International Journal of Multimedia Data Engineering and Management
Volume 10 • Issue 1 • January-March 2019

Enhancing Rating Prediction by Discovering and Incorporating Hidden User Associations and Behaviors

Ligaj Pradhan, University of Alabama at Birmingham, Birmingham, USA

ABSTRACT

Collaborative filtering (CF)-based rating prediction would greatly benefit by incorporating additional user associations and behavioral similarity. This article focuses on infusing such additional side information in three common techniques used for building CF-based systems. First, multi-view clustering is used over neighborhood-based rating predictions. Secondly, additional user behavior knowledge discovered by mining user reviews are infused into non-negative matrix factorization (NMF) techniques. Finally, the article explores how to infuse such additional behavioral knowledge into a Deep Neural Network (DNN) based DF architecture. The article also explores using term frequency-inverse document frequency (TF-IDF) vectors as the input to DNN. Since TF-IDF does not directly capture the conceptual contents of the text or the behavioral aspects of the writer, the article also proposes a novel scheme called topic proportions-inverse entity frequency (TP-IEF) that uses topics discovered from reviews instead of words to better capture semantic associations between users and items.

KEYWORDS

Collaborative Filtering, Deep Neural Networks, Matrix Factorization, Multi-View Clustering, Rating Prediction, TF-IDF, Topic Modeling, TP-IEF, User-Concerns

DOI: 10.4018/IJMDEM.2019010103

INTRODUCTION

With tremendous growth of e-commerce and information over the Internet, recommendation and rating prediction systems have become an integral part of almost every online system. Collaborative Filtering (CF) is a type of system which exploits past user behavior and preferences to recommend new items liked by users with similar behavior and preferences. This article focuses on discovering hidden associations between users and their similarities based on their behavior to enhance CF-based techniques. Discovered user associations and behavioral information are primarily infused in Neighborhood discovery, Non-Negative Matrix Factorization (NMF) and Deep Neural Network (DNN) based CF techniques and demonstrate how such systems can be improved.

The article first explains some effort to improve CF-based rating prediction by improving neighborhood discovery. This section elaborates how multi-view clustering can be used to discover better user and item clusters. Multi-view clustering refers to the clustering process where multiple modalities will be used collectively to make the neighborhood search more effective and accurate. The accuracy of cluster-based CF is expected to improve by improving the user and item clustering.

Then the article focuses on enhancing NMF by incorporating user behavioral knowledge. Different users express their concerns regarding different aspects like cost, proximity, location, time, cleanliness, etc. in their reviews. Such concerns and interests expressed in the user reviews are referred to as user concerns (UC) in this work. This section presents how user concerns and interests can be extracted from user reviews and subsequently represented with a user concern vector (UC-Vector). The article then shows how can be used as an additional side information to improve the prediction accuracy of NMF-based rating prediction model.

DNNs have achieved significant success in predicting user sentiments and ratings over various range of items. Several neural models project the initial user and item vectors into semantic spaces before predicting how much the user likes the items (Huang, He, Gao, Deng, Acero & Heck, 2013; Le, 2015). To further improve the rating prediction accuracy, the next explains how the extracted UC-vectors are incorporated into DNN based architectures. The author experiments with a deep architecture which is referred to as Deep Semantic Projection based Rating Predictor (DSPRP) and which consist of two parallel branches to learn user and item latent vectors. A vector representation of ratings given by user and received by items (rating vectors) are used as an input to DSPRP to learn the latent user and item representations. The dot product of the learned user and item latent vectors then learns to gives the corresponding user-item rating values. Later, UC-Vectors are fed into such Deep Architectures to regulate the learning of latent user vectors and enhance the DNN-based rating prediction. In addition to using only rating vectors as the initial input to the DNN model, the author also explored using only Term Frequency-Inverse Document Frequency (TF-IDF) vectors as an alternative input. However, it is based on statistical learning alone and doesn't directly capture the conceptual contents of the text or the behavioral aspects of the writer. Hence, in the next section the author explains a novel technique to extract relatively low dimensional user behavioral vectors from the same text, from which TF-IDF vectors are extracted, and used these behavioral vectors to enrich the performance of TF-IDF to train a deep rating prediction model. Experiments show that adding such conceptual knowledge to TF-IDF vectors can significantly enhance the performance of TF-IDF vectors by only adding very little complexity.

The work presented in paper are evaluated using user-restaurant review dataset from the Yelp challenge dataset (Yelp Dataset Challenge, 2014) and user-hotel review dataset crawled from TripAdvisor (Wang, Lu & Zhai, 2011). The proposed approaches for making rating predictions are compared and evaluated with rating predictions systems made by using several state-of-the-art techniques on neighborhood discovery, NMF and DNNs.

Related Content

Data Gathering with Multi-Attribute Fusion in Wireless Sensor Networks
Kai Lin, Lei Wang, Lei Shuand Al-Sakib Khan Pathan (2012). *Advancements in Distributed Computing and Internet Technologies: Trends and Issues (pp. 159-181).*
www.irma-international.org/chapter/data-gathering-multi-attribute-fusion/59682

Subjective and Objective Trustworthiness of Acquaintance Peers
Yoshio Nakajima, Alireza Goudarzi Nemati, Tomoya Enokidoand Makoto Takizawa (2009). *Handbook of Research on Mobile Multimedia, Second Edition (pp. 484-500).*
www.irma-international.org/chapter/subjective-objective-trustworthiness-acquaintance-peers/21023

Static Signature Verification Based on Texture Analysis Using Support Vector Machine
Subhash Chandraand Sushila Maheshkar (2017). *International Journal of Multimedia Data Engineering and Management (pp. 22-32).*
www.irma-international.org/article/static-signature-verification-based-on-texture-analysis-using-support-vector-machine/178931

Efficiency and Performance of Work Teams in Collaborative Online Learning
Éliane M.F. Moreau (2009). *Encyclopedia of Multimedia Technology and Networking, Second Edition (pp. 444-450).*
www.irma-international.org/chapter/efficiency-performance-work-teams-collaborative/17434

New Paradigms: A Collaborative Web Based Research Tool
Hamish Holewa (2008). *Handbook of Research on Digital Information Technologies: Innovations, Methods, and Ethical Issues (pp. 57-67).*
www.irma-international.org/chapter/new-paradigms-collaborative-web-based/19835