3. PREDICTIVE MODELS

Since our dataset is Amazon Product Review, our research aims to make analysis based on users’ purchase history and their comment. Therefore, our goal is to establish a recommender model to predict the rating with user and item pair.

For baseline model, as we mentioned in section 2, is to make predictions according to the mean of ratings for all users and each user alone. This model makes prediction firstly based on user’s own mean rating with his history collected. Then if the user is new to model, the predicting rate would be mean of whole users.

Besides that, our first intuitive thought is simple latent factor model. The simple latent factor model considers the influence of both users and items.

G:\qq\598378694\Image\SharePic\20221128144914.png

Based on that, our next step is using complete latent factor model and SVD, since both models have much more complexity than original simple latent factor models.

C:\Users\Lenovo\AppData\Roaming\Tencent\Users\598378694\QQ\WinTemp\RichOle\YP_QYI54G0CA{AY[X{NP_5J.png

The formula is shown above. Comparing to simple latent factor model, this one assumes that there might be latent factors determining users’ decisions.

Latent factor models can perform well since they can fully learn biases and preferences and find those potential determining factors.

On the second hand, since the data contains not only user’s IDs, item’s IDs and ratings, but also user’s reviews and their summaries. Since the data contains abundant context information, our team also applied text mining models such as Word2Vec and Item2Vec.

While we apply and train our models, we encountered bunch of issues.

Firstly, we found that data are listed in a sorting order, which significantly increases the probability of the pairs of users and items being new to model. This will increase the case we predict by using rating mean. Plus, specifically, the text mining model will gain no similarities that makes the model collapsed. That is to say, we need to shuffle the dataset first to train models sufficiently

Secondly, since the dataset is huge, training usually takes the whole day, we shrink the training data, and then we encountered the issue of overfitting. The balancing seems a good means.

Our team modified our model by setting different hyperparameters like the learning rate, the regression lambda, the dimensions of factors, the window size for text mining and the vector size. On some attempts, our NLP model even performs worse than just predicting as mean rating, since the model apparently learnt nothing. However, we still managed to train a good model that performs well.

4. RELATED WORK

In recent times, the recommender systems have considerable importance in academia, commercial activities, and industry. They are widely used in various domains such as shopping (Amazon), music (Pandora), movies (Netflix), travel (TripAdvisor), restaurant (Yelp), people (Facebook), and articles (TED).

Our dataset is established by Amazon and has been widely studied. In some papers, the vision has been broadly used in recommender system [1, 2]. Visual data has been trained by using deep convolutional neural network to calculate similarities to make predictions and research on fashion

3. PREDICTIVE MODELS

Since our dataset is Amazon Product Review, our research aims to make analysis based on users’ purchase history and their comment. Therefore, our goal is to establish a recommender model to predict the rating with user and item pair.

For baseline model, as we mentioned in section 2, is to make predictions according to the mean of ratings for all users and each user alone. This model makes prediction firstly based on user’s own mean rating with his history collected. Then if the user is new to model, the predicting rate would be mean of whole users.

Besides that, our first intuitive thought is simple latent factor model. The simple latent factor model considers the influence of both users and items.

G:\qq\598378694\Image\SharePic\20221128144914.png

Based on that, our next step is using complete latent factor model and SVD, since both models have much more complexity than original simple latent factor models.

C:\Users\Lenovo\AppData\Roaming\Tencent\Users\598378694\QQ\WinTemp\RichOle\YP_QYI54G0CA{AY[X{NP_5J.png

The formula is shown above. Comparing to simple latent factor model, this one assumes that there might be latent factors determining users’ decisions.

Latent factor models can perform well since they can fully learn biases and preferences and find those potential determining factors.

On the second hand, since the data contains not only user’s IDs, item’s IDs and ratings, but also user’s reviews and their summaries. Since the data contains abundant context information, our team also applied text mining models such as Word2Vec and Item2Vec.

While we apply and train our models, we encountered bunch of issues.

Firstly, we found that data are listed in a sorting order, which significantly increases the probability of the pairs of users and items being new to model. This will increase the case we predict by using rating mean. Plus, specifically, the text mining model will gain no similarities that makes the model collapsed. That is to say, we need to shuffle the dataset first to train models sufficiently

Secondly, since the dataset is huge, training usually takes the whole day, we shrink the training data, and then we encountered the issue of overfitting. The balancing seems a good means.

Our team modified our model by setting different hyperparameters like the learning rate, the regression lambda, the dimensions of factors, the window size for text mining and the vector size. On some attempts, our NLP model even performs worse than just predicting as mean rating, since the model apparently learnt nothing. However, we still managed to train a good model that performs well.

4. RELATED WORK

In recent times, the recommender systems have considerable importance in academia, commercial activities, and industry. They are widely used in various domains such as shopping (Amazon), music (Pandora), movies (Netflix), travel (TripAdvisor), restaurant (Yelp), people (Facebook), and articles (TED).

Our dataset is established by Amazon and has been widely studied. In some papers, the vision has been broadly used in recommender system [1, 2]. Visual data has been trained by using deep convolutional neural network to calculate similarities to make predictions and research on fashion trend. In specific domain like fashion, visual models can obtain better more useful information on users’ preferences.

On the other hand, another team focus on cross-domain recommendation (CDR) [3]. In this paper, they proposed using a knowledge graph to let different domains share knowledge.

