International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

SCALABLE CONTENT AWARE COLLABORATIVE FILTERING FOR LOCATION RECOMMENDATION

N. SANGEETHA[1], J. SUJATHA[2]

[1]*PG Scholar, Department of MCA, Arulmigu Meenakshi Amman College of Engineering, Anna University, Vadamavandal (near Kanchipuram), India.*
[2]*Assistant Professor, Department of MCA, Arulmigu Meenakshi Amman College of Engineering, Anna University, Vadamavandal (near Kanchipuram), India.*

---***---

ABSTRACT:- Area suggestion assumes a fundamental job in helping individuals find appealing spots. Though the researches has studied how to suggest areas with social and land data information's about the new users. Because of mobility records are often shared on social networks semantic information can be leveraged to tackle this challenge. 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 preferences is not observable in human portability. To this end, we propose scalable Implicit-based Content-aware Collaborative Filtering (ICCF) system to considered to avoid negative examining. We evaluate the effective features are size and highlight estimate, and quadractically with the measurement of latent space. We additionally build up its association with chart Laplacian regularized framework factorization. At long last, we evaluate ICCF with a large-scale LBSN (location based service network) in which users have profiles and textual content. We develop framework named as 1-injection to address the sparsity problem of recommender systems. As our planned approach is methodology agnostic it is simply applied to a spread of CF algorithms.

1. INTRODUCTION

The web program has long become the foremost vital portal for standard folks searching for helpful info on the net. However, users would possibly expertise failure once search engines come back extraneous results that don't meet their real intentions. Such un connectedness is basically thanks to the big sort of users' contexts and backgrounds, still because the ambiguity of texts. Location Based Rating Prediction (LBRP) is a general category of search techniques aiming at providing better search results, which are tailored for individual user needs. As the expense, user information has to be collected and analyzed to figure out the user intention behind the issued query. The solutions to LBRP can generally be categorized into two types, namely click-log-based methods and profile-based ones. The click-log based mostly strategies ar easy they merely impose bias to clicked pages within the user's question history. Although this strategy has been demonstrated to perform consistently and considerably well, it can only work on repeated queries from the same user, which is a strong limitation confining its applicability. In distinction, profile-based methods improve the search experience with complicated user-interest models generated from user profiling techniques. Profile-based strategies are often doubtless effective for pretty much all kinds of queries, however are according to be unstable below some circumstances. Although there are pros and cons for both types of LBRP techniques, the profile-based LBRP has demonstrated more effectiveness in improving the quality of web search recently, with increasing usage of personal and behavior information to profile its users, which is usually gathered implicitly from query history browsing history click-through data bookmarks user documents, and so forth. Unfortunately, such implicitly collected personal data can easily reveal a gamut of user's private life. In fact, privacy concerns have become the major barrier for wide proliferation of LBRP services.

2. SYSTEM ANALYSES

2.1 EXISITING SYSTEM

Among existing solutions in recommender systems RS, specifically, cooperative filtering (CF) strategies are shown to be wide effective. Based on the past behavior of users explicit ratings and implicit click logs. CF methods similarity between users behavior patterns as clicks and book marks. Most of CF methods low accuracy. These works requires collecting extra data. Sparsity problem. there are two algorithms used, 1.Collaborative DP (Discriminative Power) and collaborative IL (Information Loss) 2. The Brute Force Algorithm

2.2 PROPOSED SYSTEM

We develop a 1-injections by using a preferences for unrated items. The projected l-injection approach will improve the accuracy of top-N recommendation supported 2 strategies: (1) Preventing uninteresting things from being enclosed within the top-N recommendation. (2) Exploiting each uninteresting and rated things to predict the relative preferences of unrated things additional accurately. To identify the uninterested items into pre-use and post-use then preferences the items. There are two algorithms used, 1.Collaborative Filtering. 2. Naive Bays Algorithm.

3. PROBLEM DESCRIPTION:

The Main Modules In The Applications Are:

- User Interest Profiling
- Diversity and Concept Entropy
- User Preferences Extraction and Privacy Preservation
- Personalized Ranking Functions

3.1 USER INTREST PROFILING:

SSCFLR (Scalable Content Aware Collaborative Filtering for Location Recommendation) uses "concepts" to model the interests and preferences of a user. Concepts classified into two different types, namely, content concepts and location concepts. The concepts are modeled as ontology's, in order to capture the relationships between the concepts. We observe that the characteristics of the content concepts and location concepts are different. User query maintain and preferences the data.

