Research Article

Personalized News Recommendation System Based Preferences and Behavior Analysis

Mr.Ruturaj Shivaji Harer and Prof. Dr. S.B.Chaudhari

Department of Computer Engineering College Name: JSPM College Of Engineering Handewadi,Hadapsar ,Pune

Abstract

Now-a-days people can read news from several sources around the world. This paper investigates a novel user profile model to express users' preferences from different aspects. Then, considers the scope of the user's preferences for historical news, and propose a method to calculate the desire weight of historic news consistent with the user's analyzing behavior and the popularity of news. In proposed system,behaviour and preference (BAP) method may want to assemble user profiles greater correctly. Additionally, represents a dynamic technique for news recommendation, wherein each short-term and long-term user preferences are taken into consideration. The contribution work is to implement location-aware personalized news recommendation with explicit semantic analysis (LP-ESA), which recommends news using both the users' personal interests and their geographical contexts. The experimental consequences show that BAP technique and LP-ESA technique can fundamentally increase the recommendation outcome.

Keywords: News Recommendation, Personalization, User Profiling Method, User Behavior, Location-Aware News Recommendation

Introduction

Scientific discourse. As more and more journal articles and conference papers are published year by year, it becomes increasingly difficult to identify research articles that are related to one's field of interest. With the digitization of research publications, there has been a move to use computers to augment the search for related news which are relevant to a researcher's field of interest. Such systems are known as research news recommendation systems. A recommender system can be taken as a black box which takes in a profile of a user and matches it against a candidate set of items in order to suggest previously unseen items for a user. These items are considered to be the most relevant recommendations for that user. The existing news recommendation systems can be divided into three categories: content-based recommendation, recommendation based on collaborative filtering, and hybrid recommendation. Content-based recommendation: the recommendation system tries to find news with similar content to the news that the user has read. Content-based recommendation systems are usually easy to implement. However, in some scenarios, the profile of the user with a bag of words is not sufficient to accurately capture the user's preferences. Recommendation based on Collaborative filtering: the system recommends news by users' news ratings, and generally they are content free. Many users

do not have sufficient historical behaviors, or the number of users in the system is not high enough, which is known as a cold-start problem. Hybrid recommendations: as discussed above, content-based and collaborative filtering recommendation systems can provide meaningful results, but they have some drawbacks. This paper proposes a new news recommendation system that extends the user profile to three stages.

And we proposed a novel method called BAP to build the user profile. The method gives each historical news a corresponding weight based on user's reading Behavior And the Popularity of news, instead of 0, 1, or some fixed value. Furthermore, when dealing with short-term profiles, we propose a time function to adjust the user's preferences for all historical news rather than some of it. This helps us construct a more objective and comprehensive short-term profile of the user. The system consists of four main components: news collection and processing, user profiling method, personalized news recommendation and location-aware personalized news recommendation. There are some challenges of existing system which are solved in this paper. First, many news recommendation systems, the user profiles are one-sided, and user modeling from a single perspective cannot reflect the real preferences of users. Second, there is not yet a way to assess the degree of users' preferences for historical

news. In reality, users' preferences for news are quite different. Thus, treating these historical records equally to analyze a user's preferences is not reasonable. Third, when building a short-term profile, most research studies abandon the relatively early browsing records, or use only a few recent browsing records. This may cause many contingencies and an incorrect understanding of the user's preferences, or the recommendation results will be too similar to what the user just read.

A. Motivation • To calculate the user's preference for the news based on user behavior and news popularity using BAP method • To provide a high quality of dynamic recommendation results with the short-term and long-term preferences of users. • Users' news preferences strongly depend not only on their geographical contexts (i.e., locations), but also on their personal interests; so, both should be gives satisfactory personalized news recommendations.

