University of Dar es Salaam Library Journal
Vol 17, No 1 (2022), pp 38-53
ISSN: 0856-1818

Rating Prediction based on Optimal Review Topics: A Proposed Latent Factors-Optimal Topics Hybrid Approach

James Chambua
Computer Science and Engineering Department, University of Dar-es-Salaam
Email: james.chambua@udsm.ac.tz

Abstract

Rating prediction is an inevitable problem which recommender systems (RS) need to address. Its goal is to accurately predict the rating a user will assign to a particular item. Predictions which utilize numerical ratings and review texts are biased and have low accuracy. Also, existing topic-based rating prediction approaches focus on finding the most preferred items through the identification of latent topics expressed in users' review texts. Even though the latent topics seem to represent most user review texts, they do not necessarily capture each user's preferences. The goal of this work is then to develop a more accurate model by considering product review texts analysis so as to gain additional preference knowledge. Hence, a hybrid algorithm that optimizes the latent topics is proposed. Specifically, the proposed approach finds appropriate weights for the topics of each review text. Rating prediction is critical task for RS because slight performance enhancement of the prediction accuracy results into significant improvements in recommendations. Experimental evaluation over real-world datasets revealed performance improvements of the proposed approach compared to alternative models. The proposed model can be used by RS in various domain such as e-learning, movie and hotel rating.

Keywords: Keywords: Rating prediction, topic modeling, latent factor model, sailor assignment problem, review texts.
https://dx.doi.org/10.4314/udslj.v17i1.4

Introduction

Online sites such as stores, microblogs, e-commerce and similar sites facilitate wide access to a huge number of products, services or contents which subsequently lead to information overload problem for their users. Each product or service in such sites is associated with numerous review texts which are intended to assist users in evaluating and comparing items in order to make informed decisions. As the amount of product review texts keeps growing rapidly (Wang & Ester, 2014), it becomes impossible for users to read all reviews or even determine a sufficient number of reviews to make an informed interpretation.

Users' review texts and numerical ratings in different application domains such as hospitality, music or movie have a great impact on how people make decisions (consume/purchase products or services) (J. Chen *et al*., 2016). Some rating prediction approaches utilize numerical ratings only to measure the similarity of users' preferences towards items in order to predict ratings. But users may provide same ratings towards items for completely different reasons making such mere utilization of numerical ratings unreliable. This situation further motivates considering preference knowledge acquired from review texts in order to appropriately weight the influence of numerical ratings. Therefore, rating prediction approaches should involve review texts analysis to facilitate profound understanding of users' preferences and motives behind such preferences as well as item aspects which most users consider relevant. This will help such approaches to be more competent in constructing a full contextual representation of users' preference knowledge and consequently make more accurate predictions. This makes rating prediction to be one of the most critical tasks for recommender systems (RSs) such that slight performance enhancements of predictions may result to significant improvement in recommendations.

Recommender systems address the information overload problem by employing rating prediction approaches which enable the generation of ranked lists of most preferred items for each user. To ensure more accurate rating prediction as well as providing reliable information which explains the reasons behind users' preferences, some existing works (e.g., McAuley & Leskovec, 2013; Xu, Lam, & Lin, 2014; Bao, Hui, & Zhang, 2014; Liu, Liu, Shen, & Li, 2017; Liu & Shen, 2018) have considered review texts topics as additional knowledge to enhance rating predictions. However, review topics generated by such models may not correspond or reflect each user's preferences. This is due to the fact that each user review text is assigned to particular topics based on topic distribution of all its words irrespective of the context or meaning of the words. The generated topics are also given equal importance/weight in predicting user preferences which may lead to inaccurate results. Hence, it's necessary to ensure that review topics correspond with expressed user preferences and allocate proper weights for each review text based on their topic distributions. The main objective of the study was to devise a rating prediction approach with more accurate prediction compared to similar existing approaches. Specifically, the study intends to a) explore latent factor and latent topic models used in rating prediction; b) develop a hybrid rating prediction model based on latent factors and topics; c) evaluate the accuracy performance of the developed hybrid model.

Related work

This section reviews related studies on rating prediction specifically focusing on latent factor models and topic-based models.

Latent factor models

Numerous rating prediction approaches which are based on latent factor models have been proposed (e.g., Koren, Bell, & Volinsky, 2009; Zhao *et al*., 2015). Such approaches are aimed at approximating the scores of the rating matrix for users towards items by using low-rank matrices consisting of latent factors for users' preferences along with item properties. Matrix factorization is one of the early realizations of latent factor models which has been very successful. Numerous extensions (variants) of the matrix factorization (MF) model have been proposed because of its wide adoption and simplicity (Cheng *et al*., 2018).

The work by Agarwal and Chen, (2010) utilized MF model to infer user ratings supported by item metadata. Their factorization through Latent Dirichlet Allocation (fLDA) model regularizes both latent factors and average latent topics from items' metadata. Their model predicts a user rating towards an item by modeling it as an affinity between user's features and item's topics which are associated by a bag of words for each item. Jiang et al., (2013) proposed two boosting frameworks to predict user ratings with the MF model as a base predictor. Specifically, they adapt adaBoost boosting approach for their frameworks which combines several recommenders which are controlled by different sample weights in predicting ratings. Other works (i.e., Agarwal & Chen, 2009; Li *et al*., 2013; Li, Xu, and Cao, 2016) have employed/extended latent factor model.

