# Regularising Factorised Models for Venue Recommendation using Friends and their Comments

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

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

## ABSTRACT

Venue recommendation is an important capability of Location-Based Social Networks such as Yelp and Foursquare. Matrix Factorisation (MF) is a collaborative filtering-based approach that can effectively recommend venues that are relevant to the users' preferences, by training upon either implicit or explicit feedbacks (e.g. check-ins or venue ratings) that these users express about venues. However, MF suffers in that users may only have rated very few venues. To alleviate this problem, recent literature have leveraged additional sources of evidence, e.g. using users' social friendships to reduce the complexity of – or regularise – the MF model, or identifying similar venues based on their comments. This paper argues for a combined regularisation model, where the venues suggested for a user are influenced by friends with similar tastes (as defined by their comments). We propose a MF regularisation technique that seamlessly incorporates both social network information and textual comments, by exploiting word embeddings to estimate a semantic similarity of friends based on their explicit textual feedback, to regularise the complexity of the factorised model. Experiments on a large existing dataset demonstrate that our proposed regularisation model is promising, and can enhance the prediction accuracy of several state-of-the-art matrix factorisation-based approaches.

## 1. INTRODUCTION

With an overwhelming amount of data in Location-Based Social Networks, recommending venues that are relevant to the users' preferences is challenging. Traditional Matrix Factorisation [9] (MF) is a collaborative filtering-based approach widely used to predict ratings that users give to venues. A regularisation technique is commonly used to ensure that the MF models simple and to avoid over-fitting. MF treats all users equally, i.e. the predicted rating of a *target* user for a venue can be influenced by any other user, as long as they share similar preferences, and regardless of their relationship. However, when the users' rating data are sparse in nature, i.e. users/venues have very few ratings, the rating prediction accuracy of traditional MF approaches can be significantly degraded. To alleviate this problem, previous works [4, 10] have incorporated social network information (e.g. the user's friendships) to enhance the prediction accuracy. In particular, Ma *et al.* [10] proposed a MF approach called SoReg, where a social regularisation component assumes that users are likely to be influenced by their friends who rate similar venues with similar scores.

Apart from social network information, the textual comments left by users are another form of explicit preference information. Indeed, comments can be leveraged to enhance the prediction accuracy and provide useful insights about user's preferences. For instance, Hu *et al.* [5] showed that by incorporating the comments left by the users on venues (along with the geographical proximity of venues), a more effective rating prediction model can be achieved. Similarly, Huang *et al.* [6] showed that a representation of user's preferences extracted from comments provides useful insights for restaurant owners. They argued that users' ratings are influenced by a variety of aspects. For example, assuming the widely used rating scale range from 1 to 5 (where 1 denotes user does not like a venue and 5 is the most positive rating), the reason that a user rated a restaurant 4 rather than 5 could be that he/she liked the food and price but was disappointed by the service. Typically, comments are represented as a Bag-of-Words (BoW). However, we argue that BoW is not effective at capturing the semantic properties of comments because they ignore the ordering and the semantics of the words. An example of two comments about a venue that are semantically similar are *delicious sushi bar in Illinois* and *best Japanese restaurant in Chicago*. While these two comments have no words in common, their semantic properties are similar, however a similarity measured based upon BoW would fail to capture these properties. Recently, Mikolov *et al.* [11, 12] proposed a word embedding technique that represents a word within a multi-dimensional vector space based on the contexts surrounding the word. Previous work [1] has shown that using word embeddings to analyse comments not only captures the semantic properties but also the sentiment expressed in the comments. With such benefits, various literature [1, 13, 14] have proposed to apply word embeddings to recommendation tasks. For instance, Musto *et al.* [13] proposed a collaborative filtering technique that applies word embeddings to the textual content of items obtained from their Wikipedia pages, in order to recommend items to users. In contrast, we propose a model that models the comments of each user using word embeddings.

