

User based News Recommendation system using Twitter

Afzal Patel¹, Shreyas Paradkar², Tushar Parmar³

^{1,2,3} B.E Computer Engineering student, Vidyavardhini's College of Engineering and Technology, Mumbai, Maharashtra, India

Abstract— Number of websites over the internet display news articles to variety of users. There is a huge amount of irrelevant articles shown, which may not appeal to the user. Hence, to deal with this surge, articles are recommended in light of client interests as opposed to introducing articles by their event. A User based News Recommendation system is implemented using micro blogging activity on twitter. News articles are sorted based on the popularity based on tweets for specific subjects. This system ranks the news stories on the basis of article's popularity with the help of tweets and then displaying relevant stories based on user profile which contains user ratings. By combining these two approaches, hybrid news recommender is implemented to get relevant stories.

Key Words: popularity weight, user profile, recommender, tweet, similarity, rss.

1. INTRODUCTION

In today's world, huge amount of information is available at one click, which is accessible to users all over the world. The prime objective of the user is to find useful information from a very large data source. This has led to the evolution of the recommender systems that presents user specific contents based on their choices.

A recommendation system automatically suggests content for website readers and users. Major example is how shopping websites like Amazon offers customized recommendations to each user. Additionally, Netflix recommends movies to its users and music apps such as Wynk music or Saavn recommends music based on users' past history and preferences. Moreover, there exists various online news reading websites, such as Yahoo news, The Times of India etc. With so much news available all over the internet at different sources, the basic problem is to identify and recommend the most interesting articles to a particular user which will ensure that they are not presented with irrelevant and huge amount of data. These articles should be based on user choices as well as those news stories which are trending across the world.

We aim at developing a news recommendation system which displays news stories not only based on user's interests but also using data regarding trending twitter topics. This system ranks the news stories on the basis of article's popularity with the help of tweets and then displaying relevant stories based on user profile which contains user ratings.

2. RELATED WORK

2.1 Recommender Systems

User profiles are created by collecting information from the users by the recommendation systems. Recommendations are generated using these user profiles. Recommendation system may work on two criteria, firstly according to the ratings provided by the user. In dividing line to this approach user's behavior may be monitored to provide better recommendations [2][3]. Types of recommender systems are as follows:

- Collaborative recommender systems: User behavior and activities are collected and analyzed in this approach. It assumes that people will like similar kinds of details as they liked in the past.
- Content based recommender systems: It analyzes item descriptions to identify items that are of particular interest to the user.[4]

2.2 Popularity Based News Recommendation System:

The growing flood of information makes it difficult for the user to access the most popular news articles on the web. The popular news items can be determined using social networking sites such as Twitter and Facebook. A huge amount of information on current events can be generated as they are instant and widely available. However, the quality of data may vary as content is not moderated.

One of the popularity based news recommender is the *Buzzer*. It harnesses real-time micro-blogging activity such as Twitter data to provide news recommendations. It takes advantage of a novel content-based approach for finding news stories among a users' set list of RSS feeds. Thus articles are ranked as per the similarities to provide most trending news stories.[5]

2.3 User Profile Based News Recommendation System:

Articles are presented to the user solely on the basis of his/her interests in Profile based news recommendation system. Pazzani et al. made News Dude, an individual news-prescribing operator that utilizes TF-IDF in blend with a Nearest Neighbor calculation to achieve the end goal to prescribe news stories to clients [6]. Wouter et al. additionally depict ontology based strategies to prescribe news articles to the clients relying upon their interests [7].

3. PROPOSED SYSTEM

Displaying interesting news stories to the user by merging user preferences and hot topics is the idea behind hybrid news recommendation system. A web form is used to collect user preferences to generate user profile and popularity of an article is calculated using similarity with trending tweets. We assume that users wish to see news stories related to topics in their profile that are also creating buzz around the globe.