Latent factor models are very simple and convenient way for predicting ratings. However, they still have some pitfalls including being very sensitive to the level of sparsity in the dataset and cold-start problems (Pan, 2016; Çano and Morisio, 2017). The data sparsity and cold start challenges are usually severe because most available datasets are usually very sparse with numerous missing ratings and each user provides rating for a few items compared to the large number of existing items. To resolve the data sparsity, cold start issues and information limitation faced by rating prediction approaches researchers have considered utilization of additional information such as review texts (Tarus et al., 2017), user/item aspects topics (Wan & Niu, 2016; Tarus *et al*., 2018; Wan & Niu, 2018) and user/aspect sentiments.

Topic based models

Topic based models consider latent topics discussed in users' review texts to assist rating prediction. The discovered user/item topics are combined with latent factors from numerical ratings to enhance models' ability of understanding users' preferences. The approaches consider user review texts as a collection of separate documents where each document comprises topic distributions generated from word distribution of all the words in the corpus. Wang and Blei, (2011) utilized the LDA model to learn latent topics in order to infer scientific papers ratings. They created a latent space for topics and factors from numerical ratings and proposed a collaborative topic regression (CTR) model. Their proposed CTR model utilized maximum a posteriori (MAP) to perform parameter estimation necessary for optimizing predictions.

McAuley and Leskovec, (2013) proposed hidden factors and topics (HFT) model which combined hidden topics from review texts and hidden factors using a transformation function. The transformation function not only exponential linked the generated topics with latent factors but also addressed the weaknesses of both latent factors and topics. The HFT model utilized Gibbs sampler to perform parameter estimation. Ling, Lyu, and King, (2014) applied topic modeling methods on product reviews and associated them with latent factors enhance rating prediction. They used LDA to learn and extract item's aspect topics specific to each user and proposed ratings meet reviews (RMR) model which implements Gaussian mixtures to model infer unknown ratings. They argued that their method is capable of learning interpretable topics due to the fact that user-topic specific Gaussian distributions have clear interpretations which can describe how a user values the aspects denoted by each latent topic. The work by Liu *et al*., (2017) considered temporal dynamics of users' past numerical ratings together with review texts in predicting ratings. They focused on temporal drifts in review texts which enabled them to understand user preference variation according to time. They performed model learning by using EM algorithm with alternation between Gibbs sampler and Gradient descent.

Collecting implicit user data, cold start and data sparsity are still among the challenges which face recommender systems (Roy & Dutta, 2022). Nagy *et al*., (2021) argue that data sparsity problem requires researchers to find new ways to predict this missing data (ratings) through application and analysis of individual behavior and habits as well as collecting and analyzing implicit and suitable information such review topics. Different alternatives to address the challenges exist in the literature, however, combination/merging of two or more algorithms and utilization of alternative sources of information such as item aspects and review topics to compliment rating prediction approaches in RS have been most promising. This work focuses on harvesting the advantages of latent factor and topic-based models to form a hybrid model for rating prediction.

The main challenge with most topic based models is finding appropriate weight for each topic to use for rating prediction (Chen *et al*., 2017). This is expressed as a content and algorithmic challenge by Paraschakis, (2018) where it's difficult to quantify the topics content and define technical implementation of the generated topics. This study based on the basic topic model by McAuley and Leskovec, (2013) attempts to optimize the generated topics and combine with latent factor model for rating prediction.

Proposed approach

The proposed approach benefits from the capability of latent factor model for predicting ratings and latent dirichlet allocation (LDA) (Blei *et al*., 2001) for assigning topics to documents. The model addresses the challenges identified in the introduction section by optimizing the generated review text topics to correspond with user preferences. The proposed hybrid approach implements optimal semantic similarity matching (Rus & Lintean, 2012) which derives preference meaning of review topics by comparing them with each particular review text which explains individual user preferences. If the assigned topics semantically match the expressed preference meanings in user review text, then we assume that the topics correspond with specific user preferences.

The proposed approach identifies appropriate weights for review topics through assessing their preference knowledge contribution against expressed preferences in review texts based on the compositionality principle (Rus & Lintean, 2012) along with the sailor assignment problem (Dasgupta *et al*., 2009). The principle states that meaning of a sentence can be determined by composing meanings of its individual words while the sailor problem involves finding an optimal assignment of sailors to different navy positions (jobs) subject to resources and budget constraints. Likewise, our focus is to optimally match terms of the generated topics (sailors) to words of individual review texts (navy positions) based on WordNet word-to-word (Pedersen et al., 2004) and latent semantic analysis (Lintean, *et al*., 2010) similarity measures.

Matching assigned topics with each review texts sets appropriate weights for the topics which is referred to in this paper as optimized review topics (ORT) approach. Our proposed hybrid approach combines preference knowledge acquired from two models: latent factor model and topic-based model. The combination is realized through factorization of the optimized latent topics from user review texts and latent factors for users and items from past numerical ratings. Hence, the following subsections briefly present the latent factors, latent topics models, sailor assignment problem and then present the proposed hybrid model namely optimized review topics.

