Project Proposal

Neural Collaborative Filtering for Expert Recommendation

Mina Mousavifar, sem311, 11279515

The community question and answer (CQA) platforms, such as Stack Overflow[1], leverage the knowledge and expertise of users to answer questions posted by fellow users. Over time, these websites turn into repositories of knowledge. Knowledge acquisition and exchange are generally crucial, yet costly for both businesses and individuals, especially technical knowledge which covers a wide range of topics. CQA platforms offer an opportunity for sharing knowledge at a low cost, where community users, many of whom are domain experts, can potentially provide high-quality solutions to a given problem. However, in the era of information explosion, a large volume of questions are posted every day and it is challenging to identify the relevant qualified experts to answer these questions. To alleviate this information overload, I aim to recommend the experts who are most likely going to answer a given question in the Stack Overflow community. This task can be extended to other QA communities, such as peer learning systems.

Task

Personalized recommendation is based on modelling users' preference for items based on their past interactions, which is known as Collaborative filtering(CF)[2]. The most popular collaborative filtering technique is matrix factorization (MF), which utilizes a vector of latent features to represent a user or an item that can model user's interaction on an item as the inner product of their latent vectors[3]. This project will use a collaborative filtering based recommendation approach based on implicit feedback, that combines traditional matrix factorization and multilayer perceptron to perform expert recommendation in a community question and answer (CQA) platform[3][4].

Specifically, I will adapt the Neural matrix factorization model[2] for the expert recommendation. My proposed model will use the identity of a user and a question as the input features. The input will be first transformed into a binarized sparse vector with one-hot encoding. Afterwards, these inputs are fed to the embedding layer, which is a fully connected layer that projects the sparse representation to a dense vector, in which extracts user/answer latent feature vectors. Finally, this user embedding and item embedding are fed into a multi-layer neural architecture to map the latent vectors to prediction scores[3].

Inputs and Outputs

From the task formulation above, our proposed model uses the identity of a user and a question as the input feature which is further transformed into a binarized sparse vector with one-hot encoding. The output of our model is the predicted score, which is further used for ranking questions for an expert.

[1] https://stackoverflow.com/

[2] Neural Matrix Factorization model will further be referred by NeuMF

General Dataset info

This project will be based on StackOverflow(SO) dataset. Stackoverflow is an open community where developers share knowledge and skills through posting questions and answering each other's questions. It has a "data-dump" where its complied datasets are kept so that people can download it for research purposes. The dataset has the following files: Badges, Comments, PostHistory, PostLinks, Posts with a PostTypeId feature that is 1 for a question and 2 for an answer, Tags, Users, Votes. The dataset consists of 19m questions, 29m answers, 73m comments, and 57k tags.

Validation

In order to evaluate the performance of the expert recommender, I'll apply leave-one-out evaluation method[4]. In which, the latest interaction between user and question is kept off for the test set and the remaining data is used for training. Due to the massive amount of questions, ranking all questions for each user is highly time-consuming. Consequently, I'll randomly sample 100 questions that are not interacted by the user, and then rank the test item among these 100 questions[5]. Afterwards, rank-based metrics such as Hit Ratio(HR) and Normalized Discounted Cumulative Gain (NDCG)[6] can be applied to evaluate our model.

Related Paper

[1]Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. Proceedings of the 26th International Conference on World Wide Web - WWW 17 (2017). DOI:http://dx.doi.org/10.1145/3038912.3052569

References

[2]Badrul Sarwar, George Karypis, Joseph Konstan, and John Reidl. 2001. Item-based collaborative filtering recommendation algorithms. Proceedings of the tenth international conference on World Wide Web - WWW 01 (2001). DOI:http://dx.doi.org/10.1145/371920.372071

[3]Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. Computer 42, 8 (2009), 30–37. DOI:http://dx.doi.org/10.1109/mc.2009.263

[4]Trevor Hastie, Robert Tibshirani, and J.H. Friedman. 2009. The elements of statistical learning data mining, inference, and prediction, New York, NY: Springer.

[5]Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A Multi-View Deep Learning Approach for Cross-Domain User Modeling in Recommendation Systems. Proceedings of the 24th International Conference on World Wide Web - WWW 15 (2015). DOI:http://dx.doi.org/10.1145/2736277.2741667

[6]Xiangnan He, Tao Chen, Min-Yen Kan, and Xiao Chen. 2015. TriRank. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management - CIKM 15 (2015). DOI:http://dx.doi.org/10.1145/2806416.2806504