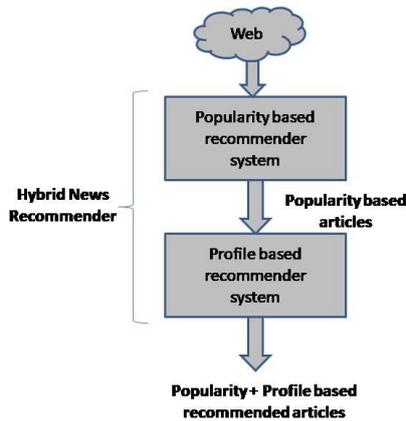


Figure 1: Hybrid News Recommender System Architecture

The system consists of three modules:

1. Popularity-Based News Recommender
2. Profile-Based News Recommender
3. Hybrid News Recommender

Figure 1 shows architectural diagram for hybrid news recommender system. Articles collected from the internet are first compared for similarity with tweets to get a weight. Articles are then arranged according to this weight and displayed to the user based on the preferences entered into the system.

3.1 Popularity Based Recommendation System:

RSS (Really Simple Syndication or Rich Site Summary) feeds are used to collect news articles from the web. It is used to collect news stories from different news sources in xml rss format. The RSS articles are collected from a news source such as TOI or NDTV that organize their stories by categories like Sports, Business, Health, Nation, Education, and Entertainment etc. These RSS feeds are pre processed to remove unnecessary contents (html tags, xml data etc) and store the article's title, link and category in the database. News from all different sources are compiled in a single database.

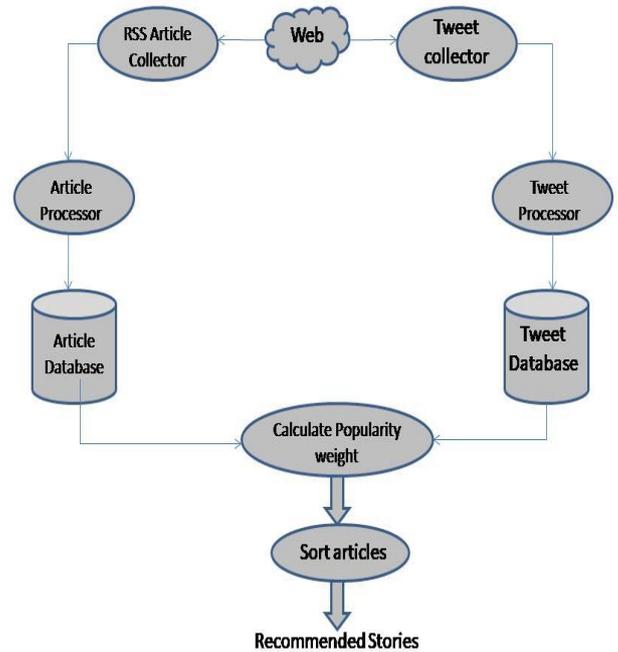


Figure 2: Popularity Based News Recommender Architecture

Figure 2 shows Architecture of popularity based recommender system. Data from micro blogging site twitter is used to identify popular news stories. Twitter's streaming API is used to request tweets from twitter's public timeline. Tweet collector collects the tweets using the authentication key. The tweets collected are initially in JSON (JavaScript Object Notation) format. Further, the tweet processor eliminates unwanted content from JSON tweet and stores only tweet id, date, and Tweet text in the database. Single article title is compared for similarity with a single tweet text to generate similarity weight. Each news title from RSS feed database is compared with all the tweets to find a similarity value known as popularity weight. The popularity weight is the average of all similarity weight between an article and a tweet.

Figure 3 shows the flowchart to compare article title and tweet to get similarity weight. In this algorithm, article title and tweets are pre processed to remove white spaces, and special characters. Strings are then compared for similarity and word count is incremented after finding each matching word. Common English words (is, the, an, but etc) are ignored while comparing. Finally similarity weight is calculated using the formula:

$$\text{Similarity Weight} = \frac{\text{No of words matched}}{\text{Total No of words in article title}}$$

Popularity weight is calculated according to the following formula:

$$\text{Popularity Weight} = \frac{\sum_{i=1}^N \text{Similarity Weight}}{N}$$

Where, N = Number of tweets in the database.