Rating prediction with latent factor models

Among successful recognitions of latent factor models are grounded on matrix factorization (Koren *et al*., 2009). Basic matrix factorization model assumes that user ratings towards items are characterized by their latent factors. Specifically, it reduces the dimension of the rating matrix by decomposing it to low feature matrices for user and item factors. This involves mapping users and items to a similar hidden factor space of size K. Hence, the model predicts the rating $\hat{r}_{ij}$ of a user u_i towards item v_j by performing a dot product of low feature matrices for user and item as follows:

$$\hat{r}_{ij} = \mu + b_i + b_j + u_i^T v_j \tag{1}$$

Where μ is the overall average rating, b_i and b_j are user and item bias while u_i and v_j represent K-dimensional user and item factors, respectively.

Given a rating matrix R with N users and M items, the objective is to predict missing ratings by choosing parameter values $\theta = \{\mu, b_i, b_j\}$ which minimizes the mean squared error (MSE). This can be achieved by optimizing the following objective:

$$L(\theta) = \sum_{i=1}^{N}\sum_{j=1}^{M}\left(r_{ij} - \hat{r}_{ij}\right)^2 + \lambda\left(\|u_i\|^2 + \|v_j\|^2 + b_i^2 + b_j^2\right) \tag{2}$$

Where r_{ij} is the actual/observed rating, $\hat{r}_{ij}$ is the predicted rating, λ is a weight which ensures that the parameters are small. The objective function can be optimized by stochastic gradient descent (SGD), alternating least square (ALS) mechanisms or others (Koren & Bell, 2011).

Latent dirichlet allocation for topic assignment

We considered each user review text towards an item as a document and followed the latent dirichlet allocation (LDA) model by Blei, Ng, and Jordan, (2003) which associates each document with different topics depending on their word probabilities. The LDA model associates each document with certain topics and each topic with certain words.

Specifically, the model initial creates a document-term matrix which consists unique terms and documents. The matrix elements are word frequencies in each document i.e. the matrix shows how frequent each word occurs in each document. Similar to latent factor models, the LDA model also assumes that latent topics associate documents and terms.

The generative LDA model links each document $d \in D$ with topic distribution θ_d with k latent topics and finds the probability of words in each document that discuss each of the k topics. That is each word $w = \langle w_1, w_2, \dots, w_{N_d} \rangle$ in review text d (with each review text having N_d words) is assigned a topic $z_{d,j}$ following the topic distribution $\theta_{z_{d,j}}$. The objective is to find topic distribution of each review text through word distribution of each topic. Final results of the model are word probabilities for each topic, topic distribution for each review text and topic assignment for each word.

Given D review texts, LDA approximates the likelihood of such text as follows:

$$p(D|\theta, \phi, z) = \prod_{d \in D} \prod_{j=1}^{N_d} \theta_{z_{d,j}} \phi_{z_{d,j}, w_{d,j}} \tag{3}$$

Where ϕ_k is the word distribution for each topic and θ_d is the topic distribution for each review text i.e. $\theta_{z_{d,j}}$ and $\phi_{z_{d,j}, w_{d,j}}$ represent probabilities of observing such specific topics and words, respectively.

Optimum assignment problem

The optimal matching of the generated topics focuses on finding maximum weights that correctly match the words in each review text with the words in each topic by using a weighted bipartite graph. Given a review text X with n words i.e., $X = \{x_1, x_2, \dots, x_n\}$ and a particular topic Y assuming it has same number of words n then $Y = \{y_1, y_2, \dots, y_n\}$. The complete weighted bipartite graph (Rus & Lintean, 2012) can be defined as follows $G = X \cup Y; X \times Y$ where the edge xy constitute a weight $w(xy)$ which is a word to word similarity between review text word x and topic word y. The semantic similarity of words is assumed to assess the preference meaning and relatedness of the topic words to the true/expressed preference of the review text.

The optimal assignment problem can then be specified as identifying a permutation π of $\{1,2, \dots, n\}$ for which $\sum_{i=1}^{n} w\left(x_i y_{\pi(i)}\right)$ is maximum. With matrix notation the assignment problem can be generally defined with a qualification matrix whose rows represents topics (sailors) and columns represent review texts (jobs/navy positions) and the matrix elements constitutes the weights between the topics and review texts (Kuhn, 2010). Considering n topics $(i = 1,2, \dots, n)$ have been generated from n review texts $(j = 1,2, \dots, n)$ and a rating (weighting) matrix $R = \left(w_{ij}\right)$ where are positive integers for all i and j. The assignment problem then tries to find a combination that satisfies a choice of one navy position j_i for each sailor (review text) i such that no position is assigned to two different people.

The full algorithm for the assignment problem is described in the work of Kuhn, (2010) and we follow its modified version for text optimization as proposed by Rus and

University of Dar es Salaam Library Journal
Vol 17, No 1 (2022), pp 38-53
ISSN: 0856-1818

Lintean (2012). Accordingly, we optimally match the words of each topic with the words of individual review texts for the set of topics that the LDA algorithm has assigned each review text. For instance, if the LDA algorithm assigns topic 2, 3 and 5 to the first review text then we optimally match the words of each topic with review text words. The resulting optimum topics will be used for rating prediction.

