A tag- and genre-based enhancement to the factorization machine approach for multimedia recommendation

Qunhao Feng[1], Zhangbin Wen[2], Shun Long[1] and Weiheng Zhu[1, a]

[1] Dept. of Computer Science, Jinan University, Guangzhou 510632, China;

[2] Guangdong Jitong Information Development Co.Ltd, Guangzhou, China.

[a] Corresponding author: tzhuwh@jnu.edu.cn

Abstract

Most of the state-of-the-art multimedia recommendation approaches take advantage of context in the form of either tags or genres. However, to the best of our knowledge, there is no published works on using both of them. This paper presents a novel approach which integrates both tags and genres in its multimedia recommendation model. Tags are first used in clustering multimedia objects, before Factorization Machine (FM) is applied to integrate first with clusters and then with genres in the recommendation model. Experimental results on the hetrec2011-movielens-2k dataset suggest that, in term of RMSE, this proposed approach achieves a performance improvement of 1%, i.e. 0.7484 in RMSE in comparison to that of 0.7567 by the original FM model.

Keywords

Multimedia Recommendation; Factorization Machine; Tag; Genre; Clustering.

1. Introduction

The past two decades has witnessed a rapid growth in availability of the multimedia information. The resulting information overload generates a growing demand for multimedia information management and recommendation. Information technologies developed in the past few decades enables users to generate a huge amount of data which can be collected and analyzed for business intelligence purpose[2]. In particular, in the area of in customized recommendation, a user's potential interests are usually predicted based on either collaborative and content-based filtering or personality-based analysis of their behaviors or traces. Since Web 2.0, tagging has become a popular increasingly popular due to the growth of social bookmarking, image sharing[8] and social network websites. It is a natural way for users to aid classification, mark ownership, note boundary and indicate online identity. Tags may be chosen informally and personally, or from a controlled vocabulary. It helps to describe an item so that it can be found easily via either browsing or searching. Most multimedia information comes in the forms of image, audio and video clips, usually with little text description. This nature makes tagging of particular importance in management and retrieval.

Various approaches have been proposed to make use of user-generated tags in multimedia recommendation. Bang et.al[1] presents a recommendation model which adopts word embedding for tags and predicts the unknown ratings in user-item matrix by collective matrix factorization with the user-tag matrix. Wei et.al[10] proposed a hybrid approach for movie recommendation via tags and ratings. A social movie network and a preference-topic model are first constructed, before the social tags are extracted, normalized and re-conditioned according to user's tag-based preference. This model is then enhanced by using user historical ratings as supplementary information in its SVD-based[5] recommendation. The success of these tag-based recommendations relies on the users' historical records. Some approaches failed to deliver satisfying results as promised in practice because the amount of such user records is insufficient. To tackle this problem, genre-based recommendation is proposed. Genres are usually formed by conventions and assigned by a group of specialists, which makes it far less affected by individual user's preference. This makes it an ideal alternative for recommendation. Zhang et.al[11] proposed a collaborative filtering algorithm which is based on similarities of both genres and ratings. Experimental results suggest that it can significantly lower the mean absolute errors in recommendation. Sang et.al[3] presents a genre correlation based multimedia recommendation algorithm which improves the precision of recommendations. However, to the best of our knowledge, there is no published work on combining both user-generated tags and genres in recommendation.

We present in this paper a novel approach which integrates both tags and genres in its multimedia recommendation model. Tags are first used in multimedia object clustering, before Factorization Machine (FM) is applied to integrate clusters with genres in the recommendation model. Experimental results on the hetrec2011-movielens-2k dataset suggest that this proposed approach outperforms the existing FM-based approach by roughly 1% in term of RMSE.

The paper is organized as follows. Section II presents a brief review of the FM[9] model. Our new model, namely Factorization Model Clustering plus Genre (FMC+G), is explained in depth in section III. Section IV presents our experimental results, before some concluding remarks.

1.1 factorization machine (FM)

First proposed by S.Rendle in 2010, factorization machine (FM) is a new model class that combines the advantages of support vector machine (SVM)[4] with factorization models. Like SVMs, FMs are general predictor working with any real-valued feature vector. However, it uses factorized parameters to model all interactions between variables by mapping them to a low dimensional space, which makes it capable of estimating interactions even in problems with huge sparsity where SVMs sometimes fail. They are regarded as a generalization of a few factorization models including matrix factorization and SVD++[7].

