

Group 49: Feedback-based news recommendation system

REPORT

Samiksha Garg (MT21074) Sakshi Sinha (MT22121) Shambhavi Pathak (MT22067) Debnath Kundu (MT22026)
Snehal Buldeo (MT22074)

Problem Statement:

The project's objective is to create a new recommendation system (**NRS**) that can identify bias in suggested news items, notify readers who could be at risk of entering a *filter bubble*, and recommend news from a diverse topic if they want. To mitigate the filter bubble, the system should be able to identify when a user is reading a majority of news articles that are biased towards a certain topic or theme and allow them to break out of it. If the user chooses to do so, the system should recommend random articles that are more diverse in perspective and less biased. Filter Bubble refers to a situation where people are shown a narrow and personalized content selection based on their past behaviour, preferences, and interests. In the context of news recommendations, users are less likely to encounter news stories outside of their interests.

The risk of entering a filter bubble is due to a personalized recommendation system that certain applications tend to use to provide a better user experience by providing the users with news articles tailored to their individual preferences and interests. In contrast, there is a responsible news recommender system that is designed to prioritize the accuracy, credibility, and diversity of news articles, to provide users with a more balanced and trustworthy news experience. Responsible news recommenders work by attempting to surface news articles that represent a diverse range of perspectives and viewpoints unknown to a user. With our work, we want to alert the reader about the biases in their reading and recommend diverse topics with their consent. The existing applications notify the users about the bias in the news or of the news source. On the other hand, we tend to identify the biases in a user's reading pattern by collecting implicit feedback such as clicks.

Updated Literature Review:

The filter bubble phenomenon is a common concern in news recommendation systems and it occurs when the system narrows the information and deprives users of diverse information. Once the recommendations have been made, they need to be evaluated, to check the diversity in the recommended items. A diversified list more likely contains the user's actual search intent [8]. Despite the rise of interest and work on the topic in recent years, we find that a clear common methodological and conceptual ground for the evaluation of these dimensions is still to be consolidated [9]. To measure the diversity of a recommendation list, a common evaluation measure is the average pairwise distance between items in the ranked list. This measure is called intra-list diversity (ILD) and a high value in ILD means the recommended list contains items with a broad range of content [8]. introduced a rank and relevance-sensitive intra-list diversity measure (RR-ILD) that shows to what extent the recommender can diversify the list and preserve the relevant items in the high ranks. The measures used to calculate diversity for a list of items depend on the type of recommendation system.

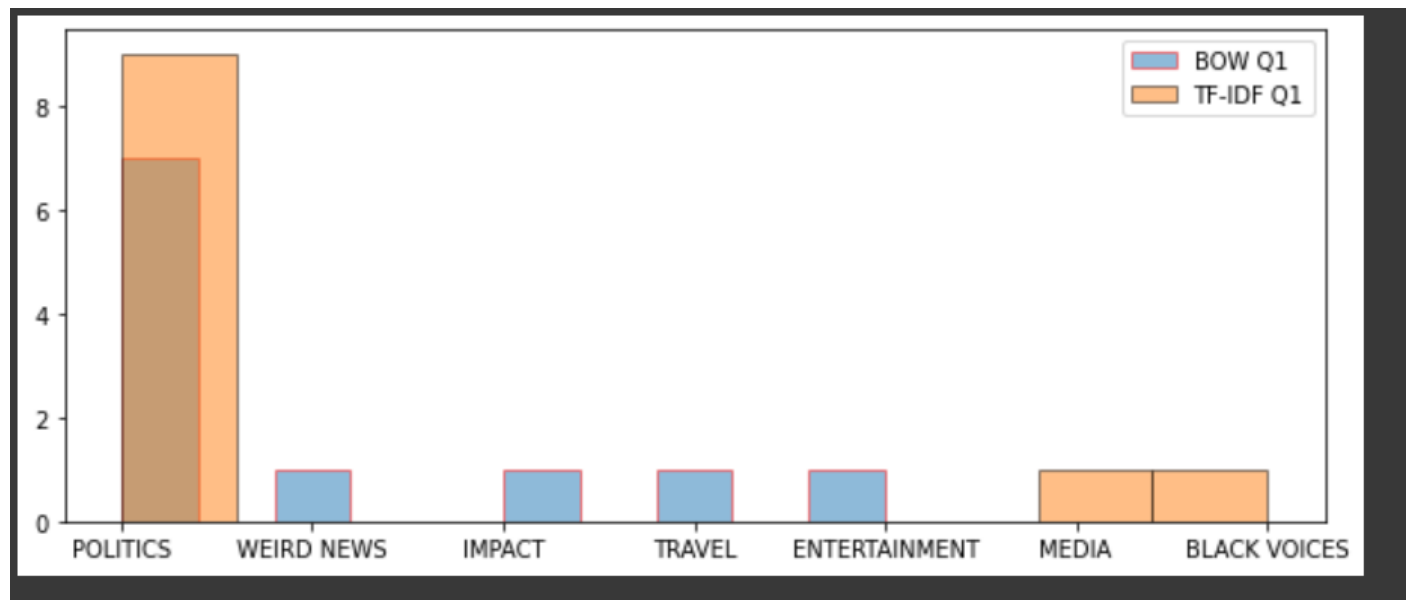
As we want to introduce diversity into an already recommended list of articles, it focuses on one user and one set of recommended articles. These articles are checked for diversity based on their content and headline.

