Efficient and Effective Location Recommendation through Content Analysis

Mr. S. Hari Kishore[1]
Ms. S. D. Gayathridevi[2] Mr. S. Bharath[3]
UG Scholars
Department of Computer Science
Muthayammal Engineering College(Autonomous),
Rasipuram, Namakkal.

Mr. T. Aravind[4]
Assistant Professor
Department of Computer Science
Muthayammal Engineering College(Autonomous),
Rasipuram, Namakkal.

Abstract -Location recommendation plays a vital role in helping people in finding beautiful places. The recent research has studied how to recommend locations with social and geographical information, but few of them addressed the cold-start problem. A typical method is to feed them into explicit-feedback-based content-aware collaborative filtering, but they require drawing negative samples for better learning performance, as users' negative preference is not observable in human mobility. Prior studies have empirically shown sampling-based methods do not perform well. Therefore, a novel approach has been implemented which recommend location based on machine learning process. The user reviews are taken into consideration as dataset. Dataset are preprocessed and meaning of user reviews is understood by system automatically through NLP. Then new recommendation has been suggested through this process and data base is loaded. Hence our system achieves more accurate recommendation compared to other existing approach. Finally, we evaluate LR-NLP with a user review dataset in which users have profiles and textual content. The results show that our proposed outperforms several competing baselines, and that user feedback is not only effective for improving recommendations but also overcomes cold-start problems.

Index Terms- Location recommendation; NLP; SVM

I. INTRODUCTION

A. Importance of Recommendation

Recommender Systems have emerged as powerful tools for helping users find and evaluate items of interest. These systems use a range of techniques to assist users identify the things that best fit their tastes or needs. While popular CF-based algorithms still to produce meaningful, personalized results in a variety of domains, data mining techniques are increasingly getting used in both hybrid systems, to enhance recommendations in previously successful applications, and in stand-alone recommenders, to supply accurate recommendations in previously challenging domains. The use of data mining algorithms has also changed the types of recommendations as applications move from recommending what to consume to also recommending when to consume. While recommender systems may have started as largely a passing novelty, theyclearly appear to have moved into a real and powerful tool in a variety of applications, and that data mining algorithms can be and will continue to be an important part of the recommendation process.

B. Location Recommendation

Prior research has mainly investigated a way to leverage spatial patterns, temporal effects spatio-temporal influence, social influence, text-based analysis, and implicit characteristics of human mobility to recommend locations. However, a number of these methods require each user to have sufficient training data while others assume locations have accumulated ample textual information (e.g., tips), making it challenging to use them to tackle the cold-start problem, specifically, and recommending locations for new users. Fortunately, users are often linked to social networks, like Twitter and Weibo, which probably collect rich semantic content from users. This semantic content is likely to imply user interest, an important element for capturing users' visiting behavior.

Therefore, they will exploited to address the cold-start challenge and even improve location recommendation. A typical method is to feed them into traditional explicit-feedback content-aware recommendation frameworks, like LibFM, SVDFeature, regression-based latent factor model or MatchBox. These frameworks require drawing negative samples from unvisited locations for better learning performance, since a user's negative preference for locations is not observable in human mobility data. However, it has been empirically shown that sampling-based frameworks don't perform yet as an algorithm that treats all unvisited locations as negative yet assigns them a lower preference confidence, since the latter one deals with the sparsity issues better.

C. Challenges and Issue in Recommendation System
Cold Start Problem

The term derives from cars. When it's really cold, the engine has problems with starting up, but once it reaches its optimal operating temperature, it'll run smoothly. With recommendation engines, the "cold start" simply means the circumstances aren't yet optimal for the engine to provide the most effective possible results. In ecommerce, there are two distinct categories of cold start: product cold start and user cold starts. News sites, auction sites, ecommerce stores and classified sites all experience the merchandise cold start. The user or visitor cold start simply means that a recommendation

engine meets a brand new visitor for the primary time. Because there is no user history about the user, the system doesn't know the personal preferences of the user.

Data Sparsity

In practice, many commercial recommender systems are based on large datasets. As a result, the user-item matrix used for collaborative filtering may be extremely large and sparse, which brings about the challenges in the performances of the recommendation. One typical problem caused by the data sparsity is that the cold start problem. As collaborative filtering methods recommend items based on users' past preferences, new users will have to rate sufficient number of items to enable the system to capture their preferences accurately and thus provides reliable recommendations.