Although the previous venue recommendation literature have shown that both social network [4, 10] and comments [5] are important sources of evidence to enhance the quality of recommendations, a model that seamlessly incorporate



*CIKM'16 , October 24-28, 2016, Indianapolis, IN, USA*



DOI: http://dx.doi.org/10.1145/2983323.2983889

these two factors has not been previously studied. In addition, unlike the approach proposed by Hu *et al.* [5], we aim to leverage comments at the user-level, rather than at the venue-level (i.e. to represent user's preferences). Moreover, we do not consider the geographical locations of venues in this model. Table 1 summarises user-venue prediction approaches and their intuitions. Inspired by the previous studies mentioned above, we argue that by considering score of rating and comments both target user and his friends left for venues, a more accurate user-venue rating prediction model can be achieved. To address these challenges, we propose a regularisation model that enables traditional MF models to incorporate additional informations. Our contributions in this paper are the following: (1) we propose a regularisation model that incorporates social and textual information, which exploits word embeddings to model friends' preferences based on the comments they have left on venues, (2) we conduct experiments on a large publicly available dataset and demonstrate that our proposed regularisation model is promising, and can enhance the prediction accuracy of several state-of-the-art rating prediction approaches.

## 2. MATRIX FACTORISATION WITH SOCIAL & TEXTUAL REGULARISATION

In this section, we first outline the traditional Matrix Factorisation (MF) and the social regularisation model proposed by Ma *et al.* [10] in details. Next we explain how we exploit word embeddings to model friends' preferences from their comments and describe our proposed social-textual regularisation model (Section 2.3). The intuitions for each model are highlighted in Table 1. Later, in Section 4, we demonstrate the usefulness of our proposed model in enhancing the prediction accuracy of several state-of-the-art MF rating prediction approaches.

### 2.1 Traditional Matrix Factorisation

Traditional Matrix factorisation-based approaches assume that users can influence each others if they share similar preferences, i.e. who rate venues similarly (see Table 1). Traditionally, user-venue ratings are represented as a matrix $R^{m\times n}$ where $m$ and $n$ are a number of users and venues, respectively, and $R_{i,j}$ indicates the 1-5 scale rating feedback by user $i$ on venue $j$. Traditional MF techniques aim to approximate the matrix $R$ by finding a decomposition of $R$, i.e. a 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:

$$R \approx U^T V$$

Next, the two latent factor matrices $U$ and $V$ are optimised by minimising the regularised square error on a set of observed ratings using decomposition technique such as singular value decomposition. Hence, a loss function is defined as:

$$L(U,V) = \min_{U,V} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{i,j}\cdot(R_{i,j} - U_i^T V_j)^2 \qquad (1)$$

where $I_{i,j}$ is an indicator variable that is 1 if user $i$ rated venue $j$, otherwise 0. To avoid overfitting, i.e. $U^T V = R$, a traditional regularisation technique is added into Eq. (1):

$$\underset{MF}{L(U,V)} = L(U,V) + \frac{\lambda}{2}(\|U\|_F^2 + \|V\|_F^2) \qquad (2)$$

where $\lambda$ is a regularisation parameter and $\|.\|_F^2$ denotes the Frobenius norm. Finally, Stochastic Gradient Descent (SGD) is applied to find a local minimum of the loss function, by optimising each of the latent factor matrices $U, V$, while fixing the other, until convergence.

$$\begin{aligned}
\frac{\partial \underset{MF}{L(U,V)}}{\partial U_i} &= \sum_{j=1}^{n} I_{i,j}(R_{i,j} - U_i^T V_j)V_j + \lambda U_i \\
\frac{\partial \underset{MF}{L(U,V)}}{\partial V_j} &= \sum_{i=1}^{m} I_{i,j}(R_{i,j} - U_i^T V_j)U_i + \lambda V_j
\end{aligned} \qquad (3)$$

