A HYBRID APPROACH FOR NEWS RECOMMENDER SYSTEM USING OPTIMIZATION METHODS

A PREPRINT

Alexander Smirnov
Department of Information and Analytical Systems
Saint Petersburg State University
Russia, Saint Petersburg
ru.alexander.smirnov@gmail.com

Elena Mikhailova
Department of Information and Analytical Systems
Saint Petersburg State University
Russia, Saint Petersburg
e.mikhaylova@spbu.ru

January 12, 2021

ABSTRACT

Recommender system is an essential part of any social media application. Most recommender systems now use a hybrid approach, combining collaborative filtering, content-based filtering, and other approaches. Most common problems in the field of hybrid recommenders are cold start and data sparsity [Çano, 2017]. In this paper we address the abovementioned problems by proposing a hybrid weighted news recommender system which combines different approaches.

Keywords hybrid recommender systems · content-based recommender · collaborative recommender · optimizations

1 Introduction

Recommender system is a crucial part of every application that operates with content and user activity. Enormous amount of information leads to the problem that user is not able to find relevant content.

Common approaches, such as collaborative filtering, has its own problems: cold start, scalability and data sparsity. Content-based approaches suffer from the fact that we have to somehow represent recommended item in feature space.

To be consistent during the paper we list some domain specific vocabulary with their meanings:

- Rating: expression or preference
  - explicit (direct from user, e.g. user rated film)
  - implicit (inferred from user activity, e.g. user stopped watching movie after 5 minutes)
- Prediction: estimate of preference
- Recommendation: selected items for user
- Content: attributes, text, etc; everything about item

The remainder of this paper is organized as follows:

- Section 2 describes the relevant related work
- Section 3 describes input data
- Section 4 explains our modular design and architecture
- Section 5 describes the implementation of the algorithms in a real system
- Section 6 provides tests and experiments validating our systems results
- Section 7 explains future work
- Section 8 presents conclusions

2 Related work

According to the study [Dacrema et al., 2019], deep learning techniques are not supposed to beat conceptually and computationally simpler algorithms, so we won't touch them.

Our goal is to choose optimal algorithm for each of the following tasks:

- **Collaborative filtering**: generating predictions about the interests of a user by collecting preferences or taste information from other users. It is based on the assumption that if a person A has the same opinion as a person B on an issue, A is more likely to have B's opinion on a different issue than that of a randomly chosen person
- **Content-based filtering**:
- **Session filtering**:
- **Popularity filtering**:
- **Demographic filtering**:
- **Time-based filtering**:

We face the problem that millions of news are theoretically suitable for being recommended, so it is not correct to use abovementioned methods on such large corpus of data. Instead, as stated in [Covington et al., 2016] paper, we want to implement `candidate generation` $\rightarrow$ `ranking` pipeline to reduce number of candidates.

2.1 Collaborative filtering

There were many studies on this topic, but paper [Rendle et al., 2019] proves that well-tuned basic SVD++ approach beats newely presented algorithms.

3 Input data

As we solving domain specific task, we have domain specific data.

metadata

item_id	*date*	*source_id*
1	2021-01-08 22:08:39	9
2	2021-01-09 10:28:58	5
3	2021-01-09 14:20:34	12
⋮	⋮	⋮

Table 1: news metadata

- *source_id*: source of news item

content

item_id	*news title*	*news content*
1	Azerbaijan denies reports on construction of Turkish air bases in the country	Information that Turkey will create air bases ...
2	Durov announced the massive transition of WhatsApp users to Telegram	Telegram developer Pavel Durov said in his channel ...
3	Passenger plane that disappeared from radar crashed	Passenger plane taking off from Jakarta, disappeared ...
⋮	⋮	⋮

Table 2: news item

shows & views

- *shows*: if *item_id* was shown to the *user_id*

user_id	*item_id*
10	1
10	2
23	1
23	3
23	2
38	3
38	1
⋮	⋮

Table 3: news shows

user_id	*item_id*
10	1
10	2
23	1
38	3
⋮	⋮

Table 4: news views

- *views*: if *item_id* was clicked by the *user_id*

emotions & comments

There is an option to react on item via leaving emotion and/or writing a comment.

user_id	*item_id*	*emotion_id*
10	1	1
10	2	3
23	1	3
38	3	2
⋮	⋮	⋮

Table 5: users' emotions

user_id	*item_id*	*comment*
10	1	that's great
10	1	wish it will continue
23	2	whatsapp is not competetive anymore
⋮	⋮	⋮

Table 6: users' comments

- *emotion_id*: one of {😄, 😲, ☹, 😣, ♥ }

users' subscriptions

If *user_id* subscribed to the *source_id*.

user_id	*source_id*
10	9
23	5
⋮	⋮

Table 7: users' subscriptions

4 Overview of our approach

Out goal is to combine state-of-the-art approaches in recommender systems.

Our approach consists of combining several techniques:

- **Collaborative filtering**:
- **Content-based filtering**:
- **Session filtering**:
- **Popularity filtering**:
- **Demographic filtering**:

- **Time-based filtering**:

Solution consists of 2 parts:

- Candidate generation
- Ranking

4.1 Architecture overview

4.2 Components overview

4.2.1 Collaborative filtering recommendation

We are using SVD++ algorithm.

For collaborative filtering recommendation we should have something known as user-item matrix which may be formed from user activity from tables (cite tables)

4.2.2 Popularity-based recommendation

For measuring news item popularity following data can be aggregated: shows, views, emotions, comments.

item_id	*shows_num*	*views_num*	*emotions_num*	*comments_num*
1	1043	231	52	7
2	828	478	78	11
3	163	25	5	0
⋮	⋮	⋮	⋮	⋮

Table 8: aggregated popularity data

5 Implementation

6 Evaluation

7 Further research

8 Summary

Results show that combining different approaches leads to rise of users' involvement.

References

Erion Çano. Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*, 21:1487–1524, 11 2017. doi:10.3233/IDA-163209.

Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. Are we really making much progress? a worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems*, RecSys '19, page 101109, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450362436. doi:10.1145/3298689.3347058.

Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, New York, NY, USA, 2016.

Steffen Rendle, Li Zhang, and Yehuda Koren. On the difficulty of evaluating baselines: A study on recommender systems. 05 2019.