Shanghai Jiao Tong University

ACM Class 2020

Tridimensional Regularization with Analysis in Social Network for Recommender Systems

CS3612, Machine Learning, Course Project

Chaofan Lin

2022. 5

Social Recommendation

SHANGHAI JIAO TONG UNIVERSITY

Task Therea are m users and n items. Given some ratings and a social network $G = (V, E)$, predict an unknown rating.

Figure: Some Ratings in Douban

Difficulty

SHANGHAI JIAO TONG UNIVERSITY

- The rating matrix R (i.e. R_{ij} : the rating of user i on item j) is of high sparisty.
- The prediction should be quick to give users a better experience.
- How to use social information to boost the recommender systems?

Matrix Factorization[1]

Idea We can use some latent features (e.g. "Character" and "Plot" for movies) to describe an item and the dimension of these features is small.

Method A user-matrix $U_{m\times d}$ and an item-matrix $V_{n\times d}$, with U_{ij} being the attention user i pays in feature j, and V_{kj} be the score of item k earns in feature j.

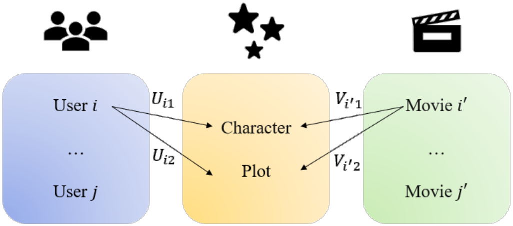


Figure: A Matrix Factorization Example with Dimension of Latent Features $= 2$

Matrix Factorization

SHANGHAI JIAO TONG UNIVERSITY

Prediction It predicts the rating matrix by $R \approx UV^T$.

Objective Function

$$\min_{U,V} \mathcal{L} = \frac{1}{2}\|R - UV^T\|_F^2$$

Rewrite

$$\min_{U,V} \mathcal{L} = \frac{1}{2}\{\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij} - U_i V_j^T)^2 + \lambda_1\|U\|_F^2 + \lambda_2\|V\|_F^2\} \tag{1}$$

Frobenius Norm

$$\|A\|_F = \sqrt{tr(A^T A)} = \sqrt{\sum a_{ij}^2}$$

Social Regularization[2]

Idea The key idea is that a movie received many positive feedbacks among your friends will more probably attract you.

Method Add a social regularization term as a contraint.

Figure: A Social Network Constraint Example

Social Regularization

Friends should be treated differently!

Social Regularization

SHANGHAI JIAO TONG UNIVERSITY

SR weighs friends by similarity.

Pearson Correlation Coefficient[3]

$$Sim(u,v) = \frac{\sum_{I_{ui}I_{vi}=1}(R_{ui}-\overline{R_u})(R_{vi}-\overline{R_v})}{\sqrt{\sum_{I_{ui}I_{vi}=1}(R_{ui}-\overline{R_u})^2}\cdot\sqrt{\sum_{I_{ui}I_{vi}=1}(R_{vi}-\overline{R_v})^2}} \tag{2}$$

Social Regularization

SHANGHAI JIAO TONG UNIVERSITY

Objective Function

$$\min_{U,V} \mathcal{L} = \frac{1}{2}\{\sum_{i=1}^{m}[\sum_{j=1}^{n} I_{ij}(R_{ij} - U_i V_j^T)^2 + \beta \sum_{v \in \mathcal{N}(i)} Sim(i,v) \|U_i - U_v\|_F^2] + \lambda_1 \|U\|_F^2 + \lambda_2 \|V\|_F^2\} \tag{3}$$

Motivation

SHANGHAI JIAO TONG UNIVERSITY

$$\min_{U,V} \mathcal{L} = \frac{1}{2}\{\sum_{i=1}^{m}[\sum_{j=1}^{n} I_{ij}(R_{ij}-U_iV_j^T)^2+\beta\sum_{v\in\mathcal{N}(i)} Sim(i,v)\|U_i-U_v\|_F^2]+\lambda_1\|U\|_F^2+\lambda_2\|V\|_F^2\}$$

Add two dimensions: influence and familiarity!