Scalability

As the numbers of users and items grow, traditional CF algorithms will suffer serious scalability problems. For example, with tens of millions of customers and millions of items, a CF algorithm with the complexity of n is already overlarge. As well, many systems have to react immediately to online requirements and make recommendations for all users regardless of their purchases and ratings history, which demands a better scalability of a CF system. Large web companies like Twitter use clusters of machines to scale recommendations for their uncountable users, with most computations happening in very large memory machines.

Gray Sheep

Gray sheep refers to the users whose opinions do not consistently agree or disagree with any group of individuals and thus do not benefit from collaborative filtering. Black sheep are the alternative group whose idiosyncratic tastes make recommendations nearly impossible. Although this is a failure of the recommender system, non-electronic recommenders even have great problems in these cases, so black sheep is a suitable failure.

In addition, supported this evaluation, we discover that user profiles and semantic content can make significant improvements over the counterpart without taking them into account. In addition to the warm-start evaluation, we also perform a cold-start evaluation with a user-based 5-fold cross validation by splitting users into five non-overlapping groups. The results indicate that both user profiles and semantic content are useful for tackling the cold-start problem in location recommendation based on human mobility data, and that user profiles are more practical than semantic content.

By understanding the need of recommendation system in current generation and its issue were also discussed. Hence a method to overcome this issue has been proposed through machine learning method called NLP. Through this NLP could recommend best place to new user by overcoming issues in recommendation process. Efficient classification SVM is employed for better result classification.

II. RELATED WORK

We propose a Machine Learning based framework for content-aware collaborative filtering from user feedback commands. Therefore, related work consists of location recommendation and content-aware collaborative filtering.

Location recommendation has been a vital topic in location-based services. From the perspective of forms of recommended items, some prior research focuses on recommending specific types of locations while others are generalized for any type of locations. For example, Defu Lian, Yong Ge, Fuzheng Zhang [1] have developed Content-aware collaborative filtering for location recommendation based on human mobility data, Vincent W. Zheng, Bin Cao, Yu Zheng, Xing Xie, Qiang Yang [2] have developed Collaborative filtering meets mobile recommendation: A user-centered approach, Mao Ye, Peifeng Yin, Wang-Chien Lee [3] have developed Exploiting geographical influence for collaborative point-of-interest recommendation, Wen-Yuan Zhu, Wen-Chih Peng and Ling-Jyh Chen [4] have developed Modeling user mobility for location promotion in location-based social networks, Bin Liu, Yanjie Fu, Zijun Yao, Hui Xiong [5] have developed Learning Geographical Preferences for Point-of-Interest Recommendation, Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu [6] have developed Exploring temporal effects for location recommendation on location-based social networks, Quan Yuan, Gao Cong, Zongyang Ma [7] have developed Time aware point-of-interest recommendation, Quan Yuan, Gao Cong and Aixin Sun [8] have developed Graph-based point-of-interest recommendation with geographical and temporal influences, Anastasios Noulas, Salvatore Scellato, Neal Lathia, Cecilia Mascolo [9] have developed A random walk around the city: New venue recommendation in location-based social networks, Pasquale Lops [10] has developed Content-based recommender systems: State of the art and trends.

In contrast to those methods, we mainly study the results of user feedback comments instead of location information on recommendation. User information should be more important than location information when addressing the cold start problem since it is available earlier for inferring user interest. Additionally, we propose a general framework for location recommendation based Natural Language Processing (NLP) and Keyword Search Algorithm.
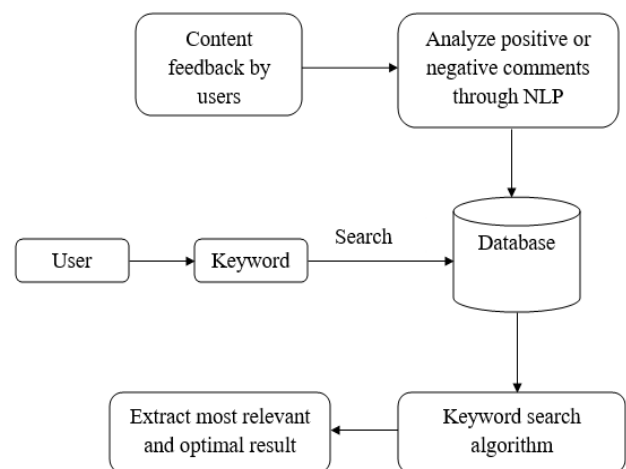

