Matrix Factorisation with Word Embeddings for Rating Prediction on Location-Based Social Networks

Jarana Manotumruksa[1], Craig Macdonald[2], and Iadh Ounis[2]

[1]j.monotumruksa.1@research.gla.ac.uk,[2]{firstname.secondname}@glasgow.ac.uk
University of Glasgow, UK

Abstract. With vast amounts of data being created on location-based social networks (LBSNs) such as Yelp and Foursquare, making effective personalised suggestions to users is an essential functionality. Matrix Factorisation (MF) is a collaborative filtering-based approach that is widely used to generate suggestions relevant to user's preferences. In this paper, we address the problem of predicting the rating that users give to venues they visit. Previous works have proposed MF-based approaches that consider auxiliary information (e.g. social information and users' comments on venues) to improve the accuracy of rating predictions. Such approaches leverage the users' friends' preferences, extracted from either ratings or comments, to *regularise* the complexity of MF-based models and to avoid over-fitting. However, social information may not be available, e.g. due to privacy concerns. To overcome this limitation, in this paper, we propose a novel MF-based approach that exploits word embeddings to effectively model users' preferences and the characteristics of venues from the textual content of comments left by users, regardless of their relationship. Experiments conducted on a large dataset of LBSN ratings demonstrate the effectiveness of our proposed approach compared to various state-of-the-art rating prediction approaches.

1 Introduction

In recent years, location-based social networks (LBSNs) such as Yelp and Foursquare have emerged as popular platforms that allow users to search for Point-of-Interest and post ratings as well as their opinions/comments about venues they have visited. This makes LBSN data very suitable for making recommendations of venues for users to visit. Matrix Factorisation (MF) – based on collaborative filtering – is a popular technique used to effectively recommend items to users by assuming that users who share similar preferences (rating positively or negatively the same items) are likely to prefer similar items [1]. Previous works [2, 3] proposed MF-based approaches that leverage such users' explicit feedback (e.g. ratings) to model their preferences, and thereby effectively suggest new venues for users to visit. However, rating data is sparse in nature, i.e. users/venues have very few ratings, hindering the quality of venue suggestions. To alleviate the sparsity problem, various MF-based approaches [3, 4] have been proposed to consider auxiliary information such as the ratings of each user's friends to effectively predict the user's ratings. In particular, Ma *et al.* [4] proposed a social-based regularisation technique that *regularises* the complexity of a MF model, by assuming that users are likely to be influenced by their friends who rate similar venues with similar scores.

Apart from the social information, the comments associated with ratings on venues left by users can provide insights about why they rated a given venue positively or negatively, while also reflecting characteristics of each venue. Previous works [2, 3, 5, 6] have shown that the textual content of comments can be leveraged to effectively model user's preferences and characteristics of venues. Recently, word embeddings are being increasingly applied in many applications due to their effectiveness in capturing semantic properties of textual content, such as text classification [7] as well as recommendation system [3, 8]. In particular, Musto *et al.* [8] apply several word embedding techniques to enhance the effectiveness of content-based collaborative approaches for tasks such as book and movie recommendation. Moreover, Manotumruksa *et al.* [3] extended the regularisation technique proposed by Ma *et al.* [4], by exploiting word embeddings to estimate the similarity between friends from their comments of venues they both visited. Their assumption is that users are not only influenced by friends who like/dislike similar venues but are also influenced by friends who share similar tastes, which can be extracted from the explicit textual feedback in the form of comments they have left on venues. Unlike previous works mentioned above [2–4], this paper contributes a novel MF-based approach that jointly models user's preferences and characteristic of venues from the textual content of comments to effectively predict user-venue ratings. Experiments on a large real-word dataset demonstrate the effectiveness of our proposed approach in comparison with various state-of-the-art user-venue rating prediction approaches. Our proposed approach is as effective as state-of-the-art rating prediction approaches that consider social information and textual content of comments.

2 Joint Linear Combinations of Matrix Factorisations

In this section, we first describe the problem of predicting user-venue rating and detail traditional Matrix Factorisation (MF) techniques. Next, we explain how we exploit word embeddings to enable MF to consider textual content of users' comments.

2.1 Traditional Matrix Factorisation

We formally describe the notations and the problem of predicting user's rating on venues. First, user-venue ratings are represented as a matrix $R \in \mathbb{R}^{m \times n}$ where m and n are the number of users and venues. Let $r_{i,j}$ and $c_{i,j}$ denote the 1-5 scale rating and textual content of comment of user i on venue j, respectively. Using the historical ratings R_i of user i, the task is then to predict the rating this user would give to venue j.