Review of Literature

The paper [1] proposes a half and half strategy called area mindful customized news suggestion with unequivocal semantic investigation (LP-ESA), which prescribes news utilizing both the clients' close to home premiums and their land settings. LP-DSA utilizes proposal arranged profound neural systems to remove thick, unique, low dimensional and compelling component portrayals for areas, clients, and news. Points of interest are: LP-DSA beats the enormous dimensionality, sparsity, and excess issues in LP-ESA by utilizing profound neural systems. Increment the news proposal execution regarding both viability and proficiency. Lessens the calculation cost. A hybrid recommender based on both, user content and collaborative filtering Wesomender framework. Wesomender [7] is comprised of two main components, a collaborativefiltering component and a content-based component. Each component evaluates the news the user has not seen or rated yet, and produces independent recommendations. Advantages are: To generate context-aware recommendations in the journalism field. A context-aware adaptive recommendation engine can fulfill the needs of journalists' daily work when retrieving timely. The paper [2] presents a rough set based collaborative filtering approach to predict a missing news category rating values of a user, and a new novelty detection approach to improve ranking of novel news items. An end-to-end system prototype that can take a collection of news articles and the user interest as input and then rerank the news articles based on novelty and CPCC similarity between the user profile database and common news articles database to provide a personalized news recommendation to the users. Advantages are: Automatically detect the novelty of news items. Efficient approach to automatically detect the missing rating value of an active user. The paper [4]

examines web server get to logs of a huge online news distributer to distinguish readership designs on the web. Specifically, the investigation is finished by first building up a model, which can be utilized to foresee in all probability articles to be perused by a specific client, trailed by breaking down what are the most significant highlights and deciphering the educated model. Advantages are: Time window selection can have a significant impact on the accuracy of predictions. Disadvantages are: The inclusion of Users feature, which is computationally the most expensive. In [6] paper, proposes PENETRATE, a novel Personalized NEws recommendation framework using ensemble hierarchical clustering to provide attractive recommendation results. Our proposed framework is beyond content-based methods and collaborative filtering, in which individual user behavior and user group behavior are simultaneously considered for recommendation. Advantages are: Improved accuracy and efficiency. Disadvantages are: Time consumption is high. In [10] paper presented two approaches that take into account the meaning of words. The methods are based on concepts and their semantic similarities, from which derive the similarities between news items. First method, Synset Frequency - Inverse Document Frequency (SF-IDF), second method, Semantic Similarity (SS). The proposed approaches to news item recommendation have been implemented as Ceryx, an extension to the Hermes News Portal news personalization service. Advantages are: Performs statistically better than TF-IDF.

Proposed Methodology

This paper proposes a new news recommendation system that extends the user profile to three stages. And we proposed a novel method called BAP to build the user profile. The method gives each historical news a corresponding weight based on user's reading Behavior And the Popularity of news, instead of 0, 1, or some fixed value. Furthermore, when dealing with short-term profiles, we propose a time function to adjust the user's preferences for all historical news rather than some of it. This helps us construct a more objective and comprehensive short-term profile of the user. The system consists of four main components: news collection and processing, user profiling method, personalized news recommendation and location-aware personalized news recommendation. 1) News collection and processing: There are three phase profile model for news. First phase, crawl the news from the Twitter API and extract keywords from news using vector space model. Second phase, extract topics by using NLP model with the help of LDA algorithm. Third phase, to identify the entities. 2) User Profiling: The profile of user is composed of four stages with help of user's browsing or reading history. The first stage represents some of the news keywords in which the user is interested. The second stage represents the topic distribution of the user's preferences. The third

stage represents the named entities in which the user is interested. And fourth stage represents the users' news preferences usually includes with the change of their locations. In this paper, the user's reading behavior is divided into several types by analysis. User's reading behaviors, popularity and location wise weight of the news calculated. 3) Dynamic Personalized News Recommendation using BAP: Recommend news by using the long-term and short-term preferences of users. A time-sensitive function is proposed to construct the short-term profile by adjusting the long-term profile of the user. After calculate the similarities between every news and profile of the user. And finally, the rank results of the long-term and short-term preferences, the news recommendation of the long-term and short-term profiles are retrieved. 4) Location-Aware Personalized News Recommendation With Explicit Semantic Analysis: A hybrid news recommendation method, called LP-ESA, which takes into account both the user's current location and also his/her personal interests. LPESA collects a set of geo-tagged documents for a given location, and estimates a local topic distribution using given location's geographical topics. LP-ESA models a localized user profile for each user at location according to both general user profile and local topic distribution. Finally the news recommendation provides to the user. A. Advantages of proposed system • This proposed approach support image data privacy protection • Using Shape-Based Invariant Texture Index (SITI) descriptor, proposed work will get more accurate result. • In proposed approach, data stored in encrypted format so more security will be provided. • Security will be maintained using access control. • Also security will be maintained using users image secret key
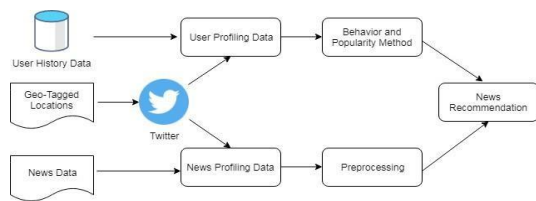
B. Architecture