A 2-way FM (degree d=2) captures all single and pairwise interactions between variables. The model is specified as

$$\hat{y}(x) \coloneqq w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} < \mathbf{v_i}, \mathbf{v_j} > x_i x_j \tag{1}$$

where n is the number of features, k is the dimension of the vector used to specify a feature, and $w_0 \in R$, $W \in R^n$, $V \in R^{n\times K}$ are the three parameters to be estimated. $< v_i, v_j >$ is the dot product of two k-dimensional vector v_i and v_j .

FM models are non-linear supervised models which can do both regression and classification. They are usually trained by stochastic gradient descent (SGD), alternative least square (ALS) or Markov Chain Monte Carlo (MCMC).

2. Factorization Model with Clustering and Genre (FM+CG)

The proposed Factorization Model with Clustering and Genre (FM+CG) model includes three major steps, as shown in Fig.1. First, a Tag-Weight matrix is created for the multimedia objects based on user-generated tags. It is then factorized via SVD and the features of these objects are extracted from the resulting U, before K-means is applied for clustering purpose. Next, each of these multimedia objects is assigned a genre. Finally, both the genre and clustering results are integrated into the users rating matrix to create the FM+CG model for recommendation purpose. It is worth noting that parameters of the FM+CG model are tuned via gradient descent. These steps are explained in depth in the following subsections.

2.1 The tag-weight matrix

Tags are assigned to multimedia objects by users to aid classification, mark ownership, note boundaries, and indicate online identity.It is worth noting that a multimedia object may be assigned multiple tags by various users. Each tag is therefore specified as a triple in the form of <multimediai, tagj, tagweightij>, where multimediai is the i-th object, tagj is the j-th tag, and tagweightij is the weight of the tagj for multimediai, which usually represents the number of times tagj is assigned to it. Considering that not every tag applies to every multimedia object, some entries are obviously missing, which leads to a sparse Tag-Weight matrix where 0 is assigned for these entries. The result is illustrated in Fig.2 below.

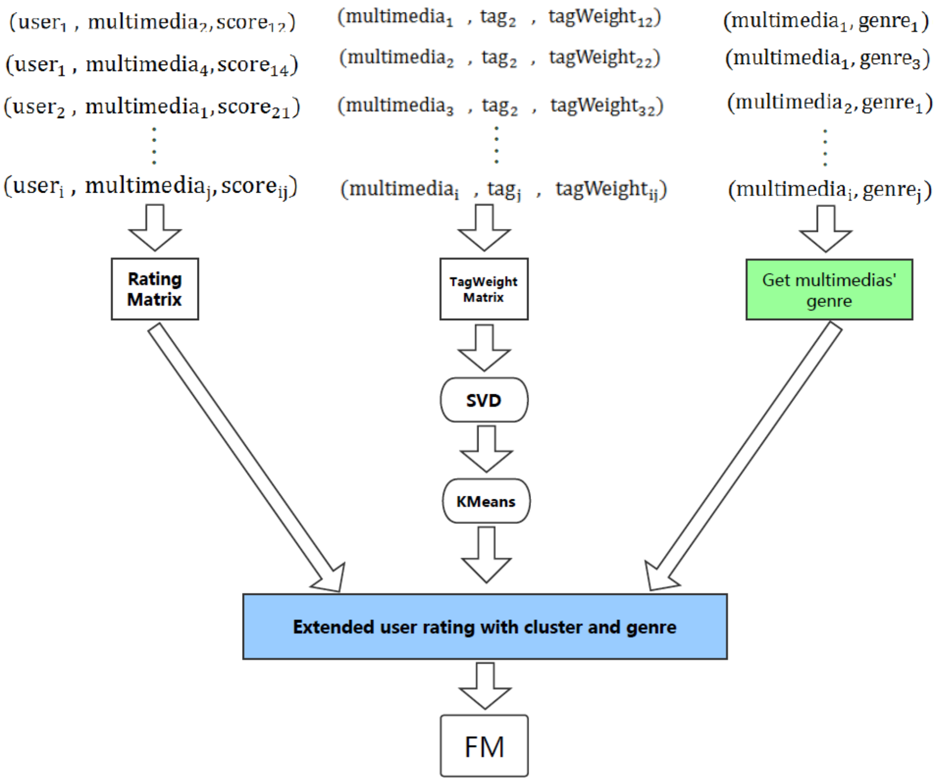

Fig.1 Basic workflow of Factorization Model with Clustering and Genre (FM+CG)