### 2.2 Social Regularisation

Based on an assumption that friends who have similar preferences are likely to rate venues similarly (see Table 1), Ma *et al.* [10] proposed a social regularisation model, denoted SoReg, which aims to minimise the distance between latent factors of target user $U_i$, and his/her friends, $U_f$, by incorporating friendship information from the social network as:

$$\underset{SoReg}{L(U,V)} = \underset{MF}{L(U,V)} + \frac{\alpha}{2}\sum_{i=1}^{m}\sum_{f\in\mathcal{F}(i)} pcc(i,f)\|U_i - U_f\|_F^2 \quad (4)$$

where $\mathcal{F}(.)$ is the set of friends of user $i$, $\alpha$ is a parameter that controls social influences. $pcc()$ estimates the similarity between the ratings of two users using the Pearson Correlation Coefficient (PCC), calculated as follows:

$$pcc(i,f) = \frac{\sum\limits_{j\in V_r(i)\cap V_r(f)} (R_{ij} - \overline{\mathrm{R}}_i)\cdot(R_{fj} - \overline{\mathrm{R}}_f)}{\sqrt{\sum\limits_{j\in V_r(i)\cap V_r(f)} (R_{ij} - \overline{\mathrm{R}}_i)}\cdot\sqrt{\sum\limits_{j\in V_r(i)\cap V_r(f)} (R_{fj} - \overline{\mathrm{R}}_f)}} \qquad (5)$$

where $V_r(i)$ are the venues that user $i$ has rated and $\overline{\mathrm{R}}_i$ is their average rating. This regularisation ensures that friends who have rated venues similarly (e.g. $pcc(i,f) = 0.9$) are predicted to give similar ratings to other venues, $U_i^T V \approx U_f^T V$.

### 2.3 Social Regularisation using Word Embeddings

As mentioned in Section 1, we argue that users are not only influenced by their friends who visit similar venues and provide similar ratings to that 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 for each venue (Table 1). An intuitive scenario for this assumption could be that $u_i$, $u_j$ and $u_k$ are friends who all enjoy a visit to a restaurant. $u_i$ and $u_j$ are impressed by the service at the restaurant and food, but $u_k$ is only impressed by the setting of the restaurant. Estimating user similarity using PCC exploited by Ma's approach [10] (Eq. (4)) can only capture the similarity of the users based on their ratings, which in this case are all positive. However, the taste of $u_k$ is quite different from the others. To tackle this problem, we propose to examine the comments left by friends, to identify those that share similar interests. In particular, instead of measuring the comment similarity using a traditional bag-of-words representation, we propose to exploit word embeddings, a popular technique that can capture semantic regularities of text [11, 12], to effectively model a user's tastes over a venue. For a comment $c_{ij}$ left by user $i$ on venue $j$, we obtain a word embedding representation by summing the vectors of the terms that occurred in the user's comment on the venue as follows:

$$w2v(c_{ij}) = \sum_{t\in c_{ij}} \vec{\nu}_t \qquad (6)$$

**Table 1: Overview of user-rating prediction approaches.**

| Models | Social | Comments | Params | Intuition |
|---|---|---|---|---|
| MF, $MF_N$ [9], $MF_P$ [15], SVD [7] | × | × | $\lambda$ | Users are likely to prefer venues rated that other similar users rate highly. |
| VMF [5] | × | ✓ | $\lambda$ | + Users are likely to prefer venues that share similar characteristics (according to textually similar comments). |
| SoReg [10], TrustSVD [4] | ✓ | × | $\lambda$, $\alpha$ | + Users are likely to prefer venues that their friends rate highly. |
| BoWReg | ✓ | ✓ | $\lambda$, $\alpha$ | + Users are likely to prefer venues visited by their friends who have similar tastes (based on textually similar comments). |
| DeepReg | ✓ | ✓ | $\lambda$, $\alpha$ | + Users are likely to prefer venues visited by their friends who have similar tastes (based on semantically similar comments). |

where $t$ is a term that occurs in the comment $c_{ij}$ and $\vec{\nu}_t \in \mathbb{R}^k$ is a vector representation of term $t$ obtained from word embeddings. $k$ is the number of dimensions in the word embedding space. Thereafter, the similarity between two users based on their comments about venues they have both visited is estimated as:

$$sim_{w2v}(i,f) = \frac{\sum_{j \in V_c(i) \cap V_c(f)} sim(w2v(c_{ij}), w2v(c_{fj}))}{|V_c(i) \cap V_c(f)|} \quad (7)$$

where $sim()$ denotes the cosine similarity between two vectors and $V_c(i)$ is the set of venues for which user $i$ has left a comment. Next, inspired by Ma *et al.* [10], we integrate the social and textual information into MF as a component of the regularisation (but using comments rather than ratings to identify similar friends):

$$\frac{\alpha}{2} \sum_{i=1}^{m} \sum_{f \in \mathcal{F}(i)} sim_{w2v}(i,f) \|U_i - U_f\|_F^2 \quad (8)$$

Similar to the traditional Frobenius norm regularisation technique in Eq. (1), our proposed social-textual regularisation model can be easily added to various matrix factorisation-based approaches. We add our proposed social regularisation model (Eq. (8)) to Eq. (2) and perform SGD on latent factors $U_i$ to obtain a local minimum of objective function (denoted as DeepReg) while SGDàÿ§ on the latent factor $V_j$ remains the same as in Eq. (3):

$$\frac{\partial \underset{DeepReg}{L(U,V)}}{\partial U_i} = \frac{\partial \underset{MF}{L(U,V)}}{\partial U_i} + \alpha \sum_{f \in \mathcal{F}(i)} sim_{w2v}(i,f)(U_i - U_f) \quad (9)$$

## 3. EXPERIMENTAL SETUP

This section describes the experimental setup used to evaluate the usefulness of our proposed regularisation model in enhancing the rating prediction accuracy of traditional MF, in comparison with various state-of-the-art rating prediction approaches. All experiments are conducted using the publicly available Yelp dataset.[1] The dataset consists of 2,225,213 ratings by 552,339 users for 77,079 venues. It contains social network information, with ∼3.5M friend links. Following Hu *et al.* [5], we perform stemming and remove standard stopwords from comments in the dataset. We report the user-rating prediction accuracy in term of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) (for both metrics, lower is better), which are widely used in previous literature [4, 5, 10]. We compare our proposed approach, called DeepReg (Eq. (9)), with a number of baselines, which can be grouped into 4 categories: traditional Matrix Factorisation (MF), social-based MF, textual-based MF and a baseline social-textual-based MF that does not consider word embeddings. The baselines are summarised below, while their parameters and sources of evidence are highlighted in Table 1.

[1] https://www.yelp.com/dataset_challenge

**MF**, **$MF_N$**, **$MF_P$** and **SVD**: Four matrix factorisation baselines that consider only the user-venue matrix to predict the ratings. We use traditional MF (denoted MF) as described in Section 2.1, Non-negative MF ($MF_N$) [9], Probabilistic MF ($MF_P$) [15] and Singular Value Decomposition (SVD) [7], a state-of-the-art MF that considers user and venue biases (e.g. low or high average ratings $R_i$). These are widely used baselines in recommendation systems [4, 5, 10].
**SoReg** and **TrustSVD**: Both consider the user-venue ratings matrix $R$ and friendship $\mathcal{F}(.)$ to predict ratings. SoReg, discussed above, is a state-of-the-art regularisation model (Eq. (4)), while TrustSVD [4] is a state-of-the-art social-based MF that considers social information, built upon SVD.
**VMF**: is a state-of-the-art textual-based MF approach that considers textual information (i.e. comments about a venue) proposed by Hu *et al.* [5]. To permit a fair evaluation, we have re-implemented the approach to consider only textual information, and ignore the geographical location properties of venues, in common with our own proposed approach that does not consider the location of venues.
**BoWReg**: Building upon SoReg, this represents the user's comments using BoW similarity, and hence considers both social friends and comments. In particular, this model is similar to our proposed model but instead the similarity between two users is estimated using an average of the cosine similarity of the user's comment BoW vectors as follows:

$$sim_{bow}(i,f) = \frac{\sum_{j \in V_c(i) \cap V_c(f)} sim(bow(c_{ij}), bow(c_{fj}))}{|V_c(i) \cap V_c(f)|} \quad (10)$$

where $bow(c_{ij})$ returns a vector that represents the term frequency of terms occurring in comment $c_{ij}$. We use **BoWReg** as a baseline that considers both social network and textual information without using word embeddings.