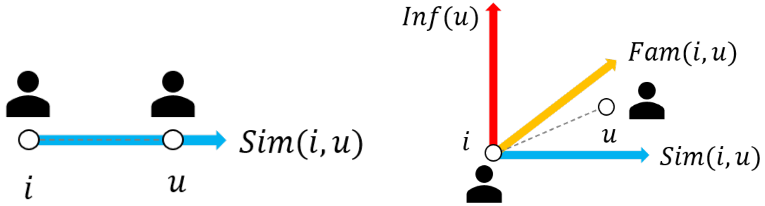
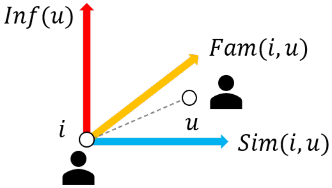


Figure: Comparsion between $Sim(i,u)$ (1 dimension) and $Tri(i,u)$ (3 dimensions)

Tridimensional Distance

SHANGHAI JIAO TONG UNIVERSITY



Tridimensional Distance Between Users

$$Tri(i,u) = \sqrt{(\alpha Inf(u))^2 + (\beta Sim(i,u))^2 + (\gamma Fam(i,u))^2} \tag{4}$$

Influence Analysis

Shanghai Jiao Tong University

Idea A recommendation from a big movie blogger is more convincible than a casual user.

Influence Analysis

Method Perform PageRank[4] in the users' social network.

$$Inf^t(i) = p \sum_{v \in \mathcal{N}(i)} \frac{Inf^{t-1}(v)}{\deg(v)} + (1-p) \cdot \frac{1}{m} \tag{5}$$

Influence Analysis

SHANGHAI JIAO TONG UNIVERSITY

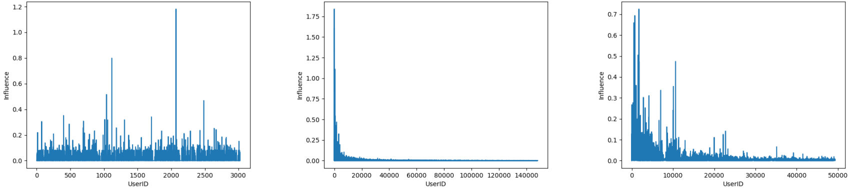


Figure: The Distribution of Influence

Familiarity

Idea Familiars not only have a greater chance to recommend but a higher credibility.

Measure[5]

$$Fam(u,v) = \frac{|\mathcal{N}(u) \cap \mathcal{N}(v)|}{\sqrt{|\mathcal{N}(u)|}\sqrt{|\mathcal{N}(v)|}} \tag{6}$$

Similarity Only VS. Tridimensional

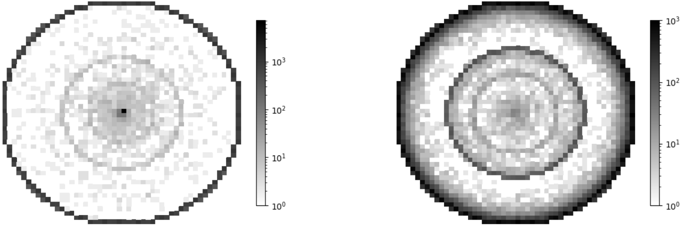


Figure: The Distribution of the Distance to the User with the Most Friends in Dianping

Our Model

We propose Tridimensional Social Regularization (TriSR). By combining (2), (4), (5) and (6), the objective function of TriSR can be formulated as

$$\min_{U,V} \mathcal{L} = \frac{1}{2}\{\sum_{i=1}^{m}[\sum_{j=1}^{n} I_{ij}(R_{ij} - U_i V_j^T)^2 + \sum_{v \in \mathcal{N}(i)} Tri(i,v) \|U_i - U_v\|_F^2] + \lambda_1 \|U\|_F^2 + \lambda_2 \|V\|_F^2\} \tag{7}$$

Datasets

SHANGHAI JIAO TONG UNIVERSITY

Table: Statistics of Datasets

Dataset	Douban	Dianping	Epinions
Users	3022	147918	49289
Items	6977	11123	139738
Ratings	195493	2149675	664824
Social Relationships	1366	629618	487181

Metrics

Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)

$$MAE = \frac{\sum_{(i,j,R_{ij})\in T} |\hat{R}_{ij} - R_{ij}|}{|T|}$$

$$RMSE = \sqrt{\frac{\sum_{(i,j,R_{ij})\in T} (\hat{R}_{ij} - R_{ij})^2}{|T|}}$$

Competitors

- **UserMean** predicts the rating $(user, item, score)$ by the average rating scores of $user$.
- **ItemMean** predicts the rating $(user, item, score)$ by the average rated scores of $item$.
- **UserCF** is a Collaborative Filtering algorithm focusing on users. It predicts the rating $(user, item, score)$ by the average rating scores of Top-N most similar users of $user$.
- **ItemCF**[6] is the same with UserCF but it focuses on items.
- **MF** is the matrix factorization method.
- **SR** is a matrix factorization with social regularization based on similarity.
- **TriSR** improves SR with a more detailed three-dimensional social regularization.