Optimized review topics model

The architecture of the proposed optimized review topics (ORT) model which comprises three steps: 1) topic modeling with LDA; 2) topic optimization based on the sailor assignment problem; 3) rating prediction using numerical ratings along with the optimized review topics. Figure 1 presents the optimized review topics model:

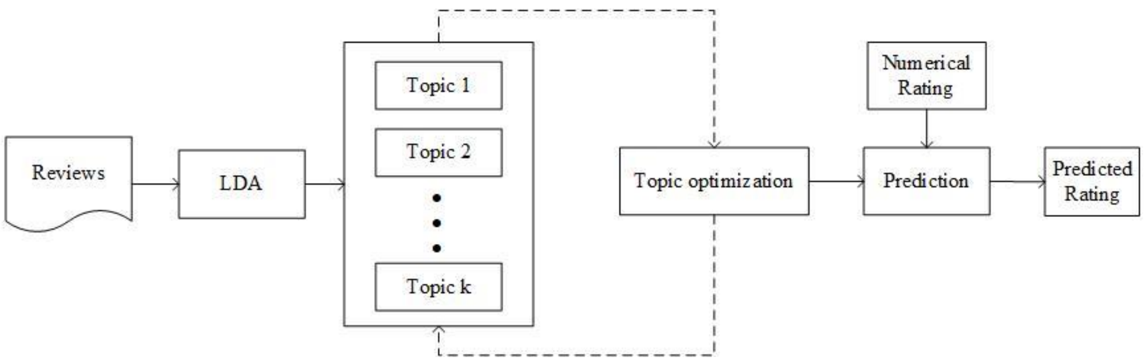

Figure 1: The architecture of the proposed model

The proposed hybrid model combines the optimized topics from the optimum assignment problem section with the latent factors from the rating prediction with latent factor section in order to make more accurate rating prediction. The combined loss function of the proposed approach constitutes the likelihood of the optimized topics from the review texts along with the error from the latent factor model. The loss is specified as follows:

$$L(R, D|\gamma, \theta, \phi, z) = \sum_{i=1}^{N} \sum_{j=1}^{M} \left(r_{ij} - \hat{r}_{ij}\right)^2 - \lambda_c \log p(D|\theta, \phi, z) \qquad (4)$$

Where R is the set of numerical ratings, $\gamma = \{\mu, b_i, b_j, u_i, v_j\}$ and λ_c is a hyper-parameter (control parameter) that harmonizes the performance of latent factors and review topics. We define a transformation to link item latent factors v_j with review topics θ_j similar to the HFT model by McAuley and Leskovec, (2013) as follows:

$$\theta_{d,k}^{u,v} = \frac{exp(Kv_{j,k})}{\sum_{k'} exp(Kv_{j,k'})} \quad (5)$$

Where $\theta_{d,k}^{u,v}$ represents the kth dimension for the topic distribution of review text d for item v_j by user u_i.
The generative prediction can be expressed with the following conditional probability:

$$p(d|u_i, v_j) = \prod_{j \in N_d^{u,v}} \sum_k \theta_{d,k}^{u,v} \beta_{k,w_j} \quad (6)$$

Where β_{k,w_j} represents kth dimension for the word distribution of word w_j.
In our proposed model, we have defined a threshold value 0.5 for the optimized topics as the final results from the integer weights a normalized to a range between 0 and 1, such that the sum of all weights is less or equal to 1. Specifically, the topic with weight above the threshold (0.5) is taken as a sole representative for the review text topics otherwise with weights less than the threshold as set of topics (two or three topics with relatively same weights close to 0.5 will be chosen) is considered to represent the review text.

Methodology

A number of experiments were conducted to evaluate the prediction performance of the proposed approach. In this section, we describe the dataset used, experimental design and baseline methods.

Dataset

Amazon products dataset developed by McAuley and Leskovec, (2013) was used to assess the proposed model performance in predicting missing ratings. The dataset comprises huge volume of product reviews spanning from May 1995 to March 2013 of approximately 9.5 million users on 2.4 million products. From the dataset we randomly selected nine different datasets. The statistics of the selected datasets are provided in Table 1. The Amazon product datasets have been experimented not only by the baseline methods described in the baseline methods and evaluation metrics section but also by many other models for different evaluation purposes.

Table 1: Dataset statistics

Dataset	**# users**	**# items**	**# ratings**	**# reviews**	**# tokens**	**#unique tokens (after filtering)**
Automotive	133,221	47,528	187,473	187,473	49,216	2,462
Gourmet Food	112,527	23,195	151,254	151,254	58,520	6,469
Musical Instruments	66,994	14,109	84,261	84,261	34,200	2,620
Amazon Instant Video	312,925	21,906	667,126	667,126	86,280	9,467
Office Products	110,468	14,098	132,258	132,258	48,308	5,536

Pet Supplies	160,491	17,483	216,476	216,476	60,270	6,422
Tools & Home Improv.	283,504	50,734	403,188	403,188	77,058	8,556
Toys and Games	290,710	52,214	390,597	390,597	68,719	7,096
Beauty	167,719	28,791	248,502	248,502	65,834	6,445

Experimental setup

The experimental setup involved two steps. First step was preprocessing the dataset to improve its quality and facilitate efficient machine learning. Second step was setting model parameters to ensure a similar setup with alternatives methods for smooth evaluation. The preprocessing conducted and parameter initialization is as follows.