We implement and conduct all experiments using LibRec [2], a Java library for recommender systems. Following previous works [4, 10], we set the latent factor dimension $d$ to 10 and $\lambda = 0.001$. Our experiments are conducted in a 5-fold cross-validation, split by user, where each fold has 60% training, 20% validation and 20% testing. We use the validation set to determine the social parameter, $\alpha$. An experiment conducted by Ma *et al.* [10] showed that the value of $\alpha$ has a marked impact on the prediction accuracy. Setting $\alpha$ extremely high can degrade the prediction accuracy as users are fully influenced by their friends (i.e. users are predicted to prefer every venue that their friends like). Following [10], we vary $0.000001 \leq \alpha \leq 1$, multiplying $\alpha$ by 10 at each iteration. For each fold, the $\alpha$ that minimised RMSE on the validation set is used for testing. For word embeddings, we use the Word2Vec tool[2], training a skip-gram model [11] using the default settings (window size 5 and $k = 100$ dimensions[3]) on the Yelp dataset. We choose the skip-gram model over the continuous-bag-of-words (CBOW) model as Mikolov *et al.* [12] reported that the skip-gram model performs better than or equally to the CBOW model.

[2] https://code.google.com/archive/p/word2vec/ [3] Initial experiments showed no marked change in prediction accuracy for larger $k$.

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 performing result, which is highlighted in bold.

| Metrics | $MF_N$ | $MF_P$ | MF | VMF | SoReg | BoWReg | $DeepReg_{MF}$ | SVD | TrustSVD | $DeepReg_{SVD}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| MAE | 1.3881 | 1.5924 | 1.1640 | 1.2198 | 1.1260 | 1.1004 | **1.0781** | 1.1067 | 1.0292 | **1.0077** |
| Δ | 22.34% | 32.30% | 7.39% | 11.62% | 4.26% | 2.03% | | 8.94% | 2.09% | |
| RMSE | 1.8994 | 2.0643 | 1.5243 | 1.5006 | 1.3870 | 1.4354 | **1.3456** | 1.4357 | 1.3034 | **1.2504** |
| Δ | 29.16% | 34.82% | 11.72% | 10.33% | 2.99% | 6.26% | | 12.90% | 4.06% | |

# 4. EXPERIMENTAL RESULTS

Table 2 reports the rating prediction accuracy, in terms of MAE and RMSE, of our proposed DeepReg and other baselines. The left hand side of the table compares systems based upon MF, while the right-hand side contains SVD-based approaches; as explained below, DeepReg can be implemented upon both MF and SVD, which we denote with subscripts, i.e. $DeepReg_{MF}$. Firstly, we note that the relative prediction accuracy of the baselines on the Yelp dataset are consistent with the results reported by Ma *et al.* [10] and Guo *et al.* [4], namely that SoReg outperforms $MF_N$ and $MF_P$, and TrustSVD outperforms SVD, respectively.