Fig. 1. Proposed system working architecture

III. METHODOLOGY

A. Content Feedback From User

The feedbacks collected from users through social networks by posting images and giving comments with respect to location information. Therefore, data are gathered from various users are considered and loaded in our database through our application will used for analysis. The embracement of the web into our daily life activities during this contemporary period has become almost inevitable and quite numbers of populace depend upon the web for different purpose range from placing on their view and read others view while also commenting on such views, e-learning, e-banking, e-library and e-commerce etc. The number of available documents on the web is enough to improve the diverse ways of educating and research need of the public now then. In this system, data is collected on the basis of our visited web pages, our activities in social networks, smart phones and through the numerous sensors of the physical world. It is this stream that forms the basis of Big Data. A data stream (or flow of various streams) basically, without interpretation, has less value, but based on analysis creates information that we can use, so the data becomes valuable. In our work it'll focus on making use of context-based approach additionally to CF approach to recommend quality content to its users. It might be exploiting available contextual information, analyzing and summarizing user queries, and linking the metadata like tags and feedback to a richer information model to recommend content.

B. Support Vector Machine

Support Vector Machine (SVM) is a noteworthy methodology for characterizing high-dimensional information with the utilization of Structural Risk Minimization (SRM) rule. SVM has been passed on as a discriminative classifier which is further precise than most previous order models. SVM gains the ideal hyper plane those parts preparing information focuses from various classes by expanding the arrangement edge. Also, SVM is utilized to information focuses with nonlinear choice surfaces by connecting with a framework recognized as the part technique that plans the information to a higher dimensional component space, where a direct isolating hyper plane can be dispatch.

C. Natural Language Processing Works

NLP entails applying algorithms to identify and extract the natural language rules such that the unstructured language data is converted into a form that computers can understand. When the text has been provided, the computer will utilize algorithms to extract meaning associated with every sentence and collect the essential data from them.

Syntax

Syntax refers to the arrangement of words in a sentence such that they make grammatical sense. In NLP, syntactic analysis is used to assess how the natural language aligns with the grammatical rules. Computer algorithms are used to apply grammatical rules to a group of words and derive meaning from them.

Lemmatization:It entails reducing the various inflected forms of a word into a single form for easy analysis.

Morphological segmentation:It involves dividing words into individual units called morphemes.

Word segmentation:It involves dividing a large piece of continuous text into distinct units.

Part-of-speech tagging:It involves identifying the part of speech for every word.

Parsing:It involves undertaking grammatical analysis for the provided sentence.

Sentence breaking:It involves placing sentence boundaries on a large piece of text.

Stemming:It involves cutting the inflected words to their root form.

Semantics refers to the meaning that is conveyed by a text. Semantic analysis is one of the difficult aspects of Natural Language Processing that has not been fully resolved yet. It involves applying computer algorithms to understand the meaning and interpretation of words and how sentences are structured.

Named entity recognition (NER): It involves determining the parts of a text that can be identified and categorized into preset groups. Examples of such groups include names of people and names of places.

Word sense disambiguation: It involves giving meaning to a word based on the context.

Natural language generation: It involves using databases to derive semantic intentions and convert them into human language.

D. Working Process

1-Preprocess Reviews: read reviews, use a morphology and part-of-speech tagging systems to:
   a. Find part-of-speech and root for each word in the text
   b. Identify adjectives in the text
   c. Check if neglected tool (word) is attached to the adjectives

2-Apply Rules: Extract attributes and associate them with their values (adjectives) that are labeled in step #1.

a.Tag up to two words headed by an adjective, stop when encountering a verb, particle or punctuation mark.

b.Use the following rules to form adjective phrases:

Adjective Phrase → <Attribute><Adjective>
| <Attribute><Neglect-Tool><Adjective>
Attribute → Simple Attribute | Compound Attribute

c.Check if <adjective> is already in adjectives table, find its classification, either positive or negative, otherwise classify it and update the adjectives table

d. Check if <attribute> either if it is a simple or compound is in attributes table, if not validate it and update attributes table

3-Update Graph: use the output from step #2 (attributes/values) to update graph by updating frequency of each node and each edge. Each node in the graph contains either an attribute or a value, attribute nodes connected to values nodes through edges.

