果があることを実験的に示している。また、レビューを用いた既存の情報推薦法と比較して推薦精度が高いことを確認している。

第6章は、CAREの複数評価軸に基づいた推薦法(Multi-Criteria Recommender System:

MCRS)への応用について述べている。MCRS ではアイテムの評価軸を複数用意し、軸ごとの評価から総合評価を推定する。評価軸は推薦するアイテムの種類によって異なるため、対象領域に適した軸を選ぶことが重要となる。本章では、CAREによって得られる部分テキストを4つのカテゴリに分類し、MCRS の評価軸に適したカテゴリとその利用可能性について論じている。

第7章では、以上の結果をまとめるとともに今後の課題を示している。

公開論文発表会において、出願者はおよそ45分で博士論文の内容を説明し、その後、15分程度の質疑応答が行われた。続いて審査員による質疑応答を行い、提案手法のパラメタのチューニング法、抽出されたコンテキストの分類、他の問題への応用可能性について質問とコメントが寄せられ、出願者は適切に回答した。

質疑応答後に審査委員会を開催し、審査委員で議論を行った。博士論文審査の結果、出願者は情報学分野の十分な知識と研究能力を持つと認められ、また研究内容は学位論文として十分なレベルの新規性、有効性があると認められた。本論文の内容に関し、国際学術雑誌に2編(IEEE Access, International Journal of Computers and Applications)、査読付き国際会議に1編の主著論文が採択されている。以上より、審査委員会全員一致で、博士論文として十分な水準の研究であると認め、学位の授与に値すると判断した。