	tag$_1$	tag$_2$	∘ ∘ ∘	tag$_n$
multimedia$_1$	tagWeight$_{11}$	?	∘ ∘ ∘	tagWeight$_{1n}$
multimedia$_2$	?	tagWeight$_{22}$	∘ ∘ ∘	?
multimedia$_3$	tagWeight$_{31}$	?	∘ ∘ ∘	tagWeight$_{3n}$
∘ ∘ ∘	∘ ∘ ∘	∘ ∘ ∘	∘ ∘ ∘	∘ ∘ ∘
multimedia$_m$	tagWeight$_{m1}$	tagWeight$_{m2}$	∘ ∘ ∘	tagWeight$_{mn}$

A

	tag$_1$	tag$_2$	∘ ∘ ∘	tag$_n$
multimedia$_1$	tagWeight$_{11}$	0	∘ ∘ ∘	tagWeight$_{1n}$
multimedia$_2$	0	tagWeight$_{22}$	∘ ∘ ∘	0
multimedia$_3$	tagWeight$_{31}$	0	∘ ∘ ∘	tagWeight$_{3n}$
∘ ∘ ∘	∘ ∘ ∘	∘ ∘ ∘	∘ ∘ ∘	∘ ∘ ∘
multimedia$_m$	tagWeight$_{m1}$	tagWeight$_{m2}$	∘ ∘ ∘	tagWeight$_{mn}$

A’

Fig.2 The tag-weight matrix

2.2 Enhancement with clustering

Once the modified Tag-Weight Matrix A’ is obtained, Singular Value Decomposition (SVD) is applied for noise reduction and feature extraction, i.e. A’ = USVT, where V and U are two orthogonal matrices of $r \times n$ and $m \times r$ respectively, S is a diagonal matrix of $r \times r$ ($r \ll n$ and $r \ll m$). A matrix U much smaller than A’ is therefore obtained. Each row of this new matrix U is considered as a feature of these multimedia objects.

Based on these features obtained, K-means is then applied to cluster these objects based on the Euclidean distance between them, which is defined as follows. Let $v_i = (v_{i1}, v_{i2}, v_{i3} \dots v_{in})$ and $v_j = (v_{j1}, v_{j2}, v_{j3} \dots v_{jn})$ are two multimedia objects, their Euclidean distance is defined as

$$d\left(v_i, v_j\right) = \sqrt{(v_{i1} - v_{j1})^2 + (v_{i2} - v_{j2})^2 + \cdots + (v_{in} - v_{jn})^2} \qquad (2)$$

The enhanced FM model can therefore be achieved by integrating the clustering results into the factorization model.

$$\mathrm{FM}(R, C) \rightarrow R' \qquad (3)$$

Where R is the rating matrix, C is clusters of these multimedia objects and R′ is the predicted values of the unknown entries.

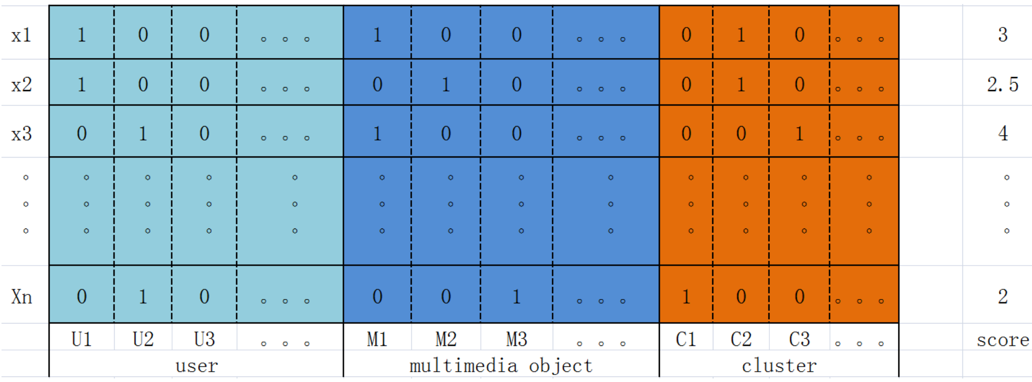

Fig.3 The user rating matrix extended with clustering

To implement this, the user-rating matrix is extended with clustering, as illustrated in the Fig.3, where the users are represented by the leftmost section (in light blue), and the multimedia objects are represented by the middle section (in blue). They are extended by the clustering results represented in the brown section. Factorization machine can be applied to this extended matrix for recommendation purpose, and the results are shown in the rightmost white section. It is worth noting that these data are One-Hot encoded[6], i.e. x3 in the 3rd row means that user U2 gave the multimedia object M1 (which belongs to cluster C3) a rating of 4.

