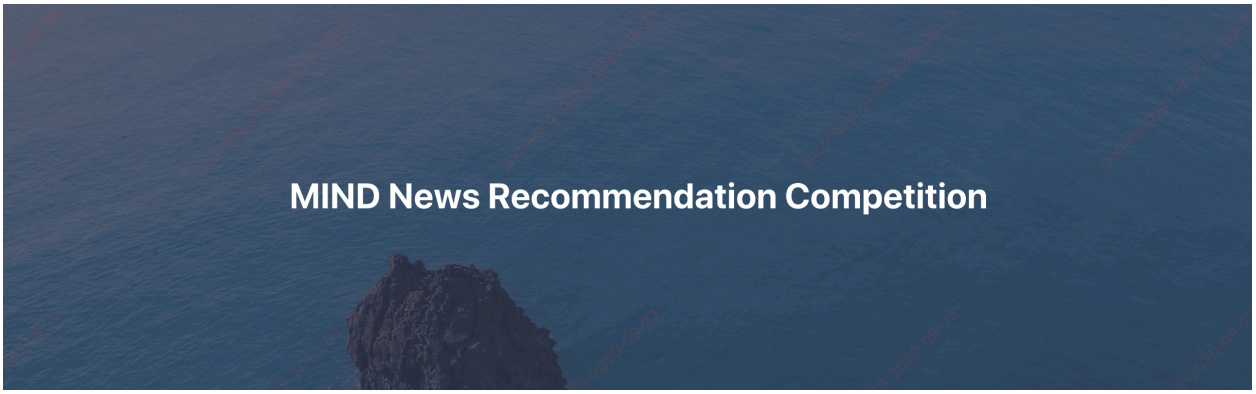


MIND News Recommendation Competition

Technical Report

--- 09.23.2020
--- Congcong Gu
--- PingAn

Introduction:

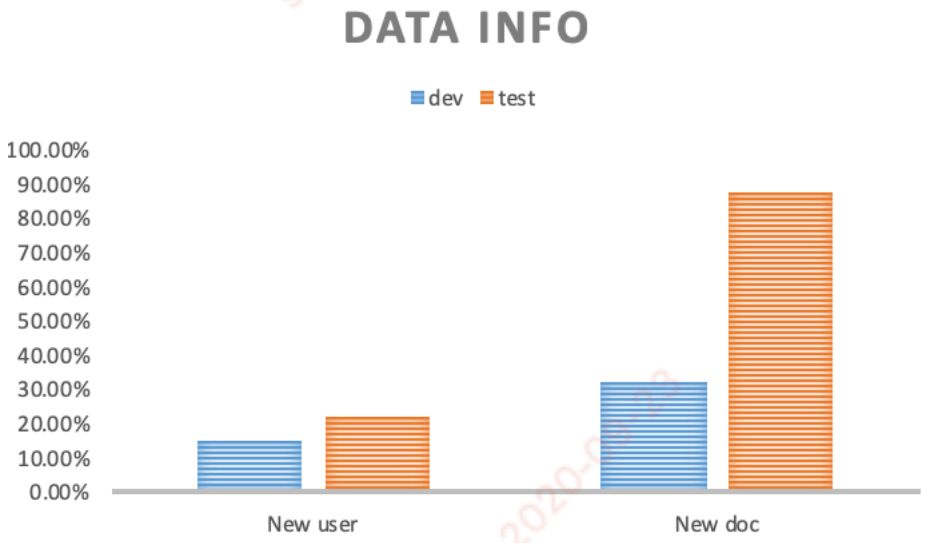
News recommendation is a challenging task。 First, many new articles are posted continuously, and existing news articles will disappear after a short period of time. So, it is a cold-start problem. Articles can't be represented by ID. Second, news articles usually contain rich information such as title and body. It is very important to understand news content from their texts using NLP techniques. Each user will pay different attention to the same article. Third, user interests change quickly. Users who read news are characterized by rapid interest transfer. So, it is a hard thing to catch user interest.

Evaluation Metrics

The primary metric for submission ranking is **AUC**.