Data preprocessing

We carried out dataset preprocessing not only as a way to remove noisy terms from review texts but also as a necessary step for effective topic modeling. The preprocessing involved *tokenization* – converting the review texts to their atomic elements, *stop word removal* – eliminating meaningless words, and *stemming* – merging words have similar meaning or lexical syntax. We also implemented an arbitrary filtering which involved ignoring terms which appeared in less than 30 review texts or in more than 15% of the total number of review texts. The number of unique tokens for each dataset prior and after filtering are as presented in table 2.

Parameter setting

Similar to baseline models, we set the dimension for latent factors to 5 (D=5) and set the predefined number of topics k = 5 and the hyper-parameters $\alpha = 0.1, \beta = 0.02, \mu_o = 0, \sigma_o^2 = 1$

Baseline methods and evaluation metric

The results of the proposed approach were compared with commonly used state of the art latent factor and topic models. The following baseline models were considered for evaluation:

Collaborative topic regression (**CTR**) (Wang & Blei, 2011) as evaluated by Ling, Lyu, and King, (2014)– which combined review texts topics with latent factors from numerical ratings in collaborative manner to from a collaborative topic regression model. CTR analyzes the latent topic space to explain both the observed ratings and the observed topics for rating prediction. Whereas:

Hidden factors as topics (**HFT**) (McAuley & Leskovec, 2013) – uses the generated review topics to regularize latent factors in the basic latent factor model.

Ratings meet reviews (**RMR**) (Ling *et al*., 2014) – learned item's aspect topics specific to each user and generated a mixture of Gaussians to infer ratings. They argued that the user-topic specific Gaussian distributions which facilitate interpretable topics.

Prediction model by Liu *et al*., (2017) referred in this paper as **LLSL** model – learned review topics with temporal dynamics of users' and items latent features. They used the review topics to regularize basic MF model.

The selected baseline methods have all reported improved performance over basic matrix factorization model described by the work of Salakhutdinov and Mnih, (2007). The work by Ling *et al*., (2014) also reported enhanced performance compared to LDAMF proposed by McAuley and Leskovec, (2013) while the work of Liu *et al*., (2017) demonstrated rating prediction improvement compared to BPTF (Xiong *et al*., 2010), TimeSVD++ (Koren, 2010) and JMARS (Diao *et al*., 2014). The prediction performances of the proposed approach as well as baseline methods were evaluated with the widely adopted mean squared error (MSE) metric. The error is a well-known measure that calculates the differences between actual and predicted ratings.

Results

Experimental results of the proposed approach reveal performance improvements achieved by optimized review topics. From the results presented in Table 2, proposed ORT approach outperforms baseline models by an overall average of 12.02% (the average of percent improvement of ORT versus each baseline). This improvement exceeds the 10% average improvement set by $1M Netflix prize (Bell & Koren, 2008) whose goal was to increase the accuracy of the Netflix's rating prediction approach by 10%. Our ORT approach demonstrates very high performance compared to the CTR, HFT and RMR models in almost all datasets accept the Beauty dataset. The proposed model also performs better than the LLSL model with an average performance of 10.76% for all datasets.

From our experiments, we analyzed the effect of varying important parameters like the number of latent factors K, control parameterλ_c and topic model hyper-parameters. To test the model, we fixed the topic model hyper-parameters first and varied either the latent factors or control parameter. With testing the ORT model produced relatively same results which implied that the proposed model is not highly affected by the number of latent factors. Similar results were obtained by varying the control parameter. This shows that the ORT model is stable to such variation and we can conclude that the optimized topics play a vital role in predicting missing ratings.

Afterwards, we varied the hyper-parameters and fixed the latent model parameters (K, λ_c). Varying the hyper-parameters did not result into performance improvement of the proposed model compared to the best results which were obtained with the hyper-parameter setting presented in section 4.2.2. The task of fine tuning of the hyper-parameters is generally left to human experts as the hyper-parameters are neither trained nor derived

University of Dar es Salaam Library Journal
Vol 17, No 1 (2022), pp 38-53
ISSN: 0856-1818

from the datasets. Apart from existing optimization algorithms, researchers usually do manual tuning of the hyper-parameters and report best results.