Matrix Factorisation (MF) is a collaborative filtering technique that assumes that users who share similar preferences (e.g. visiting similar venues and rating these venues similarly) are likely to influence each other [1]. In particular, the goal of MF is to reconstruct the rating matrix R by calculating the dot product of latent factors of users $U \in \mathbb{R}^{m \times d}$ and venues $V \in \mathbb{R}^{n \times d}$ where d is the number of latent dimensions:

$$R \approx \hat{R} = U^T V \tag{1}$$

The MF model is trained by minimising a loss function L, which consists of sum-of-squared-error terms between the actual ratings and predicted ratings, as follows:

$$L(U, V) = \underset{U,V}{min} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{i,j} \cdot (r_{i,j} - \hat{r}_{i,j})^2 + \frac{\lambda}{2} (\|U\|_F^2 + \|V\|_F^2) \tag{2}$$

where λ is a regularisation parameter and $\|.\|_F^2$ denotes the Frobenius norm, used to avoid overfitting. $I_{i,j}$ is an indicator, which gives 1 if user i rated venue j, otherwise 0.

2.2 Combining Matrix Factorisations with Word Embeddings

As mentioned in Section 1, explicit feedback by users in comments can provide insights into why users rate a venue positively or negatively and also reflect the characteristics of the venue. Unlike previous works [2, 3, 5], we propose a MF approach that leverages the textual content of comments by other users, regardless of their relationship, to alleviate the sparsity problem. A straightforward approach would be to represent comments using a bag-of-words approach, however this would not consider the context in which terms occur, and hence not model the semantic properties of comments. Instead, we follow [6, 8], by exploiting word embeddings to semantically model the user's preferences $S_u \in \mathbb{R}^{m \times k}$ and characteristics of venue $S_v \in \mathbb{R}^{n \times k}$ from the comments in a low-dimensional space, where k is the number of word embedding dimensions, as follows:

$$S_{u_i} = \sum_{c_{i,j} \in C_{u_i}} \sum_{t \in c_{i,j}} w2v(t) \times r_{i,j} \qquad S_{v_j} = \sum_{c_{i,j} \in C_{v_j}} \sum_{t \in c_{i,j}} w2v(t) \times r_{i,j} \tag{3}$$

where C_{u_i} and C_{v_j} are the sets of user i's and venue j's comments and $w2v(t) \in \mathbb{R}^k$ is a function that returns a word embedding representation of term t. Note that the $w2v()$ function in Equation (3) can be replaced with a word representation generated by more complex convolutional or recurrent neural networks (e.g. [7]). However, we consider other formulations beyond the scope of this paper and leave these as future work.

We note that the MF-based approach proposed by Hu *et al.* [2] leverages the comments to decompose latent factors of venues V into a combination of latent factors of comment terms, which can alleviate the sparsity problem for venues that have few ratings. However, their proposed approach lacks flexibility, because it requires the latent factors of comment's terms to be in the same space as the latent factors of venues V (i.e. the dimension d of these latent factors need to be equal). Instead, we argue that those two latent factors do not necessarily share the same space due to different nature of venues and comments. Intuitively, similar venues can be recognised by services provided by the venues, while similar comments can be recognised by terms appearing in the comments and their *semantics*. Therefore, the latent factors of venues and comments should not share the same dimensions, indeed the latent factors of comments should be larger due to the complexity of comments. Our work also differs from [3] in that comments are considered inherent to the matrix factorisations rather than the regularisation.

Hence – and unlike previous works [2, 3] – to incorporate the representations of users' preferences S_u and characteristics of venues S_v within a joint MF model, we modify Equation (1) to linearly combine matrix factorisations:

$$R \approx \hat{R} = \alpha U^T V + (1 - \alpha)(U_s^T S_u + V_s^T S_v) \tag{4}$$

where $U_s \in \mathbb{R}^{m \times k}$ and $V_s \in \mathbb{R}^{n \times k}$ are semantic latent factors of users and venues respectively and α is a parameter that controls the influence between the latent factors (U and V) and the semantic latent factors (U_s and V_s). To avoid overfitting, we regularise the model based on the complexity of the semantic latent factors, as follows:

$$L(U, V, U_s, V_s) = L(U, V) + \frac{\lambda}{2}(\|U_s\|_F^2 + \|V_s\|_F^2) \tag{5}$$

Table 1. Overview of user-rating prediction approaches.