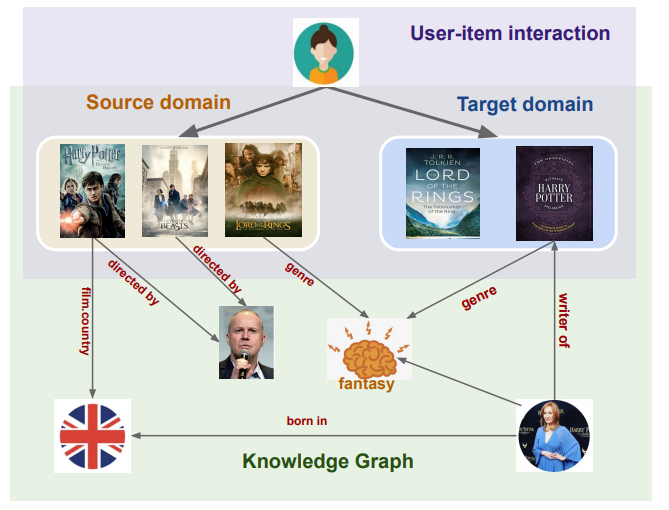


Fig 1 The Knowledge Graph

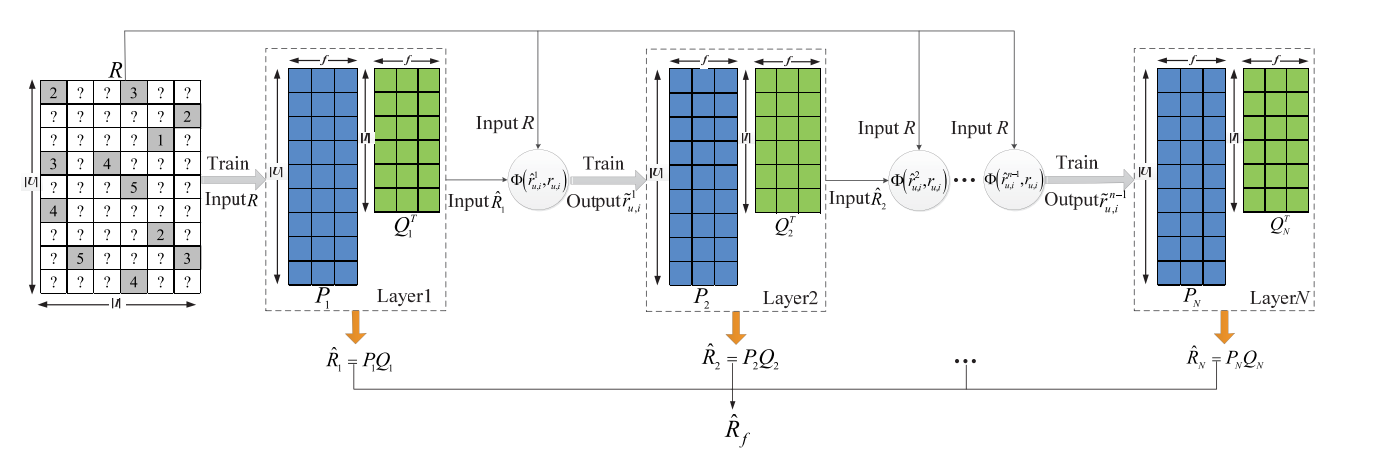


Fig 2 Deep Learning Method

However, our research mostly focuses on the rating predicting problem, therefore, we didn’t pay much attention on the graph data or different domains. Instead, the key to our model is content information and biases for users and items.

For latent factor model’s respective, there are a lot of great teams’ work on SVD.

Since the data are huge and such a matrix is usually high-dimensional and sparse with users and items exploding, Wu proposed a deep a deep-structured model by sequentially connecting multiple latent factor (LF) models instead of multilayered neural networks through a nonlinear activation function. [4].

The deep learning model is also studied by another team by introducing deep-structure in collaborative filtering model. Their work introduces the deep latent factor model (deepLFM) [5]. They found that the dimensions of factors cannot be too low, which might cause the model misses some important features, and on the contrary, it cannot be too high as well, which might introduces too much complexity.

This thought is similar to ours, since we find that in a large data pool we need

to appropriately set more factors to ensure our model have a better understanding of users’ preferences.

On the other field, some teams specifically studied content information by using text mining.

Ghuribi [6] found that collaborative filtering (CF) like SVD is a popular group of methods employed to build effective traditional techniques, while a content-based (CB) approach mines the appropriate recommendations for a user based on his recent behaviors according to what the user liked, bought or watched. Thus, they manage to combine both methods to obtain more useful information as more as possible.

Another team focus on text mining techniques as well by using named entities and topic hierarchies to exploit meaningful information in users’ comments [7].

5. CONCULUSIONS

F

REFERENCES

1. Mcauley J, Targett C, Shi Q, et al. Image-based Recommendations on Styles and Substitutes:ACM,10.1145/2766462.2767755[P]. 2015.
2. He R, Mcauley J. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering[J]. International World Wide Web Conferences Steering Committee, 2016..
3. Li Zhang, Yan Ge, Jun Ma, et al. Knowledge-aware Neural Collective Matrix Factorization for Cross-domain Recommendation [J]. <https://arxiv.org/pdf/2206.13255.pdf>
4. Wu, Di, et al. "A deep latent factor model for high-dimensional and sparse matrices in recommender systems." IEEE Transactions on Systems, Man, and Cybernetics: Systems 51.7 (2019): 4285-4296.
5. Mongia, Aanchal, et al. "Deep latent factor model for collaborative filtering." *Signal Processing* 169 (2020): 107366.
6. Al-Ghuribi, Sumaia Mohammed, and Shahrul Azman Mohd Noah. "Multi-criteria review-based recommender system–the state of the art." *IEEE Access* 7 (2019): 169446-169468.
7. Domingues, Marcos Aurélio, et al. "Exploiting text mining techniques for contextual recommendations." *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*. Vol. 2. IEEE, 2014.