

# NewsBoard Recommender System

Ruofan Chen Shanchao Wu

This project is a demo-based project, whose goal is providing a news board to users every day along with a news search system embedded in it. The final project will be available on GitHub [[https://github.com/RuofanChen/IR\\_News\\_Board](https://github.com/RuofanChen/IR_News_Board)].

## **What is exactly the function of your tool (or a method)? That is, what will it do?**

This project will have a web application built for users to sign up and subscribe to the news service. After signing up, while subscribing, users will need to answer some questions for news topic/category preference and accept to receive the news board through their email. Every day at 8:00 am, users will receive a news board sent through email. The news board includes two main parts. The first part is a click behavior-based news recommender system, which will display the top news (shown as news title and abstract with the news link) which is collected by New York Times API (<https://developer.nytimes.com/apis>) from different categories. The news recommended to users will be updated every day. For those new users, recommended news will be assigned based on the user's sign-up preference, then adjusted the recommended news based on the user's everyday click-behavior.

The second part is **optional** (depending on the time, some functions may be provided).

The second part is a news search function and news recommendations based on user searches. We also consider the case where the user has an explicit interest in some news that is not on the news board (Part 1), in which case we provide a search bar for the user to search for news. The result will be returned in the form of a news title and the link in the order by relevance. If the user chooses to follow the news/topic, then the relevant news will be displayed aside from the main news board. Users can unfollow the topic at any time. After unfollowing, no news will be displayed on the side.

## **Why would we need such a tool (or a method) and who would you expect to use it and benefit from it?**

Everyone can use the news board tool to keep track of news updates every day. This news board is a personalized recommendation, not only focusing on giving news topics that users are interested in, but also giving various other topics to cover hot news, which is different from other news tools.

Some professionals, such as investors, can follow relevant target industry news. Usually, industry news is related to industry prospects or has an impact, such as some released policy, which can help them make decisions.

Additionally, this tool will be particularly beneficial to international students as it is an additional resource for receiving news updates beyond TV and radio, as not all international students have the access to watch TV and listen to car radio news.

**Do these kinds of tools/methods already exist? If similar tools/methods exist, how is your tool/method different from them? Would people care about the difference?**

We don't find such a news board exists for free at this time.

Some news corporations have a news subscription service, e.g. New York Times and Apple. However, most news subscription services require a subscription fee. Some users are price sensitive. A free news app would be a good choice for them. The New York Times recommends several news stories to display in full text, but our news shows headlines and abstracts with news links for users to choose from.

**How hard is it to build such a tool/algorithm? What is the challenge? How do you plan to build it? You should mention the data you will use and the core algorithm that you will implement (either an existing algorithm for tools or a new algorithm for methods).**

We are considering the following challenges:

1. Find an appropriate news API that satisfies our design. After looking up APIs and some experimental tests, we choose to use the New York Times API to crawl the news. After grabbing the raw news data in JSON format, pre-processing needs to be done to serve the needs.
2. Save the news into the database (SQLite or MongoDB), insert news every day, set an update window to drop the news that is more than 10 days.
3. Calculate the similarity of upcoming news with the existing news (tf-idf, cosine similarity)
4. Use Latent factor models for news board: Matrix Factorization - decomposes the user-news rating matrix ( $n \times m$ ) into two lower-dimensional matrices representing news features ( $n \times k$ ) and user features ( $k \times m$ ), use SGD to minimize the loss function. Evaluate the model with RMSE, NDCG.
5. Use tf-idf, BM 25 for news search functions, return ranked results. If the user chooses to follow the topic/specific news, the user will also receive recommendations on the following topic. **(optional)**
6. Build a nice and neat web page for users to sign up and subscribe.
7. Create an email delivery system
8. Record the user click information
9. Design the news board.

Our plan:

| Task  | Timeline | Task will be done by                               |
|---|----------|--|
| News crawling and preprocessing   | Mar 23   | Shanchao   |
| List 1: Design the database, set update window<br>List 2: Save news                     | Mar 23   | Ruofan: List 1<br>Shanchao: List 2                 |
| Build Latent factor model<br>List 1: set initial value and do Matrix Factorization, use | April 10 | Ruofan: List 1, List 4<br>Shanchao: List 2, List 3 |

|  |          |          |
|--|----------|----------|
| SGD, train hyperparameter factor k<br>List 2: update new item(news): tf-idf cosine similarity, weighted sum old news vector<br>List 3: update new user, initial value based on preference (if exist), else use total average rate/click<br>List 4: pick news to recommendation |          |          |
| Build sign up page and user database   | April 13 | Ruofan   |
| Design news board page   | April 20 | Both     |
| Email delivery   | April 20 | Shanchao |
| Test the application   | April 23 | Both     |

### What existing resources can you use?

New York Times API (<https://developer.nytimes.com/apis>) and web app modules.

News recommender system: a review of recent progress, challenges, and opportunities

<https://link.springer.com/article/10.1007/s10462-021-10043-x#:~:text=In%20order%20to%20help%20users,interest%20for%20the%20news%20readers.>

A Singular Value Decomposition Approach For Recommendation Systems: chapter 4

<https://open.metu.edu.tr/bitstream/handle/11511/19640/index.pdf>

User authentication using Django framework

<https://docs.djangoproject.com/en/4.0/topics/auth/default/>

### How will you demonstrate the usefulness of your tool/method?

To evaluate the efficiency of our methods, we will first simulate 3-5 types of users. They will share different tastes on main news preferences and overlap in some categories/topics.

After that, we will find some volunteers to participate in the application and function testing. Surveys will be sent to the subscribed user to collect the user's feedback, including their scoring of the subscription service, and the functions that need to be improved. In addition, the user-click number is also a good indication of the application. We can keep track of their click of the news sent by email.