Digital Media Laboratory

Weekly Report October 10, 2021

Ali Jandaghi Alaee Report No. 11

alialaee98@gmail.com 00/07/12- 00/07/18

Abstract

In the article [2], several steps are introduced for the construction of a recommender system, based on available information, especially user-related ones. The aim is to include personalization to the conventional recommendation system methods. In the first two paragraphs each one is discussed, and in the last one the utilization of them is described.

Description

- Listing Embeddings
  At the first step, similar to the previous paper [1], the embeddings of listings (i.e. houses and hotels in this case) are learned with the help of skip-gram. The main difference is that during a search session, the sequence of clicks are observed, and then, every training sample would be a pair of adjacent clicks (listing ids). This differs from the method proposed in [1] which considers queries as vectors as well (our implemented method also works this way). A window size is defined, and the booked listings will appear as a positive sample for all the listings available in a session.

- User type and Listing type embedding
  This part has to do with the personalization, and the goal is to form another embedding space that associates users to listings. First, users should be divided into groups to form "user_types". The listings should be divided into groups of "listing_types" as well. The formation of the "user_types" takes place regarding user's meta-data (language, device, existence of profile picture, etc) and user's history (average spending for each night in previous bookings, etc). On the other hand, "listing_types" are formed considering listing meta-data (type, the region, number of stars it has, ratings in reviews, etc). The training procedure is still a skip-gram implementation. Training data is generated from the data set of booking histories. Positive samples are composed of tuples, in which a user_type has booked a particular listing_type. After the construction of the embedding space, for each user_type, several similar listing_types can be obtained, and this can help with the recognition of the desire of users.

Feature Name	Description
EmbClickSim	similarity to clicked listings in H_c
EmbSkipSim	similarity to skipped listings H_s
EmbLongClickSim	similarity to long clicked listings H_{lc}
EmbWishlistSim	similarity to wishlisted listings H_w
EmbInqSim	similarity to contacted listings H_i
EmbBookSim	similarity to booked listing H_b
EmbLastLongClickSim	similarity to last long clicked listing
UserTypeListingTypeSim	user type and listing type similarity

Figure 1: Features derived from Embedding spaces [2]

- Experiments [2]
  the final search rankings are obtained from the results of a gradient boosted decision tree (GBDT); its inputs are a set consisting of 104 features for each listing, which are of two main groups; First, features that are related to the query, users and listings (e.g. filters applied by user, number of rooms for listing and the device of the user). Second, results of two embedding space described above; listing embeddings

and (user_type, listing_type) embeddings. The first one computes the relevance of every listing item to the recently interacted items by user, and the second one computes the similarity of user_type of the user performing the search to listing_type of each listings

After the execution and evaluation of the GBDT, it turns out that embedding features have boosted up the performance of the model. In addition, a few of these 8 features, which are derived from embedding spaces and are illustrated in Figure 1, seem to be among the most important features of the GBDT algorithm.

Next Week

1. Deciding which type of personalization suits our project.

References

[1] Mihajlo Grbovic et al. Scalable semantic matching of queries to ads in sponsored search advertising. *39th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy*, 2016.

[2] H Cheng M Grbovic. Real-time personalization using embeddings for search ranking at airbnb. *KDD '18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2018.