2.3 the FM+CG recommendation model

It is believed that further performance improvement can be obtained if some objective information are adopted in addition to the inherently subjective tags provided by users. We integrate genres of these multimedia objects as additional dimensions or indicators for classification/recommendation.

In practice, a multimedia object may be considered of multiple genres, for instance, the movie Star Wars are considered as both action and Sci-Fi genres). To address this, various priorities are assigned to these genres in a way similar to the criteria selection used in decision tree generation, intuitively, higher priority suggests that people tend to have unanimous opinion for objects of this genre, and the corresponding recommendations are therefore easier to make. The results are illustrated in Table.1, where the genre Short is given the highest priority 1, and Film-Noir is given priority 2, and so on.

Table.1 Priorities of Genres

Genre	Priority	Genre	Priority	Genre	Priority	Genre	Priority
Short	1	Mystery	6	Romance	11	Comedy	16
Film-Noir	2	War	7	Thriller	12	Action	17
Documentary	3	Animation	8	Adventure	13	Children	18
IMAX	4	Crime	9	Fantasy	14	Sci-Fi	19
Drama	5	Western	10	Musical	15	Horror	20

The initial Factorization Model with Clustering and Genre (FM+CG) model is therefore obtained.

$$\mathrm{FM}(\mathrm{R},\mathrm{C},\mathrm{G}) \rightarrow \mathrm{R}' \tag{4}$$

Where R is the rating matrix, C is clusters of these multimedia objects, G is the genre information, and R′ is the predicted values of the unknown entries.

x1	1	0	0	∘ ∘ ∘	1	0	0	∘ ∘ ∘	0	1	0	∘ ∘ ∘	1	0	0	∘ ∘ ∘	3
x2	1	0	0	∘ ∘ ∘	0	1	0	∘ ∘ ∘	0	1	0	∘ ∘ ∘	0	1	0	∘ ∘ ∘	2.5
x3	0	1	0	∘ ∘ ∘	1	0	0	∘ ∘ ∘	0	0	1	∘ ∘ ∘	0	1	0	∘ ∘ ∘	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Xn	0	1	0	∘ ∘ ∘	0	0	1	∘ ∘ ∘	1	0	0	∘ ∘ ∘					2
	U1	U2	U3	∘ ∘ ∘	M1	M2	M3	∘ ∘ ∘	C1	C2	C3	∘ ∘ ∘	G1	G2	G3	∘ ∘ ∘	score
	user				multimedia object				cluster				genre				

Fig.4 the tag-weight matrix enhanced with categories and genres

---

ALGORITHM: FM+CG

Input: rating data R , tag matrix A’ ,genres of multimedia objects G,

Output: Model parameters w_0 , $\mathbf{W}$, $\mathbf{V}$ // parameters of the FMC+G model

1: U $\leftarrow$ sparsesvd(A′);

2: clusters $\leftarrow$ kmeans(U);

3: for each multimedia object ∈ G do

4: get its genre from the genre table

5: extend R to TrainingSet with cluster and genre // as illustrated in Fig.4

6: end // end of for all multimedia object loop

6: $w_0 \leftarrow 0$; $\mathbf{W} \leftarrow (\mathbf{0}, \mathbf{0}, \dots, \mathbf{0}, \mathbf{0})$; $\mathbf{V} \leftarrow \boldsymbol{N}(\mathbf{0}, \sigma)$ // parameter initialization

7: repeat // Parameter tuning along the negative gradient direction

8: for (X , score) ∈ TrainingSet do

9: $w_0 \leftarrow w_0 - \eta((\tilde{r} - r) + \lambda^0 w_0)$ // update parameter w_0

10 for i $\in \{1, \dots p\} \wedge x_i \neq 0$ do

11: $w_i \leftarrow w_i - \eta((\tilde{r} - r)\, x_i + \lambda^w_{\pi(i)} w_i)$ // update parameter w_i

12: for $f \in \{1, \dots k\}$ do

13: $v_{i,f} \leftarrow v_{i,f} - \eta((\tilde{r} - r)\, x_i \sum_{j=1}^{n} v_{j,f}\, x_j - v_{i,f} x_i^2 + \lambda^w_{\pi(i)} w_i$;