Platform

SHANGHAI JIAO TONG UNIVERSITY

- Reimplement all models and testing them on a self-implemented bench.
- Learning strategy: Adam (at first) and Mini-batch SGD (close to convergence).
- Take the 5-times average as the result.

Results

SHANGHAI JIAO TONG UNIVERSITY

Table: Experiment Results

Dataset	Metrics	UserMean	ItemMean	UserCF	ItemCF	MF	SR	TriSR
Douban	MAE	0.6961	0.6265	0.6874	0.7462	0.5984	0.5973	**0.5959**
	RMSE	0.8675	0.7875	0.8743	0.9440	0.7625	0.7614	**0.7597**
Dianping	MAE	0.5314	0.5127	0.5351	0.5464	0.5048	0.5034	**0.5033**
	RMSE	0.7216	0.6702	0.7044	0.7835	0.6700	**0.6674**	0.6677
Epinions	MAE	0.9369	0.9421	0.9827	0.9873	0.9251	0.9183	**0.9180**
	RMSE	1.2072	1.2243	1.2942	1.3314	1.1920	1.1832	**1.1824**

Impact of Parameters

SHANGHAI JIAO TONG UNIVERSITY

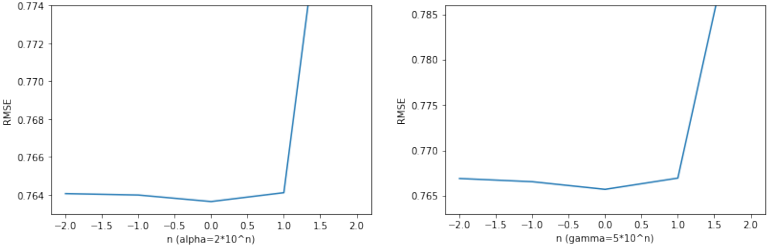


Figure: Impact of Parameters

Conclusion

SHANGHAI JIAO TONG UNIVERSITY

- A slight improvement comparing with SR and other baselines.
- Fail on the RMSE metric in Dianping: Dianping concentrates more on items while we adjust our model with too much users' information.
- Perform better in a user-oriented platform (where UserMean performs better than ItemMean).

Limitations and Gains

- My work is not an improvement on state of the art (DeepFM, xDeepFM...) in this field.
- The embedding idea behind Matrix Factorization is awesome and beautiful.
- Futher work: item network analysis, community analysis...

[1] KOREN Y, BELL R, VOLINSKY C. Matrix Factorization Techniques for Recommender Systems[J/OL]. Computer, 2009, 42(8): 30-37. https://doi.org/10.1109/MC.2009.263. DOI: 10.1109/MC.2009.263.

[2] MA H, ZHOU D, LIU C, et al. Recommender Systems with Social Regularization[C/OL]//WSDM '11: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. Hong Kong, China: Association for Computing Machinery, 2011: 287-296. https://doi.org/10.1145/1935826.1935877. DOI: 10.1145/1935826.1935877.

[3] BREESE J S, HECKERMAN D, KADIE C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering[C]//UAI'98: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence. Madison, Wisconsin: Morgan Kaufmann Publishers Inc., 1998: 43-52.

[4] BRIN S, PAGE L. The anatomy of a large-scale hypertextual Web search engine[J/OL]. Computer Networks and ISDN Systems, 1998, 30(1): 107-117. https://www.sciencedirect.com/science/article/pii/S016975529800110X. DOI: https://doi.org/10.1016/S0169-7552(98)00110-X.

[5] XIANG L. Recommendation System Practice[J], 2012: 162.

[6] SARWAR B M, KARYPIS G, KONSTAN J A, et al. Item-based collaborative filtering recommendation algorithms[C]//WWW '01. [S.l. : s.n.], 2001.

Thank You

Chaofan Lin · Tridimensional Regularization with Analysis in Social Network for Recommender Systems