3.2 DIVERSITYCONCEPT

ENNTROPHY

SCCFLR(scalable content aware collaborative filtering for location recommendation) media consists of a content facet and a location facet. What we are preference between the content facet and location facet. Then checking the integration between the content facet and location facet. Content facet more effective than location facet.

3.3 USER EXTRACTION PRIVACY PRESERVATION

Given that the concepts and click through data are collected from past search activities, user's preference can be learned. SCCFLR and then discuss preserves user privacy. SpyNB learns user behavior models from preferences extracted from click through data. spyNB to predict a negative set of documents from the unlabeled document set. The details of the SpyNB method can be found in. Let P be the positive set, U the unlabeled set and PN the predicted negative set obtained from the SpyNB method.

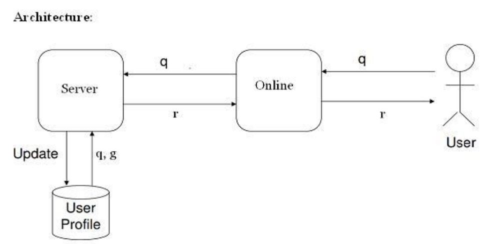
3.4 PERSONALISED RANKING FUNCTIONS

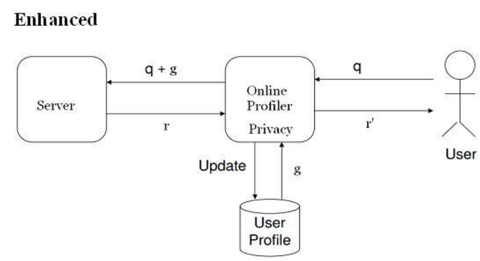
peoples can be learning personalized ranking functions are adaption of the search results according to the user content location preferences. Then we are creating to the ranking technique (SVM) RSVM. These ranking functions classified the two types:

- top- N recommendation
- top- Least recommendation.

4. SYSTEM DESIGN

SYSTEM ARCHITECTURE





4. SOFTWARE SPECIFICATIONS

Java is a high-level language that can be characterized by all of the following exhortations.

- Simple
- Object Oriented
- Distributed
- Multithreaded Dynamic
- Architecture Neutral
- Portable
- High performance
- Robust
- Secure

JAVA PLATFORM:

A platform is that the hardware or package surroundings within which a program runs. The java platform has two components:

1. The Java Virtual Machine.

2. The Java Application Programming Interface (API)

Development Tools:

The development tools offer everything you'll would like for collection, running, monitoring, debugging, and documenting your applications. As a new developer, the main tools you'll be using are the Java compiler (javac), the Java launcher (java), and the Java documentation (javadoc).

Application programming Interface (API):

The API provides the core practicality of the Java artificial language. It offers a large array of helpful categories prepared to be used in your own applications. It spans everything from basic objects, to networking and security.

MYSQL Server 2008

Microsoft MYSQL Server is a relational database management system developed by Microsoft.

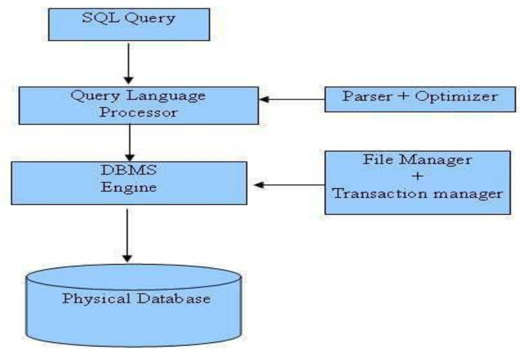
History

The history of Microsoft MYSQL Server begins with the primary Microsoft MYSQL Server product - MYSQL Server one.0, a 16-bit server for the OS/2 operating system in 1989 - and extends to the current day. As of Dec 2016 the subsequent versions square measure supported by Microsoft:

- MYSQL Server 2008
- MYSQL Server 2008 R2
- MYSQL Server 2012
- MYSQL Server 2014
- MYSQL Server 2016.

