

Location Based Services using Productive Filtering and Review Analysis

Praveen Raut, V. L. Kolhe

Department of Computer Engineering, D. Y. Patil College of Engineering Akurdi, Pune, India

ABSTRACT: In today's world, location based services are used in many applications. In these systems, any user request for a particular service through a query to a server and the server gives feedback according to location. For providing this facility, a push model or server initiate model is used. For using this facility, the user has to enter their interest for finding their need, because of this model, the user gets related information which is going to help them to get an exact match. The system creates an index, based on an R-tree by collecting the tokens into R-tree nodes. Using this model, it reduces some unnecessary information by filtering an algorithm which helps to increase the efficiency of the system and also reduces the cost of publisher/subscriber application. Our approach is to design a system which helps the user to get the relevant information through trusted reviews. Manipulated reviews are identified using Naive Bayes algorithm which increases the trustworthiness between the publisher and subscriber. Naive Bayes algorithm is applied on results generated by R-tree which helps to find fake reviews and ranking of results. It helps to increase the accuracy of results.

KEYWORDS: Indexing, Location, Naive Bayes, publish/subscribe, R-Tree, Review,

I. INTRODUCTION

Services which are made based on location are very useful for academic and industrial purposes. There are many other location based services available and it has to be located and accepted because they offer location aware experience. In this type of services, the user enters their choices and it responds to the user with a location aware answer. For providing this facility to the user, one push model is designed. For using this model, the user has to register spatio-text for finding their interest and the publisher sends text messages. This model gives a new experience to the user for finding their queries about location. Consider an example, a user wants to find an ATM machine, so he/she just puts a query 'ATM machine' and the LBS (location based services) system gives output which depends on the keyword and location which the user entered. There are a number of models used for finding locations. Following are some steps in the models:

Group on:

In this, the user inserts their interest with keywords and location, then the service provider replies to the user through a message who is interested in the message.

Location aware AdSense:

In this, the service provider registers his/her advertisement so that if any user wants to search any particular location.

Feedback:

In this, the service provider checks their feedback about their service and their product. In all these models, there are a lot of big problems of getting high performance. Because the service provider/system should support millions of customers and send them messages according to their queries. The replies of customers' queries contain both text information and location based information. For solving the above problem, the system builds a token based R-tree based R-tree index structure, (this is known as R-tree). This can be done by collecting every R-tree node with a pack of tokens which are selected with the help of R-tree and then apply a strategy of filter and verification so that messages get delivered efficiently. The system uses some good tokens for decreasing the number of tokens. Because of this system, it easily increases the performance and also decreases the index size. The publish-subscribe pattern of message, and

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 9, September 2017

particularly its content based incarnation, has shown its competence in large figure of domains. In most fields in which content-based publish-contribute catches its normal submission, an active message paradigm wants to take into account the position in which the material to be connected is consumed and produced.

II. REVIEW OF LITERATURE

Publish/Subscribe systems is a key technology for information dissemination. Each participant in a pub message system can also take the role of a publisher as well as the subscriber of information. Publishers create information in form of processes, which is used by subscribers issuing subscriptions representing their interest only in specific events.

A. Topic-based Model :

Notifications and messages are grouped in topics i.e., a subscriber declares its interest for a particular topic and will receive all procedures and events related to that topic. Each topic corresponds to some channel ideally which are logical connecting to each possible publisher to all interested subscribers. For the broadness, the difference between topics and channel is that topics are carried within an event as a special attribute. Using coarse grain correspondence, whichever network multicast services or diffusion trees, one for each topic, can be used to distribute events to interested subscribers [12]. The main disadvantages of the topic-based model is the limited articulateness it offers to subscribers. A subscriber interested in a division of events related to a specific topic receives also all the other events that belong to the same topic. To address problems related to very low expressiveness of topics, several solutions are abused in pub/sub implementations.

B. Content based Model:

Subscribers tell their interest by specifying conditions over the content of messages and notifications they want to receive. A filter in a subscription is a query formed by a set of constraints over the values of attributes of the notification construct through conjunction or disjunction operators. Possible limitations depend on the subscription language and on the attribute type. Most subscription languages comparison operators and comprise equality as well as regular expressions. The complexity of the subscription language obviously effects the complexity of matching operation. So it is not common to subscription of languages allowing queries more complex than those in conjunctive. A complete specification of content based subscription models explain in following examples of systems that fall under the content-based category are Subscribe, Ready [1], Gryphon SIENA [2], JEDI [3], Hermes [4], Elvin [5].

C. Type-based:

The type-based [8, 9] pub/sub variant events are actually objects belonging to a specific domain, which will thus encapsulate attributes as well as methods. With respect to unstructured models, simple Types represent a more robust data model for enforcing type safety application developer at the pub/sub system, rather than inside the application [10]. In a type-based subscription the declaration of a desired type is the main perceptive attribute. That is, with respect to the good models, type-based pub/sub sits itself somehow in the middle, by giving a coarse-grained structure on events on which fine-grained constraints can be applied over attributes (like in content-based) or over methods.