14: end // end of f-loop

15: end //end of i-loop

16: end // end of for (X, score) ∈ T loop

17: until $\mathrm{loss}^{\mathrm{R}}(\mathrm{r}, \tilde{\mathrm{r}})$ is small enough

---

Algorithm 1: FM+CG MODEL

Based on the extended user rating matrix presented in Fig.3, the user rating matrix further extended with genres is illustrated in Fig.4, where the One-Hot-encoded genres are represented in the second to the right section (in green), For instance, x3 in the 3rd row means that user U2 gave the multimedia object M1 (which belongs to cluster C3 and of genre G2) a rating of 4.

The resulting FM+CG model is a regressive one which can be tuned in an iterative manner, as illustrated in Algorithm 1.

Minimum mean square error (MMSE) is used as a metric during the course of tuning this regressive FMC+G model, with a loss function defined as

$$\mathrm{loss}^{R}(r,\tilde{r}(\theta)) = \frac{1}{2}\sum_{i=1}^{n}(r^{(i)} - \tilde{r}^{(i)})^2 + \lambda\|\theta\|^2 \quad (5)$$

where n is the size of the test set and $\theta = (w_0, W, V)$. The L2 norm $\lambda\|\theta\|^2$ is used to avoid over-fitting and increase generalization of the model. Parameters of the proposed model are tuned via the application of the stochastic gradient descent (SGD) algorithm which tries to find the minima of the above loss function by an iterative update against its partial derivative until the loss function is reduced below a predefined threshold.

3. Experimental Results and Analysis

To evaluate its performance, we applied the proposed FM+CG model to a public dataset called hetrec2011-movielens-2k, which is an extension of MovieLens10M dataset published by GroupLens research group. This dataset contains 855598 ratings of 10197 movies given by 2113 users, together with other associated data such as directors, genres, tags and others. We skipped those movies without tags and used the remaining 821440 ratings in our experiments. The experiments were carried out on a computer armed with an Intel® Xeon(R) CPU E3-1226 v3 @ 3.30GHz 4 cores, 16GB memory, 500GB hard disk space and Ubuntu 16.04 LTS 64bit system. Root-mean-square error (RMSE) defined below is used as the metric for performance evaluation.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{u\in U, i\in I}(r_{u,i} - \tilde{r}_{u,i})}{n}} \quad (6)$$

First, the maximum number(s) of clusters in K-means was decided via the elbow method. The sum of distances of samples to their closest cluster centers are plotted against the number of clusters, and the results are given in Fig.5. The points of inflections are considered as appropriate candidate Ks. They are K=40, K=55, K=100, K=119, K=134, and K=159 respectively.

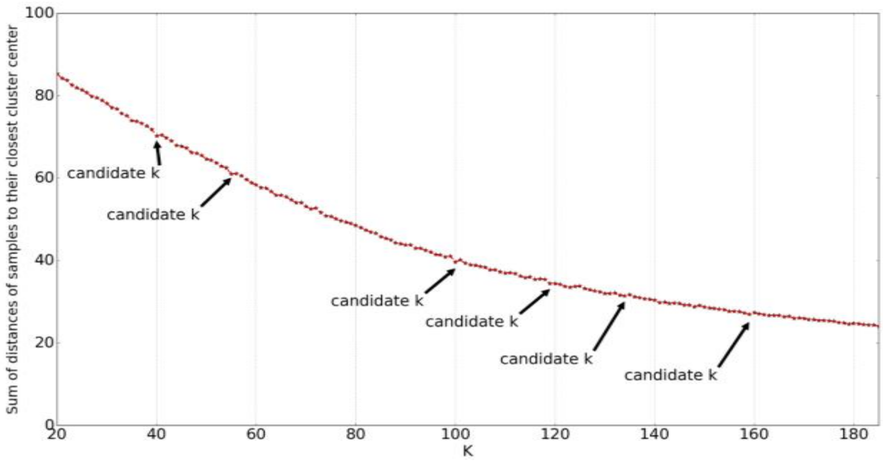

Fig.5 Identification of proper Ks (numbers of clusters) for FM+CG

Once Ks were decided, we carried out an empirical study about how the number of clusters (K) affects the performance of the proposed FM+CG model. Four models FM, FMC, FMG, and FM+CG are considered in our comparison, which refer to the original factorization machine (FM) model, the FM model enhanced by clustering, by genre, and by Clustering and Genre respectively. Fig.6 presents part of the results which are obtained with the number of latent factors is set 19 and the number of iteration is set 100. The results suggest that FM+CG achieved the best result (RMSE=0.7484) when

K is set 134, whilst the performance of both FM and FM+G are not affected by K because no clustering is involved.