Models	Social	Comments	Params	Intuitions
MF [1]	$\times$	$\times$	λ	Users are likely to prefer venues rated that other similar users rate highly.
VMF [2], JMF [5]	$\times$	✓	λ	+ Users are likely to prefer venues that share similar characteristics (according to textually similar comments).
SoReg [4]	✓	$\times$	λ, α	+ Users are likely to prefer venues that their friends rate highly.
BoWReg, DeepReg [3]	✓	✓	λ, α	+ Users are likely to prefer venues visited by their friends who have similar tastes.
MFw2v	$\times$	✓	λ, α	+ Users' preferences can be extracted from their comments on venues and users are likely to prefer venues that share similar characteristics.

Finally, we apply Stochastic Gradient Descent (SGD) to find a local minimum of the loss function (Equation (5)), by optimising each of the latent factor matrices U, V, U_s, V_s, while fixing the other, until convergence as follows:

$$\frac{\partial \underset{W2V}{L(U,V)}}{\partial U_i} = \sum_{j=1}^{n} I_{i,j}(r_{i,j} - \hat{r}_{i,j})\alpha V_j + \lambda U_i \qquad \frac{\partial \underset{W2V}{L(U,V)}}{\partial V_j} = \sum_{i=1}^{m} I_{i,j}(r_{i,j} - \hat{r}_{i,j})\alpha U_i + \lambda V_j$$

$$\frac{\partial \underset{W2V}{L(U,V)}}{\partial U_{s_i}} = \sum_{j=1}^{n} I_{i,j}(r_{i,j} - \hat{r}_{i,j})(1-\alpha)S_{v_j} + \lambda U_{s_i}$$

$$\frac{\partial \underset{W2V}{L(U,V)}}{\partial V_{s_j}} = \sum_{i=1}^{m} I_{i,j}(r_{i,j} - \hat{r}_{i,j})(1-\alpha)S_{u_i} + \lambda V_{s_j}$$

3 Evaluation

In this section, we evaluate the effectiveness of our proposed model $MFw2v$ by comparing with state-of-the-art rating prediction approaches. In particular, we aim to address the following research question **RQ**: Can we exploit word embeddings to effectively model user's preferences and characteristics of venues and improve the prediction accuracy of traditional MF-based approach?

3.1 Experimental Setup

We first describe the experimental setup used to evaluate the effectiveness of our proposed approach (MFw2v) and summarise baselines in details. We conducted experiments using the publicly available Yelp dataset[1], which consists of 2,225,214 ratings by 552,339 users for 77,079 venues. We conduct 5-fold cross-validation experiment where each fold has 60% training, 20% validation and 20% testing. We implement all experiments using LibRec [9], a Java library for recommendation systems. For each fold, the

[1] www.yelp.com/dataset_challenge

α in Equation (4) is determined using the validation set. Following [2, 5], we set the dimension of latent factors d to 10 and $\lambda = 0.001$. For word embeddings, we use the Word2Vec tool[2], to train a skip-gram model [10] using the default settings (window size 5 and word embedding dimensions $k = 100$) on the Yelp dataset. Previous work by Mikolov *et al.* [11] showed that the skip-gram model performs better than or equally to the CBOW model. The baselines used in this our experiments are summarised below, while their parameters and sources of evidence are highlighted in Table 1.

MF [1] is the traditional matrix factorisation approach, which only considers the user-venue matrix to predict the ratings (described in Section 2.1).

VMF [2] is a state-of-the-art bag-of-words based MF approach that considers geographical and textual information (i.e. comments about a venue). To permit a fair evaluation, we re-implement their approach to consider only textual information, and ignore the geographical location of venues, in common with our own proposed approach.

JMF [5] is a state-of-the-art rating prediction approach that jointly models comments and user's ratings by exploiting skip-thought vectors [12][3] to represent the content of comments. Instead of skip-thought vectors, we re-implement their approach to exploit word embeddings to permit a fair comparison with our proposed approach.

SoReg [4] is a social-based regularisation approach that enhances the rating prediction of MF-based approaches by assuming that friends who rate similar venues with similar scores can be influenced by each other.

BoWReg & **DeepReg** [3] are textual-social based regularisation approaches that leverage both social information and comments to reduce the complexity of MF-based approaches. In particular, DeepReg exploits word embeddings to estimate similarity between friends; BowReg is an orthogonal bag-of-words based regularisation approach.