Next, from the left portion of Table 2, we observe that $DeepReg_{MF}$, our proposed social-textual regularisation model using word embeddings, outperforms all MF baselines for both MAE and RMSE. In particular, our proposed regularisation model can enhance the prediction accuracy of the traditional MF (MF) by 7% for MAE and 11% for RMSE. These imply that social information and textual content of comments are useful in enhancing the accuracy of rating prediction. Moreover, by comparing $DeepReg_{MF}$ and SoReg, the MAE and RMSE results show that minimising the distance between latent factors of friends based on the semantic similarity of their comments is more effective than estimating user similarity solely based upon similar ratings (i.e. PCC, as per Eq. (5)). Comparing $DeepReg_{MF}$ with BoWReg, we find that the prediction accuracy is again enhanced (2% for MAE and 6% for RMSE). This shows that the word embeddings offer a more useful, semantic space for modelling comments by users upon venues, which results in a more effective regularisation. Although the improvements in Table 2 are relatively small, we note that small improvements in MAE and RMSE can lead to marked improvements in the quality of recommendations in practice [8]. Moreover, the observed performances are evaluated over 2.2M ratings. Naturally, we believe that refined word embedding models (e.g. obtained by varying window size or background training dataset) may provide further improvements.

Finally, the right hand side of Table 2 reports the application of our proposed regularisation model to an approach based upon SVD. This allows comparison with TrustSVD [3], a state-of-the-art MF that considers social information built upon SVD. We note that DeepReg can be easily applied to SVD by just replacing the loss function of MF in Eq. (9) with the loss function of SVD[4]. On analysing the right hand side of Table 2, we find that these implementations are consistent with results reported by Guo *et al.* [4], as TrustSVD outperforms SVD by ∼7-10% on both metrics. Moreover, SVD with our proposed regularisation model outperforms TrustSVD by 2% and 4% on MAE and RMSE, respectively. These results imply that not only friends can influence each others, but that the semantic properties of comments left by friends also play an important role in generating more effective matrix factorisation models.

[4] For more details of SVD's loss function, see [7].

# 5. CONCLUSIONS

In this paper, we explored whether social network information and textual content of comments can be leveraged to enhance the accuracy of rating prediction. We proposed a social-textual regularisation model that seamlessly incorporates both the user's social network and comments as a source of evidence, to minimise the latent factors of friends based on their semantic similarity. Our comprehensive experiments on a public dataset of 500k users and 2.2M ratings demonstrated the usefulness of our proposed regularisation model in enhancing the prediction accuracy of several MF-based approaches, and can outperform the state-of-the-art for social regularisation (SoReg) by 3-4%, and the state-of-the-art for social matrix factorisation (TrustSVD) by 2-4%.

# 6. REFERENCES


[1] X. Fu and X. Li. From movie reviews to restaurants recommendation. Technical report, Stanford University, 2015. http://stanford.io/2aB0T0f.

[2] G. Guo, J. Zhang, Z. Sun, and N. Yorke-Smith. LibRec: A Java library for recommender systems. In *Proc. of UMAP*, 2015.

[3] G. Guo, J. Zhang, and N. Yorke-Smith. A novel Bayesian similarity measure for recommender systems. In *Proc. of IJCAI*, 2013.

[4] G. Guo, J. Zhang, and N. Yorke-Smith. TrustSVD: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings. In *Proc. of AAAI*, 2015.

[5] L. Hu, A. Sun, and Y. Liu. Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. In *Proc. of SIGIR*, 2014.

[6] J. Huang, S. Rogers, and E. Joo. Improving restaurants by extracting subtopics from Yelp reviews. *iConference (Social Media Expo)*, 2014.

[7] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proc. of SIGKDD*, 2008.

[8] Y. Koren. Factor in the neighbors: Scalable and accurate collaborative filtering. *Transactions on Knowledge Discovery from Data (TKDD)*, 4(1), 2010.

[9] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), 1999.

[10] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King. Recommender systems with social regularization. In *Proc. of WSDM*, 2011.

[11] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv:1301.3781*, 2013.

[12] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Proc. of NIPS*, 2013.

[13] C. Musto, G. Semeraro, M. De Gemmis, and P. Lops. Word embedding techniques for content-based recommender systems: an empirical evaluation. In *Proc. of RecSys*, 2015.

[14] M. G. Ozsoy. From word embeddings to item recommendation. *arXiv:1601.01356*, 2016.

[15] R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In *Proc. of ICML*, 2008.