International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

News Recommendation based on User Preferences and Location

Dhanashree More[1], Ajinkya Phand[2], Nagarjun Komarashetty[3], Sampada Choudhari[4], P.N. Vengurlekar[5]

[1,2,3,4]BE Student, SITS, Pune
[5]Prof., Dept. of Computer Engineering, SITS, Pune, India

Abstract - Recommender systems are becoming an essential part of smart services. When building a news recommender system, we should consider special features different from other recommender systems. Hot news topics are changing every moment; thus, it is important to recommend right news at the right time. Reading the news is the favourite hobby for many people anywhere in the world. With the popularity of the Internet and social media, users are constantly provided or even bombarded, with the latest news around the world. With numerous sources of news, it has become a real challenge for users to follow the news that they are interested. In this work we follow two steps. First, to take the user's personal interest and their location preference respectively. We are taking the user's personal interest implicitly and explicitly. Second, by taking into consideration the above preferences, our system will map the news articles with respect to the location and provide them to the user. Further we are also providing dictionary as well as summarized news articles.

Key Words **User Profile, User Location, News articles similarity function, News recommendation System**

1. INTRODUCTION

NOWADAYS, news reading is an indispensable daily activity of many people. With their recent popularity of smart mobiles and the rapid development of the mobile Web, more and more people tend to read news online via their mobiles or other handheld devices, e.g., tablets. However, due to the huge volume of news articles generated every day, readers cannot afford to go through all the news online. So, news recommendation systems, which aim to filter out irrelevant online information and recommend to users their preferred news, have been widely studied [1], [2], [3], [4].

In classic personalized news recommendation systems, a user's news preferences are usually learned using his/her news reading history or other online activity histories; therefore, the user's news preferences are static and dynamic in these systems. For the dynamic user preferences, can be obtained in efficient way using user's Facebook profile respectively. So personalized news recommendations will recommend the news to the people based on their preferences. However, in real-world contexts, users' news preferences usually evolve with the change of their locations.

So, the solution is system to provide news based on location preferences of the user, so that user will get updated local news also. For the user's preferred location, Geolocation can be used to fetch the location with the help of its Latitude and Longitude. For further more additional features providing summarized news articles using Open NLP API's as well as providing a search dictionary to search a meaning of the word which user want to search can be a better option.

The rest of this paper is organized as follow. Section II problem statement. Section III describes our literature review of news recommendation based on user preferences and location. Section VI describes the System architecture.

Section V describes about architecture explanation. Section VI tells about the modules implemented in the system. Section VII describes the results of the system. Finally, Section V concludes.

2. PROBLEM STATEMENT

To develop a system that will analyse user's personal interest as well as location preferences and recommend appropriate news to them.

3. LITERATURE REVIEW

Sr No.	Paper Title	Paper theme/Idea	Advantages/limitations
1	Location - aware personalized news recommendation with deep semantic analysis	LP-ESA, LPDSA. Deep semantic analysis Location based news recommendations	**Advantages:** News recommendation based on user interest and their location. **Limitations:** Data redundancy, limitation in physical system-based method.

2	Content based news recommendations using fuzzy logic	news recommender system using fuzzy logic Web crawler	**Advantages** News recommendation based for future prediction. **Limitations:** Limited size of data set
3	Recommending news based on hybrid user profile, popularity, trends and location	Personalize news recommendation using popularity, trends, location of user profile	**Advantages:** Provides accuracy to news recommended system **Limitations:** More complex.
4	Personalized News Recommendation using classified keywords to capture user Preference	Neural Network Preference Mining Keyword Classification	**Advantages:** Accurate user interest analysis through keywords. **Limitations:** Learning and Modeling Limitations in neural network model.