Finally, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used to measure rating prediction accuracy (for both metrics, lower is better).

3.2 Experimental Results

Table 2 reports the rating prediction accuracy, in terms of MAE and RMSE, of our proposed MFw2v and other baselines. Firstly, we observe that our proposed approach, MFw2v, outperforms all MF baselines in terms of MAE and is comparable with DeepReg in terms of RMSE. In particular, comparing with the traditional MF (MF), the prediction accuracy of MFw2v is $\sim$12% more effective than MF for both MAE and RMSE. This implies that the users' preferences and the characteristics of venues extracted from the textual content of comments using word embeddings can enhance the rating prediction accuracy of traditional MF. Indeed, textual content of comments of venues are publicly available in LBSNs, while social information maybe not available for privacy reasons. Our proposed approach (MFw2v), which exploits comments of venues to model characteristics of venues, outperforms SoReg for both metrics (9.52% and 2.97% for MAE and RMSE respectively). Moreover, by comparing MFw2v and DeepReg, the experimental results demonstrate that MFw2v is as effective as the state-of-the-art rating prediction approach (DeepReg). Note that MFw2v only take users' comments into account, while DeepReg considers both social information and users' comments. Although the improvements in Table 2 are relatively small, Koren [1] pointed that small improvements in MAE and RMSE can lead to marked improvements in the quality of recommendations in practice.

[2] code.google.com/archive/p/word2vec [3] a state-of-the-art deep learning approach.

Table 2. Prediction accuracy in terms of MAE and RMSE of various approaches. Percentage differences of prediction accuracy are calculated with respect to the best performance achieved for that metric, which are highlighted in bold.

Metrics	MF	VMF	JMF	SoReg	BoWReg	DeepReg	MFw2v
MAE	1.1640	1.2198	1.1795	1.1260	1.1004	1.0781	**1.0188**
Δ	12.47%	16.48%	15.77%	9.52%	7.42%	5.50%	
RMSE	1.5243	1.5006	1.5073	1.3870	1.4354	**1.3456**	1.3458
Δ	11.72%	10.33%	12.02%	2.99%	6.26%		0.01%

4 Conclusion

This paper proposed a MF-based approach that leverages the textual content of comments to effectively model users' preferences and characteristic of venues, and exploits word embeddings to captures the semantic properties of comments. Our comprehensive experiments conducted using a large existing dataset (2.2M ratings from 500k users) demonstrate the effectiveness of our proposed approach in comparison with state-of-the-art rating prediction approaches. Indeed, our proposed approach (MFw2v) is shown to be as effective as state-of-the-art approaches, while only requires venues' comments as auxiliary information. For future work, we plan to apply various deep learning techniques to our proposed approach and explore the impact of word embedding parameters.

References

1. Koren, Y.: Factor in the neighbors: Scalable and accurate collaborative filtering. Transactions on Knowledge Discovery from Data (TKDD) **4**(1) (2010)
2. Hu, L., Sun, A., Liu, Y.: Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. In: Proc. of SIGIR. (2014)
3. Manotumruksa, J., Macdonald, C., Ounis, I.: Regularising factorised models for venue recommendation using friends and their comments. In: Proc. of CIKM. (2016)
4. Ma, H., Zhou, D., Liu, C., Lyu, M.R., King, I.: Recommender systems with social regularization. In: Proc. of WSDM. (2011)
5. Jin, Z., Li, Q., Zeng, D.D., Zhan, Y., Liu, R., Wang, L., Ma, H.: Jointly modeling review content and aspect ratings for review rating prediction. In: Proc. of SIGIR. (2016)
6. Manotumruksa, J., Macdonald, C., Ounis, I.: Modelling user preferences using word embeddings for context-aware venue recommendation. arXiv preprint arXiv:1606.07828 (2016)
7. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
8. Musto, C., Semeraro, G., de Gemmis, M., Lops, P.: Learning word embeddings from wikipedia for content-based recommender systems. In: Proc. of ECIR. (2016)
9. Guo, G., Zhang, J., Sun, Z., Yorke-Smith, N.: LibRec: A Java library for recommender systems. In: Proc. of UMAP. (2015)
10. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv:1301.3781 (2013)
11. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Proc. of NIPS. (2013)
12. Kiros, R., Zhu, Y., Salakhutdinov, R.R., Zemel, R., Urtasun, R., Torralba, A., Fidler, S.: Skip-thought vectors. In: Proc. of NIPS. (2015)