Table 2: MSE for the Evaluated models (best results are highlighted)

Dataset	CTR a	HFT b	RMR c	LLSL d	ORT e	% imp e vs a	% imp e vs b	% imp e vs c	% imp e vs d
Musical Instr.	1.422	1.395	1.374	1.407	**1.252**	11.95%	10.25%	8.88%	11.01%
Gourmet Food	1.482	1.457	1.465	1.452	**1.288**	13.09%	11.60%	12.08%	11.29%
Office Products	1.733	1.669	1.638	1.6065	**1.185**	31.62%	28.99%	27.65%	26.23%
Automotive	1.492	1.432	1.403	1.4205	**1.183**	20.71%	17.38%	15.68%	16.71%
Pet Supplies	1.613	1.583	1.562	1.563	**1.550**	3.90%	2.08%	0.76%	0.83%
A. I. Video	1.291	1.260	1.270	**1.1854**	1.213	6.04%	3.73%	4.48%	-2.32%
Beauty	1.361	1.358	**1.334**	1.349	1.413	-3.82%	-4.05%	-5.92%	-4.74%
Tools & Home	1.513	1.510	1.491	1.510	**1.475**	2.51%	2.31%	1.07%	2.31%
Toys & Games	1.389	1.370	1.372	1.332	**0.859**	38.15%	37.29%	37.39%	35.51%
Average Impr.						13.79%	12.17%	11.34%	10.76%

Tables 3, 4 and 5 show top 5 words for each topic in the Automotive, Office products and Gourmet food datasets. From the generated topics we can easily explain the predicted ratings, taking the topics in Automotive dataset as an example: topic 1 can be interpreted as talking about engine servicing and performance; topic 2 is about car washing; topic 3 is concerned with power accessories; topic 4 is mostly about washing/cleaning accessories; while topic 5 is concerned with accessory fitting/installation. Additionally, topic 3 and 5 in Gourmet food dataset illustrated in Table 5 are concerned about food ingredients/materials and food recipe/cooking. Hence, given the cold-start and data sparsity scenarios, we can confidently predict user interests or recommend items based on the learned topics. For instance, if we are sure that a user cares about the content food ingredient such as sugar and fat or he likes certain food recipe then we can recommend food products which best match the two topics. Such additional preference knowledge helps to gain more insights on users and items which is crucial for more accurate rating prediction.

Table 3: Top 5 words in Automotive dataset

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
oil	towel	battery	blade	light
filter	Clean	power	wiper	install
price	Dry	charge	hose	fit
change	Wash	device	water	bulb
engine	Wax	light	windshield	tire

Table 4: Top 5 words in Office Products dataset

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
pen	Desk	scan	tape	label
binder	Chair	set	box	laminate
write	Stapler	epson	note	sheet
pencil	Staple	photo	post	Folder
erase	Pad	cartridge	sticker	File

Table 5: Top 5 words in Gourmet food dataset

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
coffee	Popcorn	sugar	chocolate	sauce
Tea	Oil	calori	bag	cook
cup	Coconut	ingredi	eat	pasta
drink	Salt	bar	package	rice
water	Butter	fat	snack	season

Discussion

The success of the proposed model assumes that matching latent topics with expressed user preferences will enhance prediction performance of the model. Additionally, the optimized topics provide reliable explanations for predicted ratings which is very important for effective model understanding as well as assisting stakeholders in taking appropriate actions to support user experience and utilization of the online systems.

One of core task for rating prediction approaches employed by RSs is to ensure accurate prediction as noted in Netflix challenge (Bell & Koren, 2008) and numerous recommender systems challenge such as RecSys Challenge 2021 and 2021 (Belli *et al*., 2021). This reveals the importance of the proposed ORT model whose prediction accuracy improvement suggests massive difference in the associated recommender system. Furthermore, the proposed model addresses several concerns such as generation of explainable recommendations, superiority of preference relation-based matrix factorization algorithms, and data sparsity as described by (Kuanr & Mohapatra, 2021). Hence, the proposed model architecture and combination of later factor and topic-based models reveal the effectiveness and interpretation of the represented semantics behind user preferences as extracted from the review texts and applied in rating prediction to assist recommendation task.

Experimental results suggest that optimizing review texts topics results to more accurate prediction as the proposed ORT model outperformed the evaluated topic based models by McAuley and Leskovec, (2013), C. Wang and Blei, (2011), Ling *et al*., (2014) and Liu *et al*., (2017). The optimized topics is assumed capture relevant preference for each user and hence provide appropriate weight in predicting ratings in a collaborative filtering setting. The effectiveness of the proposed ORT model also corresponds to specified objectives as outlined in the introduction section.

Conclusion

In this paper, a hybrid approach that combines preference knowledge from latent factor models and topic-based models is presented. Specifically, the proposed approach performs factorization of optimized topics obtained through review texts analysis and latent factors for users and items past numerical ratings. The proposed approach optimizes the topics to fit expressed user preferences by conducting optimal semantic similarity matching based on compositionality principle and sailor assignment problem. The matching provides appropriate weights for the topics according to the preference meaning or relatedness of the words in review texts. Experimental results verify that our proposed model not only outperforms alternative models but also provides reliable explanations of the predicted ratings through the generated optimized topics. The proposed ORT model can be used in different recommender systems to assist users in choosing/finding most preferred items / best products according to their preference. It can be used in various fields such as e-commerce sites to help consumers select the right items and can act as a vital company/firm tool for transforming the e-commerce environment. It could also be applied to customize RSs in agriculture, health and education. One limitation the proposed model is that it can't be applied when the dataset contains only numerical ratings with no review texts. Also, the model is implemented under the assumption that all the numerical ratings of users towards items are independently and identically distributed (iid) which is rarely satisfied in practice. Future work will focus on exploring more effective ways of interpreting and matching the topics to properly align with users' preferences and develop appropriate model for predicting ratings.