E. SVM Classifier

The Support Vector Machine (SVM) being proposed as a classifier to solve the problems for recognizing pattern

between two groups. Support Vector Machine (SVM) aims to identify the best margin separation of the hyperactive plane between two groups of data. It was originally intended for solving separable cases, but can be extended to solve the linearly non-separable case by mapping the original data vector to spaces of higher dimensions. So also, SVM material to information focuses with nonlinear choice surfaces by connecting with a framework recognized as the part technique that plans the information to a higher dimensional component space, where a direct isolating hyper plane can be dispatch. Hence SVM is used to classify accurately that particular comment is positive, negative or neutral comment. Based on trained dataset new comment words were analyzed and for accurate classification SVM is used.

F. Identifying Positive Feedback

In this module, the loaded comment database is utilized for analysis of positive comments with respect to location. Data is extracted first from online sources like weblogs or websites. Data consist of comments and feedback posted by users on the website. The data which is fetched is unstructured and not in a useful form. This data is properly extracted from web pages and stored in database. These comments are stored according to location. The comments are stored in string format and are processed and broken into tokens to analyze each word in the string. These tokens are then sorted to remove repetitive words and prepositions which are not useful for determining polarity of comments. The positive and negative responses are grouped into separate database which were used for deciding the comments to be positive or negative. The useful tokens are stored in a list to determine their polarity i.e. positive or negative comments. The tokens are then used to determine for which feature the comment is made and the accordingly the ratings are assigned to that feature. The comment is related to a particular feature or not is determined by the words used in the comments provided. Based on user feedbacks either it is good place to visit or not is determined through average calculation. If maximum average calculation attains negative value then it'll be not considered to be recommended.
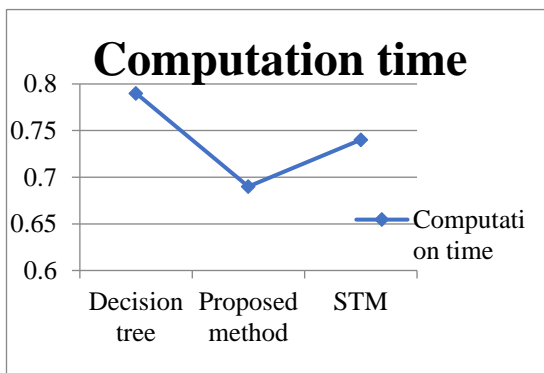

Fig.2. Prediction accuracy

G. User Query And Result

In this module, user enters their keyword and search in our analyzed and processed database. The keyword search plays a major role in searching particular content from a huge database. Most relevant content should be extracted from a

huge database and the recommendation of particular content is done through NLP. In general reviews are available in different format such as rating, likes and comments among these three comment-based reviews is most valuable therefore processing it automatically through machine is not easy task. In our proposed we are going to process it through Machine learning. Here recommendation has been done through recommending best places suggested by users after visiting those places. Therefore, most related content has been retrieved to respective user.

IV. RESULT AND DISCUSSION

The accurate recommendations of locations were done through NLP and efficient classification process (SVM). The secure and accurate prediction of highly recommended location through user feedback is shown below.

The computation time of our proposed work and existing method is shown below as graph. It shows our proposed method consumes minimum computation time and it increases performance of our system.
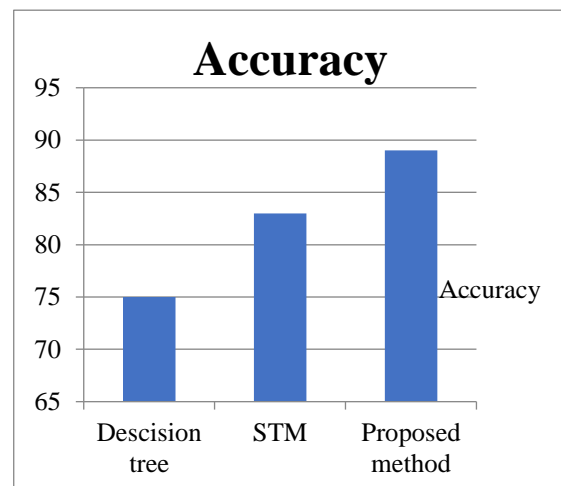