Popularity weight is the average of N similarity weights. Hence, each article has a corresponding popularity weight. Then the database is sorted according to the popularity weight in descending order and top news articles are displayed in sequence of the popularity of each article.

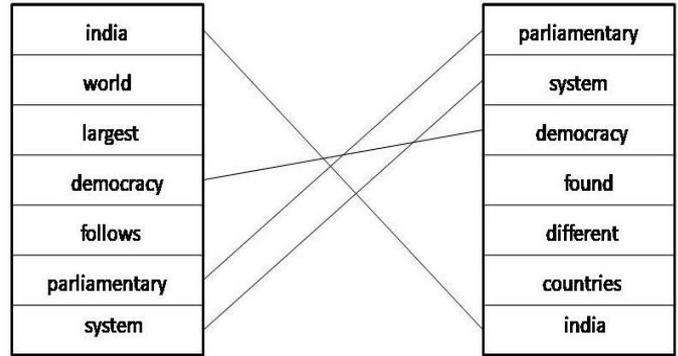


Figure 4: Comparison of 2 strings by matching words

Figure 4 shows an example for comparison of matching words between 2 strings. Here, Number of matches = 4 and Total no of words = 7

Therefore, Similarity weight = $4/7 = 0.571$.

3.2 User Profile Based Recommendation System:

News articles are presented to the user according to entered choices. A web form is presented to the user which contains name of the category and space to fill rating criteria. The number of articles presented to the user is directly proportional to the rating entered by the user. In this system ratings range from 1 to 5, where 1 is the lowest rating while 5 being the highest. Data from RSS feeds is pre processed to store title, category and link. The category stored in database is compared with the user profile and according to the best matches, articles are recommended to the user. Figure 5 shows architecture of profile based recommender system.