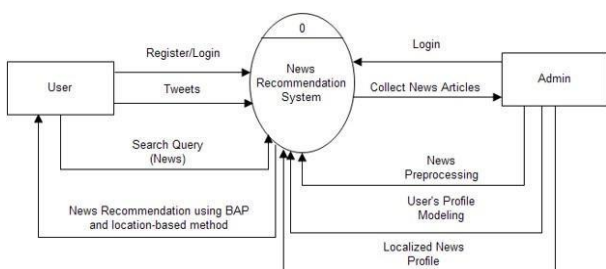**Fig. 1.** Proposed System Architecture

C. Flow Diagram

Fig. 2. Proposed Flow Diagram

D. Algorithm 1. LDA Algorithm First and foremost, LDA provides a generative model that describes how the documents in a dataset were created. In this context, a dataset is a collection of D documents. Document is a collection of words. So our generative model describes how each document obtains its words. Initially, let's assume we know K topic distributions for our dataset, meaning K multinomials containing V elements each, where V is the number of terms in our corpus. Let i represent the multinomial for the ith topic, where the size of i is V: $—i—=V$. Given these distributions, the LDA generative process is as follows: Steps: 1. For each document: (a) Randomly choose a distribution over topics (a multinomial of length K) (b) for each word in the document: (i) Probabilistically draw one of the K topics from the distribution over topics obtained in (a), say topic j (ii) Probabilistically draw one of the V words from j E. Dataset Twitter Twitter dataset is used for the classification purpose.In this social networking service users can freely communicate. They post and communicate with messages known as "tweets". Originally there was restriction of tweets characters that is 140, but from November 7, 2017, this limit was increased to 280 for all languages except Chinese, Japanese, and Korean. Registered users can post, like, and retweet tweets, but unregistered users can only read the messages. Users access Twitterthrough its website interface, through Short Message Service (SMS) or its mobile-device application software ("app"). Twitter, Inc. is based in San Francisco, California, and has more than 25 offices around the world. F. Mathematical Model Construct the profile of the user by three different but related stages: news keywords, news named entities, and news topic distributions. Each profile of the user can be expressed as

$$V u =< F u, tt u, E u$$

which corresponds to the three-stage model of news. However, there are some differences between this and the profile of news, where represents the keywords collected from the historical news that the user accessed in the past, and each entry consists of a representative word, the corresponding weight, and the last time the user accessed it;

$$F\_u = f\_1, w\_1, t\_1, f\_2, w\_2, t\_2, ……, F\_u$$

$$E\_u = e\_1, w\_1, t\_1, e\_2, w\_2, t\_2, ….. , E\_u$$

represents the named entities collected from the historical news that the user accessed in the past, and each entry consists of a representative named entity, the corresponding weight, and the last time the user accessed it; and represents the topic distribution collected from the historical news that the user accessed in the past, and each entry consists of a representative topic id, the corresponding weight, and the last time the user accessed it.

$tt\_u = g\_1, w\_1, t\_1, g\_2, w\_2, t\_2, ….. , tt\_u$

Results and Discussion

Experiments are done by a personal computer with a config- uration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and jdk 1.8. The application is web application used tool for design code in Eclipse and execute on Tomcat server. The real time news posts collection for dataset of this application using Twitter API with the help of Twitter4j-core and Twitter4j-stream jars. Some functions used in the algorithm are provided by list of jars like standfordcore-nlp jar for POS tagging etc. Proposed work is expected to implement news posts rec- ommendation system which collects input dataset of list of news posts from Twitter API. Keywords extracts using LDA algorithm. Topics extract using POS tagging method. And, finally entities extract using Named Entity Recognition (NER) method. Apply all the keywords extraction, topics extraction and entities extraction methods for news posts recommendation provide to user side. Expected outcome of this project is providing news recommendations to users with the help of Behavior-and-Popularity (BAP) method. Also providing news recommendations to the users with the help of collaborative filtering method using user's personal interests as well as their GPS tracking (geographical contexts) locations. Fig. 2 represents number of clusters generated for each aspect. The Fig. 3 shows performance on combination of features for combination of BAP+GPS+Interest proposed system effectively gives recommendations to the users as compared to existing methods.
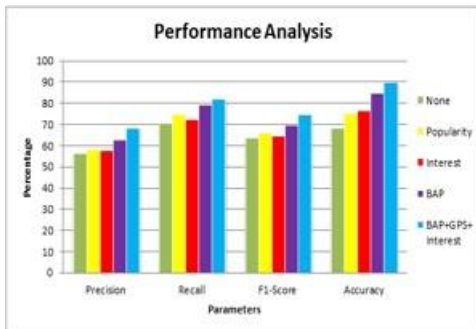