References

Agarwal, D., & Chen, B. (2009). Regression-based latent factor models. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 19–27. https://doi.org/10.1145/1557019.1557029

Agarwal, D., & Chen, B. (2010). fLDA : Matrix factorization through latent dirichlet allocation. *Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010*, 91–100. https://doi.org/10.1145/1718487.1718499

Bao, Y., Hui, F., & Zhang, J. (2014). TopicMF: Simultaneously exploiting ratings and reviews for recommendation. *Proceedings of the Twenty-Eighth AAAI Conference on*

Artificial Intelligence, 2–8. http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8273

Bell, R. M., & Koren, Y. (2008). Lessons from the Netflix prize challenge. *ACM SIGKDD Explorations*, 9(2), 75-79. https://doi.org/10.1145/1345448.1345465

Belli, L., Tejani, A., Portman, F., Lung-Yut-Fong, A., Chamberlain, B., (…) & Shi, W. (2021). *The 2021 RecSys challenge dataset: Fairness is not optional*. 1–6. https://doi.org/10.1145/3487572.3487573

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2001). Latent dirichlet allocation. *Advances in Neural Information Processing Systems 14 Neural Information Processing Systems: Natural and Synthetic, NIPS 2001*, 601-608. http://papers.nips.cc/paper/2070-latent-dirichlet-allocation

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993-1022. http://jmlr.org/papers/v3/blei03a.html

Çano, E., & Morisio, M. (2017). Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*, 21(6), 1487–1524. https://doi.org/10.3233/IDA-16320

Chen, J., Zhang, C., & Niu, Z. (2016). Identifying helpful online reviews with word embedding features. *Knowledge Science, Engineering and Management, 9th International Conference (KSEM) 2016, Passau, Germany*, 123–133. https://doi.org/10.1007/978-3-319-47650-6_10

Chen, Y., Zhang, L., Li, X., Zong, Y., & Liu, G. (2017). Incorporating user preferences across multiple topics into collaborative filtering for personalized merchant recommendation. *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data, 3*, 575–590. https://doi.org/10.1007/978-3-319-63579-8_44

Cheng, Z., Ding, Y., Zhu, L., & Kankanhalli, M. S. (2018). Aspect-aware latent factor model : rating prediction with ratings and reviews. *Proceedings of the 2018 World Wide Web Conference on World Wide Web,WWW 2018*, 639–648. https://doi.org/10.1145/3178876.3186145

Dasgupta, D., Nino, F., Garrett, D., Chaudhuri, K., Medapati, S., Kaushal, A., & Simien, J. (2009). A multiobjective evolutionary algorithm for the task based sailor assignment problem. *Proceedings of Genetic and Evolutionary Computation Conference, GECCO 2009*, 1475--1482. https://doi.org/10.1145/1569901.1570099

Diao, Q., Qiu, M., Wu, C.-Y., Smola, A. J., Jiang, J., & Wang, C. (2014). Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14*, 193–202. https://doi.org/10.1145/2623330.2623758

Jiang, X., Niu, Z., Mustafa, G., Lin, Z., Chen, B., & Zhou, Q. (2013). Novel Boosting Frameworks to Improve the Performance of Collaborative Filtering. *Asian Conference on Machine Learning, ACML 2013,* 87–99. http://jmlr.org/proceedings/papers/v29/Jiang13.html

Koren, Y. (2010). Collaborative filtering with temporal dynamics. *Communications of the ACM, 53*(4), 89-97. https://doi.org/10.1145/1721654.1721677

Koren, Y., & Bell, R. M. (2011). Advances in collaborative filtering. In: Ricci, F., Rokach, L., Shapira, B. (eds) *Recommender Systems Handbook.* Springer. https://doi.org/10.1007/978-1-4899-7637-6_3

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8), 30–37. https://doi.org/10.1109/MC.2009.263

Kuanr, M., & Mohapatra, P. (2021). Recent challenges in recommender systems: A survey. advances in intelligent systems and computing, 1199(May), 353–365. https://doi.org/10.1007/978-981-15-6353-9_32

Kuhn, H. W. (2010). The Hungarian method for the assignment problem. In *50 Years of Integer Programming 1958-2008 - From the Early Years to the State-of-the-Art* (pp. 29-47). Springer. https://doi.org/10.1007/978-3-540-68279-0_2

Li, F., Xu, G., & Cao, L. (2016). Two-level matrix factorization for recommender systems. *Neural Computing and Applications*, 27(8), 2267–2278. https://doi.org/10.1007/s00521-015-2060-3

Li, F., Xu, G., Cao, L., Fan, X., & Niu, Z. (2013). CGMF: Coupled group-based matrix factorization for recommender system. *Web Information Systems Engineering - WISE 2013 - 14th International Conference*, 189-198. https://doi.org/10.1007/978-3-642-41230-1_16