Fig.3. Computation Time Graph

V. CONCLUSION

A Machine Learning based framework for content-aware collaborative filtering from user feedback commands is proposed. Our experiment results indicate that NLP is superior to five competing baselines, including two state-of-the-art location recommendation algorithms and ranking-based factorization machine. Initially user feedbacks were collected from different websites are preprocessed and stored in database for processing. In preprocessing the unwanted content, irrelevant data are removed from database. Once preprocessed content is utilized by NLP for separating the content as string and analyzing it through both syntax and semantic analyzes. The meaning of words was analyzed and classified. By studying the effects of user profiles and semantic content, we find that they improve recommendation in warm-start cases and help address the cold-start problems. Once commands nature has been identified as either positive or negative through NLP next it was subjected to classification. SVM will efficiently classifies its category accurately and used for better prediction compared to existing system. Finally, user will retrieve their data based on

Special Issue - 2020

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCICCT - 2020 Conference Proceedings

recommended content through keyword-based search process which retrieve data with more relevancy.

VI. REFERENCES

[1] D. Lian, Y. Ge, F. Zhang, N. J. Yuan, X. Xie, T. Zhou, and Y. Rui, "Content-aware collaborative filtering for location recommendation based on human mobility data," in Proceedings of ICDM'15, pp. 261–270, IEEE, 2015.

[2] V. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang, "Collaborative filtering meets mobile recommendation: A user-centered approach," in Proceedings of AAAI'10. AAAl Press, 2010.

[3] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborativepoint-of-interest recommendation," in Proceedings of SIGIR'11, pp. 325–334, ACM, 2011.

[4] W.-Y. Zhu, W.-C. Peng, L.-J. Chen, K. Zheng, and X. Zhou, "Modeling user mobility for location promotion in location-based social networks," in Proceedings of KDD'15, pp. 1573–1582,ACM, 2015.

[5] B. Liu, Y. Fu, Z. Yao, and H. Xiong, "Learning geographical preferences for point-of-interest recommendation," in Proceedings of KDD'13, pp. 1043–1051, ACM, 2013.

[6] H. Gao, J. Tang, X. Hu, and H. Liu, "Exploring temporal effects for location recommendation on location-based social networks," in Proceedings of RecSys'13, pp. 93–100, ACM, 2013.

[7] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Timeaware point-of-interest recommendation," in Proceedings of SIGIR' 13, pp. 363, ACM, 2013.

[8] Q. Yuan, G. Cong, and A. Sun, "Graph-based point-of-interest recommendation with geographical and temporal influences," in Proceedings of CIKM'14. pp. 659–668, ACM, 2014.

[9] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, "A random walk around the city: New venue recommendation in location-based social networks," in Proceedings of SocialCom'12. pp. 144–153, IEEE, 2012.

[10] P. Lops, M. De Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in Recommender systems handbook. Springer, pp. 73–105, 2011.

[11] D. Yang, D. Zhang, Z. Yu, and Z. Wang, "A sentiment-enhanced personalized location recommendation system," in Proceedings of HT'13. pp. 119–128, ACM, 2013.

[12] B.Liu and H. Xiong, "Point-of-interest recommendation in location based social networks with topic and location awareness," in Proceedings of SDM'13, pp. 396–404, SIAM, 2013.

[13] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, "Geomf: joint geographical modeling and matrix factorization for point-ofinterest recommendation," in Proceedings of KDD'14. pp. 831–840, ACM, 2014.

[14] C. Cheng, H. Yang, I. King, and M. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks," in Proceedings of AAAI'12, 2012.

[15] Y. Liu, W. Wei, A. Sun, and C. Miao, "Exploiting geographical neighborhood characteristics for location recommendation," in Proceedings of CIKM'14. pp.739–748, ACM, 2014.

[16] C. R. Cloninger, T. R. Przybeck, and D. M. Svrakic, The Temperament and Character Inventory (TCI): A guide to its development and use. center for psychobiology of personality, Washington University St. Louis, MO, 1994.

[17] S. Rendle, "Factorization machines with libfm," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 3, no. 3, p. 57, 2012.

[18] T. Chen, W. Zhang, Q. Lu, K. Chen, Z. Zheng, and Y. Yu, "Svdfeature: a toolkit for feature-based collaborative filtering," Journal of Machine Learning Research, vol. 13, no. 1, pp. 3619–3622, 2012.

[19] D. Agarwal and B.-C. Chen, "Regression-based latent factor models," in Proceedings of KDD'09. pp. 19–28, ACM, 2009.

[20] D. H. Stern, R. Herbrich, and T. Graepel, "Matchbox: large scale online bayesian recommendations," in Proceedings of WWW'09. pp. 111–120, ACM, 2009.