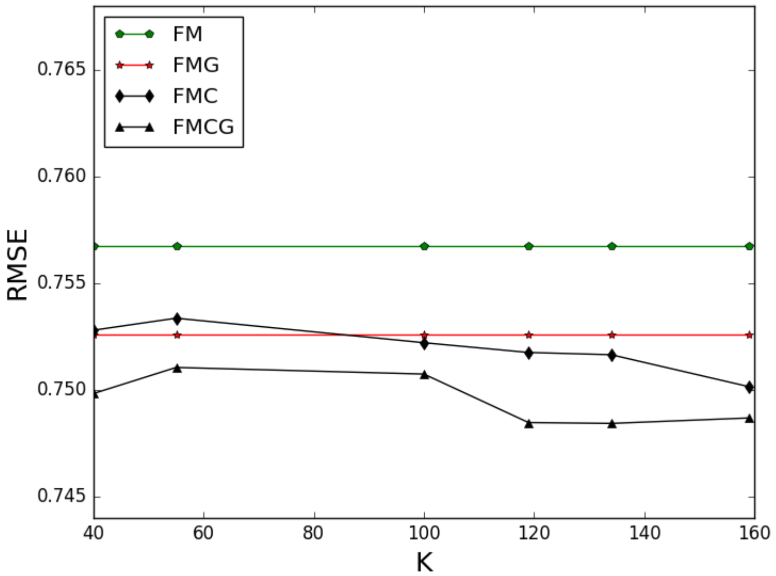

Fig.6 Performance with various numbers of clusters

Furthermore, we evaluated how the number of latent factors and number of iterations during tuning affect the performance of the proposed FM+CG model. The results are given in Fig.7 and .8 respectively.

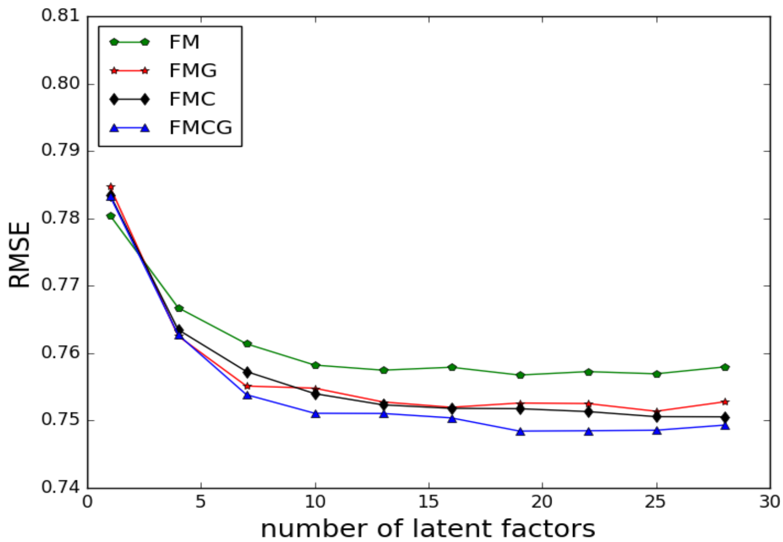

Fig.7 Performance with various numbers of latent factors

Once the best number of clusters is decided as above, we evaluate how the proposed model perform when given different number of latent factors, and the results are illustrated in Fig.7. It suggests that the number of latent factors has a significant impact on all four models. The RMSEs of all four models all falls when the number of latent factors increases. It achieves its minimum in FM+CG when the number of latent factors is set 19, which suggests that there are 19 hidden features provided in the categories and genres of these movies. Intuitively, a small number may stop the model from learning before insufficient knowledge are obtained, whilst a larger number of latent factors may lead to over-fitting in the obtained model, with additional time and space cost for learning. In brief, 19 is shown to be the most appropriate number of latent factors in this case. However, it is worth noting that the swings of RMSE turn almost negligible when the number of latent factors goes beyond 10. This suggests that a modest number of latent factors will be enough in practice.