MY SQL PROCESS

When you are executing an MYSQL command for any RDBMS, the system determines the best way to carry out your request and MYSQL engine figures out how to interpret the task. There are various components included in the process. These components are Query Dispatcher,



4. THEORETICAL MODEL

Each web server will keep the user's access information to it. Usually, this information is called WEB Log including web server access log, proxy server log records, Browser log records, Users' brief introduction, users' registration information and users' dialogue or transaction information and so on. Establishing User Interests Model As we all know the real intent of this system is to achieve personalized information retrieval. So a data model must be created to do it. Each interested node is marked with a triad (pi, wi, xi) abbreviated Node (pi)In above expression, the value range of pi is P, marked with pięP, and P is words sets, marked with P= {p1! p2 !...!pm} 'in which p1 !p2 !...!pm are the interested words and m is the number of words. The wi is the weight of interested word pi; the xi is the fresh degree of word {pi}.

For the sake of the fact that different location of word in the document reflects different importance, the location word appears is taken into account, which is called location weight marked with sign . When calculating fresh degree of words, we use a fresh degree function f (n) to document d w if it n, n∧dnęD, Sign n refers to the nth document in buffers. Sign D is the document collection in buffers∨. The function f (n) is monotonous and non-decreasing which can assure that the more recent a document is visited, the more users are interested in it. So the weight and fresh degree of Node (pi) are calculated as follows.

Mining Association Rule

Through correlation analysis, such as algorithm a prior, relationships hidden among data are uncovered. Classification Analysis in the web log mining, the input set of classification analysis is group of record collection and several types of tags. First, each record is given a type tag. Then system checks these tags and describes the common features of these tags. Clustering Analysis

Clustering analysis is different from classification analysis. It is the method of classifying information things or users with similar characteristics. Sequential Pattern

Sequential pattern refers to seek out information things that ar sequent in time from the time-series information sets. In the web log mining, sequential pattern recognition means to find the user's requests for pages which are successive in time among user session.

5. ALGORITHM IMPLEMENTATION

Implementation is that the stage of the project once the theoretical style is clad into a operating system. Thus it may be thought-about to be the foremost important stage in achieving a undefeated new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful coming up with, investigation of the present system and it's constraints on implementation, coming up with of ways to attain transition and analysis of transition ways.

5.1 NAIVE BAYES ALGORITHM

Given that the concepts and click through data are collected from past search activities, user's preference can be learned. In this section, we first review a preference mining algorithms, namely SpyNB Method, that we adopt in PWS, and then discuss how PWS preserves user privacy. SpyNB learns user behavior models from preferences extracted from click through data. Assuming that users only click on documents that are of interest to them, SpyNB treats the clicked documents as positive samples, and predict reliable negative documents from the unlabeled (i.e. un clicked) documents.

To do the prediction, the "spy" technique incorporates a novel voting procedure into Naive Bayes classifier to predict a negative set of documents from the unlabeled document set. The details of the SpyNB method can be found in. Let P be the positive set, U the unlabeled set and PN the predicted negative set ($PN \subset U$) obtained from the SpyNB method. SpyNB assumes that the user would always prefer the positive set over the predicted negative set. Abstractly, the chance model for a classifier could be a conditional model over a dependent class variable with a small number of outcomes or classes, conditional on several feature variables through . The problem is that if the number of features is large or if a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We thus explicate the model to form it a lot of tractable.

Using Bays' theorem, this can be written

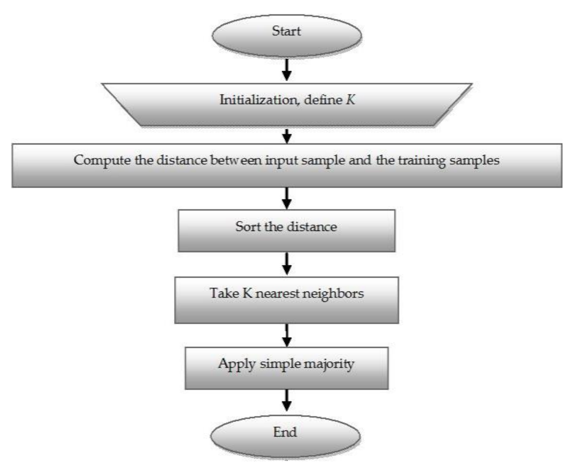
In plain English, using Bayesian Probability terminology, the above equation can be written as

In practice, there is interest only in the numerator of that fraction, because the denominator does not depend on and the values of the features are given, so that the denominator is effectively constant. The dividend is adore the probability model which may be rewritten as follows, exploitation the chain rule for perennial applications of the definition of conditional probability:

Now the "naive" conditional independence assumptions come into play: assume that each feature is conditionally independent of every other feature for , given the category . This means that and so on, for Thus, the joint model can be expressed as this means that under the above independence assumptions, the conditional distribution over the class variable is:

Where the evidence is a scaling factor dependent only on , that is, a constant if the values of the feature variables are known.