4. SYSTEM ARCHITECHTURE

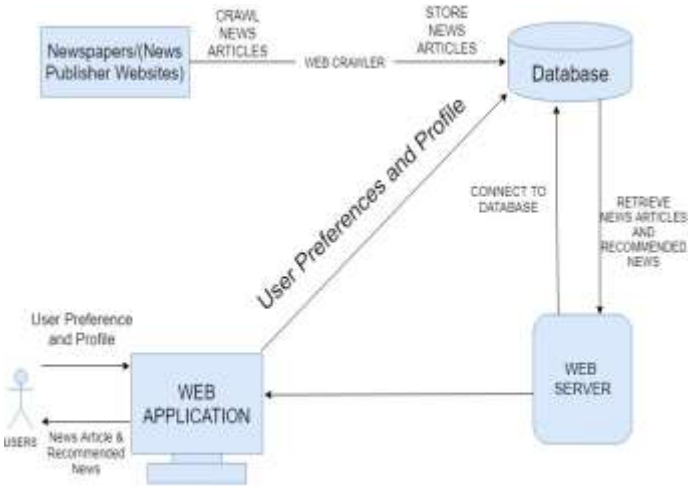

FIG : SYSTEM ARCHITECTURE

5. ARCHITECHTURE EXPLANATION

As per the architecture the main objective was to acquire real time news articles. There was no such real time dataset available online for the news articles. So, we developed a Java based web crawler using jsoup library to fetch real time news articles from various news portals. The collected news articles fetched from news portals were stored in MySQL Database for further processing. The next step was to create User profile. The user preference's for different news categories and also user's respective location was stored in the MySQL database. The proposed architecture follows the MVC (Model View Controller Architecture) standards where Model contains all the Database Logic, View tells us about the frontend which is developed in HTML, CSS and bootstrap, and the Servlet acts as a controller between model and view. After creating the User profile and News articles appropriate mapping was done and further the interested news is recommended to the user.

6. MODULES

There are four main modules in the propose system:

I. News Extraction

This module was the first step to start with the news recommendation system. In this module a java based jsoup library was used for web crawling purpose. jsoup is a Java library for working with real-world HTML. It provides a very convenient API for extracting and manipulating data, using the best of DOM, CSS, and jQuery-like methods.

jsoup implements the WHATWG HTML5 specification, and parses HTML to the same DOM as modern browsers do.

- scrape and parse HTML from a URL, file, or string
- find and extract data, using DOM traversal or CSS selectors
- manipulate the HTML elements, attributes, and text

A Web crawler, sometimes called a spider or spiderbot and often shortened to crawler, is an Internet bot that systematically browses the World Wide Web, typically for the purpose of Web indexing.

The basic functioning of our web crawler is as follows:

1. The home page link of NDTV news portal is given as an input to the crawler.
2. The crawler then visits the home page and extracts all the necessary links present on the page and stores it into the MySQL database along with the timestamp.
3. Next the crawler takes the stored links one at a time and visits the same link and extracts all the appropriate metadata like news article heading,

sub heading, publish date, image and article body and stores it into MySQL database.

4. If any metadata is missing then whole article is deleted from database. This is repeated for all the links.

II. User and location Preferences (Implicit and explicit)

In this module the main focus was to develop user profile. The user preferences were captured both implicitly and explicitly.

For explicit user preferences, user was asked to provide their interest via a form, provided to them after login. The news categories selected by the users as per their interest are then stored into MySQL database. For the location part, explicitly user is been provided with a list of locations to be selected.

For implicit preferences users are provided with a Facebook login option. After login with Facebook the proposed system will extract the interest, liking, events, posts, day to day activity of the users respectively. According to extracted data, users' profile was generated and eventually appropriate news articles was recommended. Implicit location was fetched with the HTML5 Geolocation (It is an API through which we can get the current coordinates of the user in terms of latitude and longitude) Latitude and longitude fetched from the geo location API was then mapped to city/area to provide the local news.

III. Summarized News articles

It is always better to read a summarized content then to read the whole news article. So, we developed a java-based summarizer which displays a precise summary of the news article to read. Automatic summarization is the process of reducing a document by a computer program to create summary which retains the important aspects of the original document. The proposed summarizer consists of 3 major steps, Pre-processing, Extraction of feature terms and ranking the sentence.