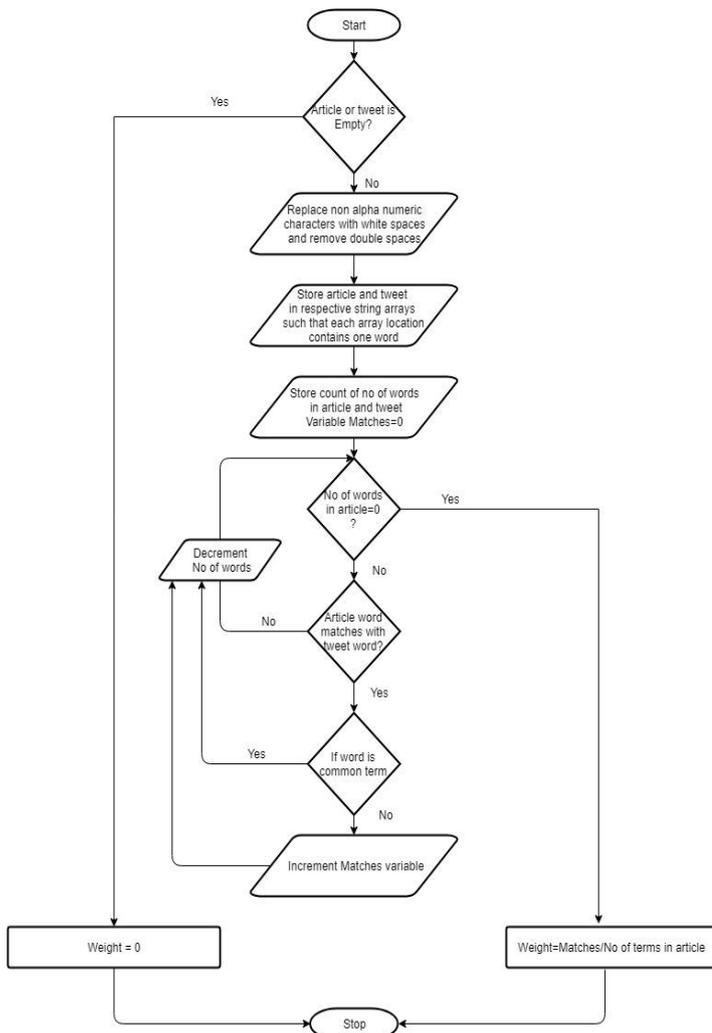


Figure 3: Algorithm to calculate Similarity Weight

Example for comparing Article title and tweet:

Assuming the following two strings:

String 1: India is world's largest democracy which follows parliamentary system.

String 2: Parliamentary system of democracy is found in different countries such as India.

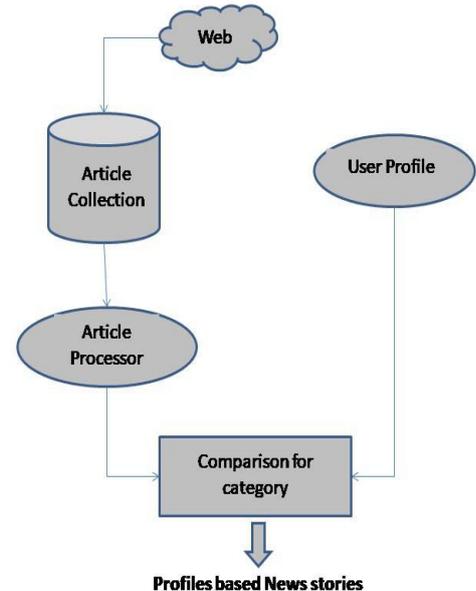


Figure 5: User Profile based News Recommender Architecture

3.3 Hybrid News Recommendation System:

In hybrid approach both the above recommenders are used. Firstly popularity based system is implemented and articles are sorted according to their popularity weight and then the sorted database is compared for the similarity with the user profile. Thus it gets both the advantages of popularity and User profile based recommender system.

4. CONCLUSIONS

The design and implementation of three different approaches to recommend interesting news articles based on the popularity and user preferences. We reach to the following conclusions:

1. Hybrid news recommender gives better results as compared to the other two system as it is an extension of adding popularity to user profile based recommender.
2. The user profile based recommender outperforms the traditional popularity based system as user popularity is not taken into consideration while calculating popularity weight.

REFERENCES

- [1] Nirmal Jonnalagedda, Susan Gauch "Personalized News Recommendation Using Twitter" 2013 IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT)
- [2] Tuzhilin and G. Adomavicius, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," in IEEE Transactions on Knowledge and Data Engineering, pages 734-749, New Jersey, USA, 2005.
- [3] M. Balabanovic and Y. Shoham, "Fab: Content-Based, Collaborative Recommendation," in Communications of the ACM, pages 66-72, New York, New York, USA, March 1997.
- [4] Michael J. Pazzani, Daniel Billsus "Content-based Recommendation Systems" FX Palo Alto Laboratory, Inc., 3400 Hillview Ave, Bldg. 4 Palo Alto, CA 94304
- [5] Owen Phelan, Kevin McCarthy, and Barry Smyth "Buzzer - Online Real-Time Topical News Article and Source Recommender" L. Coyle and J. Freyne (Eds.): AICS 2009, LNAI 6206, pp. 251-261, 2010.
- [6] D. Billsus and M. J. Pazzani, "A Personal News Agent that Talks, Learns and Explains," in The Third Annual Conference on Autonomous Agents, pages 268-275, May 1999, ACM.

- [7] W. IJntema, F. Goossen, F. Frasincar, Hogenboom, "Ontology- Based News Recommendation," in International Workshop on Business intelligence and the Web, New York, USA, 2010