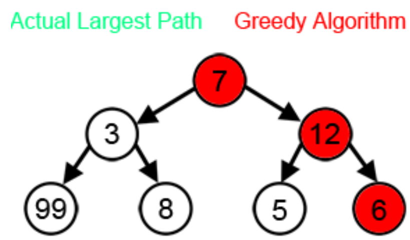
FLOW CHART NAIVE BAYES



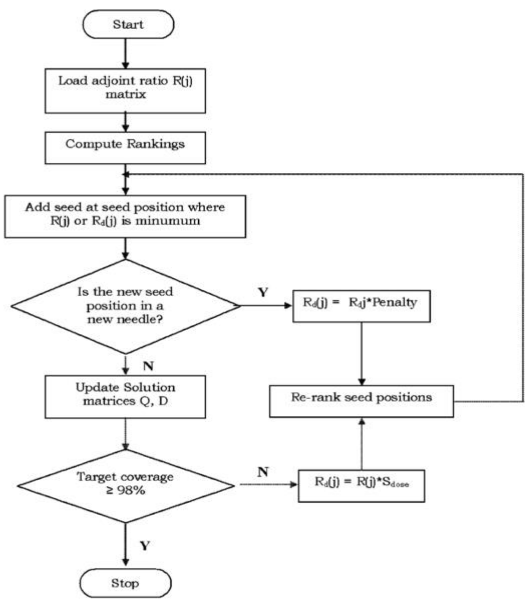
5.2 COLLABORATION FILTERING ALGORITHM

Collaboration algorithms mostly (but not always) fail to find the globally optimal solution, because they usually do not operate exhaustively on all the data. They can build commitments to bound decisions too early that stop them from finding the simplest overall answer later. For example, all known Collaboration coloring algorithms for the graph coloring problem and all other NP-complete problems do not consistently find optimum solutions. Nevertheless, they're helpful as a result of they're fast to dream up and infrequently offer sensible approximations to the optimum. Collaboration algorithms appear in network routing as well.

Using Collaboration routing, a message is forwarded to the neighboring node which is "closest" to the destination. The notion of a node's location (and hence "closeness")



FLOW CHART OF COLLABORATION ALGORITHM



6. CONCLUSION

It given a client-side privacy protection framework known as UPS for customized net search. UPS may probably be adopted by any PWS that captures user profiles in a very stratified taxonomy. The framework allowed users to specify customized privacy necessities via the stratified profiles. In addition, UPS conjointly performed on-line generalization on user profiles to safeguard the private privacy while not compromising the search quality. We proposed two Collaboration algorithms, namely Collaboration DP and Collaboration IL, for the online generalization. Our experimental results revealed that UPS could achieve quality search results while preserving user's customized privacy requirements. The results conjointly confirmed the effectiveness and potency of our resolution.

7. FUTURE WORK

For future work, I will attempt to resist adversaries with broader information, like richer relationship among topics (e.g., snobbishness, consecutive, and so on), or capability to capture a series of queries (relaxing the second constraint of the adversary) from the victim. We will also seek more sophisticated method to build the user profile, and better metrics to predict the performance (especially the utility) of UPS.

REFERENCE

[1] Defu Lian, Yong Ge, Fuzheng Zhang, Nicholas Jing Yuvan,Xing Xie, Tao Zhou, and Yong Rui, "Scalable Content-Aware Colloborative Filtering for Location Recpmmendation", IEEE Transactions on Knoweledge and ENGINEERING, 2018

[2] Q.Yuvan, G.Cong, Z.Ma, A. SUN, and N.M. Thalmann, "Timeware point-of-interest recommendation," in Proceedings of SIGIRI'1. ACM, 2013, pp. 363-72.

[3] Q.Yuvan, G. Cong, and A. Sun, "Graph-based point-of-interest

[4] 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. IEE, 2012,pp.144-153.

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

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

[7] K.Zheng, S.Shang, N.J.Yuvan, and Y.Yang,"Towards efficient search for activity trajectories," in Data Engineering(ICDE),2013.