Ling, G., Lyu, M. R., & King, I. (2014). Ratings meet reviews, a combined approach to recommend. *Proceedings of the 8th ACM Conference on Recommender Systems - RecSys'14*, 105–112. https://doi.org/10.1145/2645710.2645728

Lintean, M., Moldovan, C., Rus, V., & Mcnamara, D. (2010). The role of local and global weighting in assessing the semantic similarity of texts using latent semantic analysis. *Proceedings of the Twenty-Third International Florida Artificial Intelligence Research Society Conference,* 235–240. http://www.aaai.org/ocs/index.php/FLAIRS/2010/paper/view/1310

Liu, Y., Liu, Y., Shen, Y., & Li, K. (2017). Recommendation in a changing world: Exploiting temporal dynamics in ratings and reviews. *ACM Transactions on the Web*, 12(1), 1–20. https://doi.org/10.1145/3108238

Liu, Y., & Shen, Y. (2018). Personal tastes vs. fashion trends: Predicting ratings based on visual appearances and reviews. *IEEE Access*, *6*, 16655–16664. https://doi.org/10.1109/ACCESS.2018.2811463

McAuley, J., & Leskovec, J. (2013). Hidden factors and hidden topics:Understanding rating dimensions with review text. *Proceedings of the 7th ACM Conference on Recommender Systems - RecSys' 13*, 165–172. https://doi.org/10.1145/2507157.2507163

Nagy, F., Haroun, A., Abdel-Kader, H., & Keshk, A. (2021). A review for recommender system models and deep learning. *IJCI. International Journal of Computers and Information*, 8(2), 170–176. https://doi.org/10.21608/ijci.2021.207864

Pan, W. (2016). A survey of transfer learning for collaborative recommendation with auxiliary data. *Neurocomputing*, *177*, 447–453. https://doi.org/10.1016/j.neucom.2015.11.059

Paraschakis, D. (2018). *Algorithmic and ethical aspects of recommender systems in e-commerce* [Licentiate thesis**,** Malmö university]. https://doi.org/https://doi.org/10.24834/2043/24268

Pedersen, T., Patwardhan, S., & Michelizzi, J. (2004). WordNet:: similarity - Measuring the relatedness of concepts. *Proceedings of the Nineteenth National Conference on*

Artificial Intelligence, 1024-1025. http://www.aaai.org/Library/AAAI/2004/aaai04-160.php

Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, 9(1). https://doi.org/10.1186/s40537-022-00592-5

Rus, V., & Lintean, M. (2012). A comparison of greedy and optimal assessment of natural language student input using word-to-word similarity metrics. *Proceedings of the seventh workshop on building educational applications using nlp*, 157–162. http://www.aclweb.org/anthology/W12-2018

Salakhutdinov, R., & Mnih, A. (2007). Probabilistic matrix factorization. *Proceedings of the advances in neural information processing systems 20 (NIPS 07)*, 1257–1264. https://doi.org/10.1145/1390156.1390267

Tarus, J. K., Niu, Z., & Kalui, D. (2018). A hybrid recommender system for e-learning based on context awareness and sequential pattern mining. *Soft Computing*, 22(8), 2449--2461. https://doi.org/10.1007/s00500-017-2720-6

Tarus, J. K., Niu, Z., & Yousif, A. (2017). A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining. *Future Generation Computer Systems*, *72*, 37–48. https://doi.org/10.1016/j.future.2017.02.049

Wan, S., & Niu, Z. (2016). A learner oriented learning recommendation approach based on mixed concept mapping and immune algorithm. *Knowledge-Based Systems*, *103*, 28–40. https://doi.org/10.1016/j.knosys.2016.03.022

Wan, S., & Niu, Z. (2018). An e-learning recommendation approach based on the self-organization of learning resource. *Knowledge-Based Systems*, *160*, 71--87. https://doi.org/10.1016/j.knosys.2018.06.014

Wang, C., & Blei, D. (2011). Collaborative topic modeling for recommending scientific articles. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '11*, 448-456. https://doi.org/10.1145/2020408.2020480

Wang, H., & Ester, M. (2014). A sentiment-aligned topic model for product aspect rating prediction. *Proceedings of the 2014 conference on empirical methods in natural language processing, EMNLP 2014*, 1192–1202. http://aclweb.org/anthology/D/D14/D14-1126.pdf

Xiong, L., Chen, X., Huang, T.-K., Schneider, J., & Carbonell, J. G. (2010). Temporal collaborative filtering with Bayesian Probabilistic Tensor Factorization. *Proceedings of the SIAM international conference on data mining,SDM 2010*, 211–222. https://doi.org/10.1137/1.9781611972801.19

Xu, Y., Lam, W., & Lin, T. (2014). Collaborative filtering incorporating review text and co-clusters of hidden user communities and item groups. *Proceedings of the 23rd ACM international conference on information and knowledge management, CIKM 2014*, 251–260. https://doi.org/10.1145/2661829.2662059

Zhao, X., Niu, Z., Chen, W., Shi, C., Niu, K., & Liu, D. (2015). A hybrid approach of topic model and matrix factorization based on two-step recommendation framework. *Journal of Intelligent Information Systems,* 44(3), 335–353. https://doi.org/10.1007/s10844-014-0334-3