A) Pre-processing:

This step involves Sentence segmentation, Sentence tokenization, stop word Removal and Stemming.

Sentence Segmentation is the process of decomposing the given text document into its constituent sentences along with its word count, Tokenization is the process of splitting the sentences into words by identifying the spaces, comma and special symbols between the words, stop words are common words that carry less important meaning than keywords. This word should be eliminated otherwise sentence containing them can influence summary generated. Stemming is the process of reducing inflected (or sometimes derived) words to their word stem, base or root form—generally a written word form.

B) Feature Extraction

The sentences are ranked based on four important features: Frequency, Sentence Position value and Similarity with the Title.

Frequency is the number of times a word occurs in a document, Position of the sentence in the text, decides its importance. Sentences in the beginning defines the theme of the document whereas end sentences conclude or summarize the document, the similarity with the title consists of the words in titles and headers. These words are considered having some extra weights in sentence scoring for summarization. Final Score is given to the sentence after all the processing.

C) Sentence Ranking

After scoring of each sentence, sentences are ranked in descending order of their score value. After ranking the sentences based on their total score the summary is produced by selecting top ranked sentences.

IV. Dictionary

To provide a dictionary was an extra feature in the proposed system. In this module user can search any English word and find the meaning their itself.

For this we created a MySQL database with 1 Lakh Plus words.

6. RESULTS

(Signup form)

(Login form)

(User profile-selecting categories)

(User profile-selecting location)

(Summarized News displayed to the user)

(Word search using dictionary)

7. CONCLUSION AND FUTURE SCOPE

Recommender systems have made significant progress in recent years and many techniques have been proposed to improve the recommendation quality. However, in most cases, new techniques are designed to improve the accuracy of recommendations, whereas the recommendation diversity has often been overlooked. Proposed system will not only observe the news content on user preference or popularity basis but also provides news according to the user's preferred location. Proposed system will also provide summarized news articles and dictionary for user's

Comfort and betterment. After performing the extensive literature survey and recognizing the various methods used by the various authors, we have proposed our system architecture and the basic design workflow of our system. For instance, we have selected Jsoup library for scrapping and HTML geolocation for fetching user's location.

Proposed system will also help to refine popular and effective news content according to user desire.

Our future scope includes improving the performance of the system, generating analysis report of the news articles and keeping track of users who spend time on a particular category of news articles and learn accordingly.

ACKNOWLEDGEMENT

Authors want to acknowledge Principal, Head of department and guide of their project for all the support and help rendered. To express profound feeling of appreciation to their regarded guardians for giving the motivation required to the finishing of paper.

REFERENCES

[1] "Content Based News Recommendation System Based on Fuzzy Logic" (Md. Nuruddin Monsur Adnan, Mohammed Rashid Chowdury, Iftifar Taz, Tauqir Ahmed, Rashedur M Rahman).

[2] Recommending News Based on Hybrid User Profile, Popularity, Trends, and Location Suraj Natarajan Department of Computer Science San Jose State University San Jose, CA, USA surajnm@gmail.com, Melody Moh Department of Computer Science San Jose State University San Jose, CA, USA melody.moh@sjsu.edu2016 International Conference on Collaboration Technologies and Systems.

[3] Location-Aware Personalized News Recommendation with Deep Semantic Analysis Cheng Chen, Xiangwu Meng, Zhenghua Xu, and Thomas Lukasiewicz 2016

[4] J. Bao, M. Mokbel, and C.-Y. Chow, "GeoFeed: A locationaware news feed system,"in Proceedings of the IEEE
International Conference on Data Engineering, 2012,54–65.

[5] J. Bao and M. F. Mokbel, "GeoRank: An efficient locationaware news feed ranking system," in Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2013, pp. 184–193.

[6] Priyanka Jain, Savita Rathod, "A STUDY OF NEWS RECOMMENDATION SYSTEM: A REVIEW" International Journal of Research in Science & Engineeringe-ISSN:2394829 Volume: 3 Issue: 3 May-June 2017.