

氏 名 NGUYEN THAI BINH

学位(専攻分野) 博士  
(情報学)

学位記番号 総研大甲第 2081 号

学位授与の日付 平成 31年 3 月 22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 Implicit Feedback Embeddings for Recommender Systems on  
Sparse Data

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  
教授 宮尾 祐介 東京大学  
大学院情報理工学系研究科

(Form 3)

## Summary of Doctoral Thesis

Name in full: NGUYEN THAI BINH

Title: Implicit Feedback Embeddings for Recommender Systems on Sparse Data

In information recommendation, the preferences of users and the attributes of items to be recommended are represented by feature vectors. The items are recommended based on the similarity of the corresponding feature vectors. Characteristics of users and items are learned from the rating for items of users, etc. In general, the amount of ratings is limited, and extracting effective features from sparse data will be necessary. Also, for users who newly participate in the system, it is difficult to obtain user characteristics because the information is limited. The aim of this thesis is to propose effective models of recommender systems for such sparse data.

We first look into the rating prediction problem, one of the essential tasks in recommender systems. Rating is another kind of feedback known as explicit feedback. Different from the implicit feedback, the amount of rating data is limited because it requires users to provide the ratings explicitly. I first proposed a feature extraction method that utilizes the data that is easy to obtain, such as the click history recorded in the log when a user examines an item. In this method, the features of items are extracted based on two sources of feedback: rating data and click data. We show that exploiting the click data can supplement the shortage of rating.

We further advance this research and proposed a method to analyze the similarity of items in more detail. The proposed model can identify two kinds of relationships between items, (1) items that can be replaced, such as products of company A and products of company B, and (2) items that are often purchased together, such as bread and butter. In this research, the item features are extracted under the assumption that there is a strong correlation between the items that are positively evaluated by each user.

Finally, we introduce a model to learn the representations of the products based on their titles to model a collection of shopping transactions. The learned representations can help identify two kinds of relationships between products: similar products and “also-buy” products. The proposed model can be used for multiple purposes: next product recommendation, similar product recommendation, also-buy product recommendation, and product search by keywords. Because the model learns the representations from product titles, it can deal with the cold-start problem, the problem of modeling new products which have not appeared in any shopping transactions.

## 博士論文審査結果

Name in Full  
氏名 NGUYEN THAI BINH

Title  
論文題目 Implicit Feedback Embeddings for Recommender Systems on Sparse Data

出願者は、利用者から得られる限られた情報を有効活用しアイテムを推薦するシステムの研究を行い、その成果を博士論文としてまとめた。博士論文における研究は、推薦システムで主に用いられる利用者のアイテム評価データに加え、アクセスログなどの暗黙的なデータやアイテムの説明テキストを用いて特徴ベクトルを構成する方法を提案し、既存手法と比較し、アイテム推薦精度の向上が図れることを実験的に示している。

博士論文は6章で構成され英語で書かれている。第1章では、近年急速に普及している推薦システムの概要を述べるとともに、既存推薦システムの課題を論じ、本研究で取り組んだ問題と貢献をまとめている。

第2章では、推薦システムを分類し、本研究の基礎となる協調フィルタリングおよび利用者やアイテムの特徴ベクトルの構成法についてサーベイしている。またアイテム集合およびアイテム列を対象とした推薦システムの研究を概説している。

続く3つの章で本論文の主たる貢献を述べている。第3章ではアクセスログを活用した推薦法を提案している。ここでは、利用者によるアイテム評価データに加え、アイテムの情報へのアクセスログを併せて用い、利用者およびアイテムの特徴ベクトルを構成する確率モデルおよびアイテムと利用者の特徴ベクトルから評価値を推定する方法を提案している。アイテムの特徴ベクトルに加え、アイテム間の関係を表すコンテキストベクトルを導入し、アクセスログよりコンテキストベクトルを構成するところに特徴がある。大規模な評価用データを用い、提案手法の有効性を実験的に示している。

第4章では、ニューラルネットワークを用いて利用者とアイテムの特徴ベクトルを構成する方法を提案している。ここでは、ニューラルネットワークをエンコーダとして用いているが、第3章で提案した手法と同様にアイテムのコンテキストベクトルを導入し、利用者の購入履歴からコンテキストを抽出する。評価用データを用いた実験では、推薦性能の向上に加え、特に購入履歴の少ない利用者に対する推薦に有効であることを示している。

第5章では、アイテムの購入トランザクションから特徴ベクトルを構成する方法を提案している。ここでは、ニューラルネットワークを用い、アイテムの説明テキストからアイテムの特徴ベクトルを構成すると同時に購入されたアイテム集合からアイテム間の関係を表すコンテキストベクトルを構成する。評価実験により、ともに購入することが多いアイテム集合や置き換え可能なアイテム集合の抽出に有効なモデルであることが示されている。

第6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、20分程度の質疑応答が行われた。審査委員からは、メモリベース協調フィルタリングとの違い、ニューラルネットワークを用いたモデルとその学習アルゴリズムの詳細、評価実験

に用いられたデータを提案手法に用いることの妥当性等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、情報学専攻が定めるトップ国際会議 IJCAI を含む4編の論文が査読付き国際会議に採択されている。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。