Fig. 3. Accuracy Graph

	Existing System	Proposed System
Accuracy	79.32	89.77

Conclusion

In this paper, the BAP based user profiling method is proposed to calculate the user's preference for the news based on user behaviour and news popularity. The profiles of users can be constructed more accurately with this method. This system can provide a high quality of dynamic recommendation results with the short-term and long-term preferences of users. Experimentally, performance analysis of both BAP method and LP-ESA news recommendation system performs well. For future work, to increase heterogeneity of the recommendation outcomes and elaborate more preferences of users and also improve the quality of user experience.

References

[1] C. Chen, X. Meng, Z. Xu, and T. Lukasiewicz, "Location-aware personalized news recommendation with deep semantic analysis," IEEE Access, vol. 5, pp. 16241638, 2017.

[2] K. G. Saranya and G. S. Sadasivam, "Personalized news article recommendation with novelty using collaborative filtering based rough set theory," Mobile Netw. Appl., vol. 22, no. 4, pp. 719729, 2017.

[3] P. Lv, X. Meng, and Y. Zhang, "FeRe: Exploiting influence of multidimensional features resided in news domain for recommendation," Inf. Process. Manage., vol. 53, no. 5, pp. 12151241, 2017.

[4] B. Fortuna, P. Moore, and M. Grobelnik, "Interpreting news recommendation models," in Proc. 24th Int. Conf. WorldWideWeb (WWWCompanion), 2015, pp. 891892.

[5] L. Li, L. Zheng, F. Yang, and T. Li, "Modeling and broadening temporal user interest in personalized news recommendation," Expert Syst. Appl., vol. 41, no. 7, pp. 31683177, 2014.

[6] L. Zheng, L. Li, W. Hong, and T. Li, "PENETRATE: Personalized news recommendation using ensemble hierarchical clustering," Expert Syst. Appl., vol. 40, no. 6, pp. 21272136, 2013.

[7] Montes-Garc´ıa, J. M. A´ lvarez-Rodr´ıguez, J. E. Labra-Gayo, and M. Mart´ınez-Merino, "Towards a journalist-based news recommendation system: The Wesomender approach," Expert Syst. Appl., vol. 40, no. 17, pp. 67356741, 2013.

[8] M. Lu, Z. Qin, Y. Cao, Z. Liu, and M.Wang, "Scalable news recommendation using multi-dimensional similarity and JaccardKmeans clustering," J. Syst. Softw., vol. 95, pp. 242251, Sep. 2014.

[9] H. Wen, L. Fang, and L. Guan, "A hybrid approach for personalized recommendation of news on the Web," Expert Syst. Appl., vol. 39, no. 5, pp. 58065814, 2012. [10] M. Capelle, F. Frasincar, M. Moerland, and F. Hogenboom, "Semanticsbased news recommendation," in Proc. 2nd Int. Conf. Web Intell., Mining Semantics, 2012, pp. 2736.

[11] L. Li, D.Wang, T. Li, D. Knox, and B. Padmanabhan, "SCENE: A scalable two-stage personalized news recommendation system," in Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. (SIGIR), 2011, pp. 125134.

[12] Y. Xie, L. Chen, K. Jia, L. Ji, and J. Wu, "iNewsBox: Modeling and exploiting implicit feedback for building personalized news radio," in Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage., Nov. 2013, pp. 24852488.

[13]J. Bao and M. F. Mokbel, "GeoRank: An efficient location-aware news feed ranking system," in Proc. 21st ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst., Nov. 2013, pp. 184193.

[14]J.-W. Son, A.-Y. Kim, and S.-B. Park, "A location-based news article recommendation with explicit localized semantic analysis," in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Aug. 2013, pp. 293302.

[15] Y. Zhou and J. Luo, "Geo-location inference on news articles via multimodal pLSA," in Proc. 20th ACM Int. Conf. Multimedia, Nov. 2012, pp. 741744