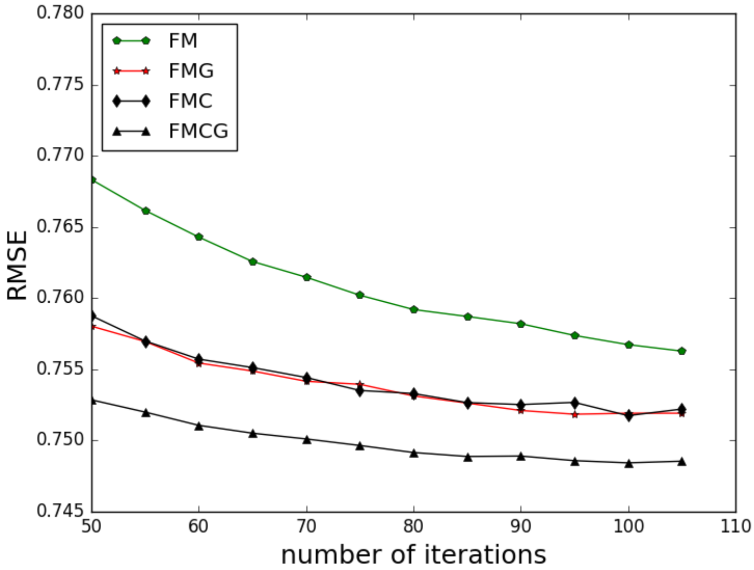

Fig.8 Performance with various numbers of iterations during tuning

Given the best number of latent factors, we evaluated how the FM+CG model performs during iterative tuning. Fig.8 suggests that the RMSEs of all four models fall when the number of iterations increases. And the variances turn negligible when the number of iterations goes beyond 100. This suggests that additional iteration in tuning brings no extra performance improvement but additional cost in time. It is worth noting that the performance of FM+CG is least affected by the number of iterations than its counterparts. This suggests that, in practice, additional cost (time for example) can be saved without loss of performance.

4. Conclusion

This paper presents a novel approach which integrates both tags and genres in its multimedia recommendation model. Tags are first used in clustering multimedia objects, before Factorization Machine (FM) is applied to integrate first with clusters and then with genres in the recommendation model. Experimental results on the hetrec2011-movielens-2k dataset suggest that, once tuned, this proposed approach can achieve 0.7484 in term of RMSE, in comparison to that of 0.7567 by the original FM model, i.e. an additional increase of 1%.

Acknowledgements

This work is supported by the Science and technology projects of Guangdong Province (C1432310400005, 2016A020224001, 2014B010103004, 2014B050505011, 2014B010117002 and 2015A030401052).

References

[1] Bang, H., & Lee, J. H. (2016). Collective Matrix Factorization Using Tag Embedding for Effective Recommender System. Joint, International Conference on Soft Computing and Intelligent Systems (pp.846-850).
[2] Cho, Y. H., Kim, J. K., & Kim, S. H. (2002). A personalized recommender system based on web usage mining and decision tree induction. Expert Systems with Applications, 23(3), 329-342.
[3] Choi, S. M., Ko, S. K., & Han, Y. S. (2012). A movie recommendation algorithm based on genre correlations. Expert Systems with Applications, 39(9), 8079-8085.
[4] Vapnik, V., & Cortes, C. (1995). Support vector networks. Machine Learning, 20(3), 273-297.
[5] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American society for information science, 41(6), 391.

[6] Harris, D., & Harris, S. (2007). Digital Design and Computer Architecture, Second Edition. Chian Machine Press.
[7] Koren, Y. (2010). Factor in the neighbors:scalable and accurate collaborative filtering. Acm Transactions on Knowledge Discovery from Data, 4(1), 1-24.
[8] Li, J., & Wang, J. Z. (2007). Real-time computerized annotation of pictures. IEEE Transactions on Pattern Analysis & Machine Intelligence,30(6), 985-1002.
[9] Rendle, S. (2010, December). Factorization machines. In Data Mining (ICDM), 2010 IEEE 10th International Conference on (pp. 995-1000). IEEE.
[10] Wei, S., Zheng, X., Chen, D., & Chen, C. (2016). A hybrid approach for movie recommendation via tags and ratings. Electronic Commerce Research and Applications, 18, 83-94.
[11] Zhang, Y., & Song, W. (2009). A collaborative filtering recommendation algorithm based on item genre and rating similarity. InComputational Intelligence and Natural Computing, 2009. CINC'09. International Conference on (Vol. 2, pp. 72-75).