The paper "A Serendipity Model for News Recommendation" presents a model for improving news recommendation by incorporating serendipity into the recommendation process. The authors define serendipity as a desirable property of recommendation systems that involves suggesting unexpected but relevant items to users. The proposed model uses a hybrid approach that combines content-based filtering and collaborative filtering. The authors also introduce a serendipity score that is used to measure the level of serendipity in the recommended news items. The score is based on the novelty and diversity of the recommended items, as well as their relevance to the user's interests. The authors evaluate the proposed model using a dataset of news articles and user interactions. The results show that the model improves the serendipity of the recommended news items while maintaining a high level of accuracy in terms of relevance to the user's interests. Overall, the paper presents an interesting approach to incorporating serendipity into news recommendation systems, which could help users discover new and interesting news articles that they may not have otherwise come across.

The paper "Improving Recommendation Lists Through Topic Diversification" explores the problem of recommendation systems that are unable to recommend a diverse range of items to users. The authors note that traditional recommendation algorithms often recommend the same popular items to a large percentage of users, which can lead to a lack of exploration and discovery. To address this issue, the authors propose a topic diversification algorithm that recommends items from a variety of topics, rather than just the most popular items. They evaluate the proposed algorithm on a dataset of movie ratings and show that it outperforms traditional recommendation algorithms in terms of diversity while maintaining comparable accuracy. The paper provides insights into the challenges of recommendation systems and presents a novel solution that can improve user satisfaction by providing a more diverse set of recommendations.

Updated baseline results (system/prototype):

As per the baseline model, our results from the Query of the user. The results from the Bag of Words, and Tf-Idf were as follows:



We did a diversity measure on the following recommendations and the values were as follows:

BOW Query 1	BOW Query 2	Tf-Idf Query 1	Tf-Idf Query 2
0.438	0.306	0.314	0.24

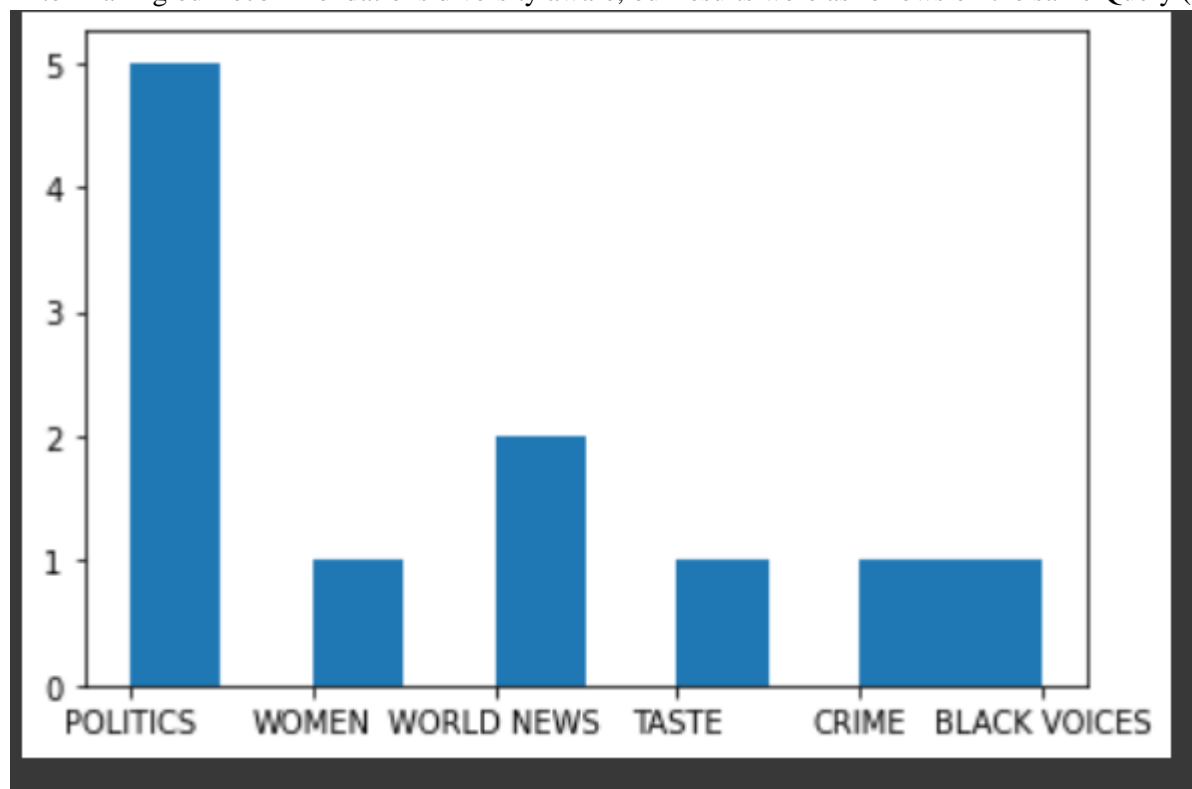
Metric used, ILD (Intra-List Diversity) which is calculated as the 1-ILS where ILS is Intralist Similarity. ILS is the average cosine similarity of all items in a list of recommendations. This calculation uses features of the recommended items (such as movie genre) to calculate the similarity.

NOTE:

How do we read an ILD measure?

It varies from 0 to 1. Higher the value more is the diversity in the recommendation.

After making our recommendations diversity aware, our results were as follows on the same Query (Q1)



The ILD measures were also better than before:

BOW Query 1	BOW Query 2	Tf-Idf Query 1	Tf-Idf Query 2	BOW with random function
0.438	0.306	0.314	0.24	0.727

The dataset used for testing had the below features:

- category: category of the news
- headline: headline of the news article
- link: source link of the news
- authors: Author of the article

The test data was of size 200853 * 4

The diversity in the news article was introduced with respect to the category.

The proposed method (features/ data analysis) [need not be fully completed]: Propose a method for solving the problem statement. Please provide details on the features and data analysis techniques you plan to use in your proposed method. You do not need to complete the proposed method at this stage fully. However, you should provide enough information to demonstrate a solid understanding of the methods and techniques required to solve the problem.

Our entire system has two parts of recommendations. One which recommends items based on the user's choice (choice of topic) and the other which will recommend a diverse range of news to the user. To implement the second recommendation system, we will be introducing topic diversification for the users who are willing to break out of their bubble of reading similar items. To introduce topic diversification we will be implementing Item based Collaborative Filtering.