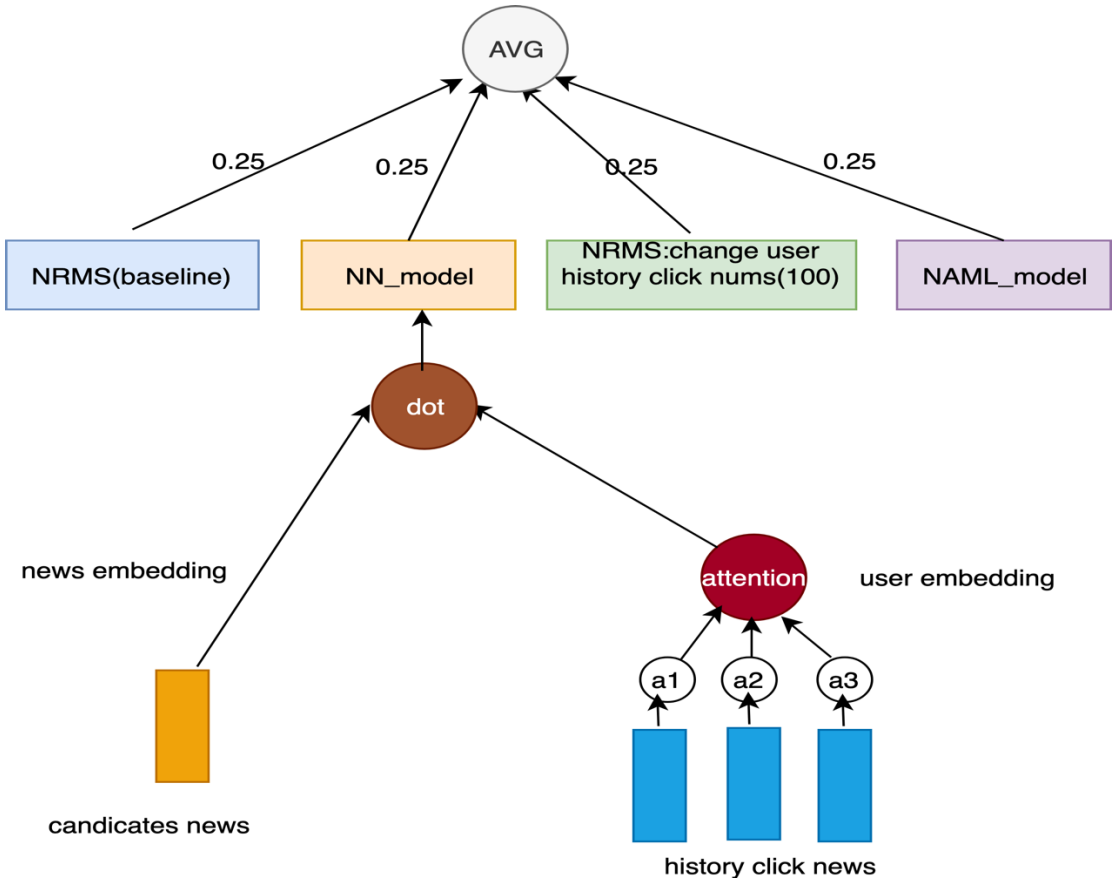
Dataset explore

1. There are huge difference between dev set(only one day) and test set(one week). Looking at the data set, I found a huge difference between the Dev set and the Test set. The dev set only one day , but the test set have seven days.



2. Auc reflects the ranking ability of the overall sample. But what we actually measure is the ranking ability of different users to different news, and what we actually pay more attention to is the ranking ability of the same user to different news.
3. The data set does not include the time and body of the article, which needs to be crawled by yourself.

Model structure :



Milestones:

Model description	Test auc
Baseline(NRMS model)	0.6817
Model_A(baseline+ change user history click nums(100))	0.6886
0.5*Baseline+0.5* Model_A	0.6719
0.3*Baseline+0.35* Model_A+0.35* Model_B(naml)	0.6741
0.25*Baseline+0.25* Model_A+0.25*Model_B(naml)+0.25*NN_model	**0.6974**

do not work on the test:

In the development phase, I use yesterday's data to make a lot of features such as news ctr, users' click count, and so on. But in the test phase, these features are invalid. Because the next 7 days don't know yesterday's data. Besides, I user LGB model in the development phase get 0.6922 score. So, it’s a pity that I spent a lot of time on feature construction. The integration of tree model and deep learning model will be greatly improved.

Some Ideas:

1. Downsampling of negative samples. Only about 4% positive ratio in test dataset. Ensuring that the proportion of positive and negative samples is close will make the model have stronger anti-overfitting ability
2. Use all available contents informations. Content information contains text and image information. Using multimodal can improve the accuracy of recommendation system.
3. Adds dev data for training.

Thanks

Thanks to the mind organizer for holding a great competition and providing the large mind dataset which also can do some researches about recommendation.Besides, I’m very fortunate to have the chance to focus on the rank model in Baidu feed team before and now in PingAn content recommendation team. And great thanks to all my teammates. Also learn a lot from the excellent open source code Microsoft provides(https://github.com/microsoft/recommenders) and papers about NRMS, NAML,NPA models and so on.