Our system will be only restricted to a single user profile and on initial startup of the web app, the user will be recommended the news of its choice. Its choice will be collected by clickstream data and few selections on the UI screen. After User reads a series of 5 articles which are similar (have low ILD value) a pop-up on the UI shows asking if the user would like to explore other topics.

On clicking yes, the Item based Collaborative Filtering model will recommend diverse topics and the user can further read it.

On clicking no, the original recommendations will continue.

Future Work:

1. Evaluation:

The performance of the content-based news recommendation system built into this project can be evaluated using various evaluation techniques. The system's effectiveness can be measured using metrics such as **precision, recall, and F1-score**, which can provide insights into the system's ability to recommend relevant articles. With evaluation, we can identify areas of improvement and refine the recommendation algorithms to provide even more relevant recommendations to users.

2. Collecting Real-Time Data:

Currently, the system uses a pre-collected dataset of news articles, but collecting news data in real-time can ensure that the system is always up-to-date with the latest news. An approach that we explored is to use the **Inshorts API** to collect news articles. By combining this with the recommendation system, we can create a more comprehensive news recommendation system that provides real-time recommendations based on the latest news articles. This can improve the

relevance and usefulness of the system, as users can receive recommendations based on the most recent news articles available.

3. **Collecting implicit feedback:**

Collecting implicit feedback such as clicks allows us to identify the user's reading pattern and preferences without requiring explicit user input. By analyzing the user's clicks, we can determine their interests and biases and use that information to personalize their news recommendations. This approach can help create a more accurate and effective recommendation system while also respecting the user's privacy.

4. **Serendipity Model:**

We also want to allow users to break out of their filter bubbles and explore different content. To do this, we can implement a serendipity model, which recommends random articles to users who choose to break out of their usual content. This can help users discover new topics and perspectives they might not have encountered otherwise.

Conclusion:

- Our study examined the effectiveness of a content-based recommendation system for news articles based on their attributes.
- The system used two popular content-based recommendation techniques: Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF).
- The goal of the recommendation system was to recommend news articles similar to a given article based on their attributes, such as article headline, category, author, and publishing date.
- Our study showed that BoW and TF-IDF techniques could effectively recommend news articles based on their attributes. The BoW technique is simple and effective, while the TF-IDF technique considers the frequency of words and produces more relevant recommendations.
- However, both techniques have their limitations and do not capture the semantic and syntactic similarity of words. Word embedding techniques such as Word2Vec, GloVe, and fastText can capture the semantic similarity between words and address the limitations of BoW and TF-IDF techniques.
- The choice of recommendation technique should be based on the specific requirements and characteristics of the data and application domain.
- Based on the analysis conducted in this study, it can be concluded that incorporating diversity into recommendation systems can lead to improved user satisfaction and engagement. In particular, the use of the Intra-List Diversity (ILD) metric can provide a useful measure of the diversity of recommendation lists, allowing for the identification of potential areas for improvement. In our study, we calculated the average of all rows in the cosine similarity matrix to obtain the ILD score, and also used the sum of pairwise distances as a supplementary measure. Overall, our findings suggest that incorporating diversity into recommendation algorithms and utilizing metrics such as ILD can lead to more effective and satisfying recommendation systems for users.

References

1. [News recommender system: a review of recent progress, challenges, and opportunities | SpringerLink](#)
2. [\[2106.08934\] Personalized News Recommendation: Methods and Challenges](#)
3. [Fairness in Recommendation: A Survey](#)
4. [News recommender system: a review of recent progress, challenges, and opportunities | SpringerLink](#)
5. [A Serendipity Model for News Recommendation | SpringerLink](#)
6. [Personalized News Recommendation Based on Click Behavior](#)
7. [FairMatch: A Graph-based Approach for Improving Aggregate Diversity in Recommender Systems](#)
8. [Diversification in session-based news recommender systems | SpringerLink](#)
9. [Rank and relevance in novelty and diversity metrics for recommender systems](#)
10. [Improving Recommendation Lists through Topic Diversification](#)