D. Spatial Keyword Search follow:

The first problem is knn based keyword search, which, given a set of keywords and location, finds top-k nearest neighbors by considering the textual relevancy and distance. Felipe et al. [6] integrated signature files and R-tree. Cong et al. [7] combined inverted files and R-tree. Author studied how to find top-k respect based spatial objects. The second problem is region based keyword search, which, given a key word query and region as well as finds the

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 9, September 2017

relevant objects in this region. Classifying review manipulation has become one of the hot research subjects in e-commerce because more and more customers make their purchase choices based on some individual comments from virtual societies and e-business websites. Customers consider these personal reviews are more reliable than the prevailing internet advertisements. Consequently, some organization attempts to create fake individual comments to affect customers mind and increase their deals. But, how to classify those manipulated reviews is a tough task for customers. As in many websites there are lengthy reviews related to any product but users are mainly concerned about the important and useful information related to that product, so the proposed system will provide relevant information related for any product which will save users time as they need not have to read those lengthy reviews. Manipulated reviews get detected which increases the customer trust. Today there is a trend that lots of social networking websites provide the users review on any product. The customers read all the reviews written by other users and get the idea about the product. These personal comments and reviews might influence the masses thinking and idea about the product. It is the human tendency that by reading the reviews of other users, one tries to look for that product and might change his/her thinking about that. Thus because of this reason many enterprises start generating their untrue reviews about their websites, products. Also today many websites manipulate the reviews i.e. when any customer writes the negative review about any website, enterprise better knows the value of those reviews and thus they manipulate those reviews [11]. The proposed model will detect these manipulated and fake reviews which will increase the customers trust towards an enterprise.

III. SYSTEM ARCHITECTURE / SYSTEM OVERVIEW

Publisher/subscriber system uses R-tree algorithm with filtering and verification which increases the efficiency of system and capture the user interest according to their locations and their reviews. The efficient filtering algorithm and effective pruning technique helps to achieve high performance. For classification of subscriber reviews naive bayes algorithm is used.

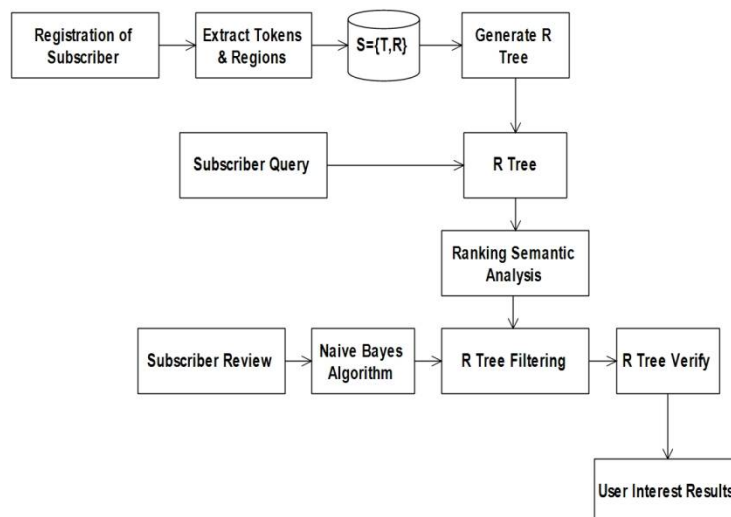


Fig. 1. Block Diagram of Proposed System

A. Registration of a subscriber:

In this proposed system, subscribers register subscriptions to capture their interests. T as token and subscriptions includes a textual description. The spatial information is used to capture a subscriber's most interested region.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 9, September 2017

B. Extracts Region and Tokens from subscriber data:

The textual description is used to capture a subscribers content-based interests, denoted by a set of tokens .

C. Extracts tokens and regions form publisher message:

A message m posted by a publisher also contains a textual description $m.T$ and spatial information $m.R$, denoted by $m=(T,R)$, which respectively have the same meaning as those of subscriptions. Note that the spatial information $m.R$ of a message can be a point, e.g., mobile users location. If the spatial information of a message is a point, call it point message; otherwise call it range message.

D. Generate R^t -tree:

A token based R-tree called R^t -tree, by mixing tokens of subscriptions into R-tree nodes. R^t -tree is a stable searchtree. Every leaf node covers between b and B data entries, where every entry is a subscription. Each internal node has between b and B node entries. Each entry is a triple (Child, MBR, Token Set), where Child is a pointer to its child node, MBR is the minimum bounding rectangle of all entries within this child, and Token Set is a set of tokens selected from subscriptions (or a pointer to the token set). A leaf nodes token set is the union of tokens of all subscriptions within this node and an internal nodes token set is the union of token sets of all entries within this node. As an entry corresponds to a node, for simplicity a node is mentioned interchangeably with its corresponding entry if the context is clear.

E. R^t -tree filtering and verification:

Algorithm 1. R^t -tree filtering

Input: S: A subscription Set; M: A message
Output: R Answer of M
Built an R^t -tree with root r
Initialize a candid node $CN=\emptyset$
 R^t FILTER (r,M,CN)
 $R= R^t$ Verify (M,CN)

For R^t -tree pruning system uses two filters, first is MBR filters which is invalidates spatial constraint as subscriptions and second one is token filters which is invalidates textual constraint. The nodes that are not pruned by the MBR filter and token filter are called candidate nodes. For pruning preorder traversing is used. After finding candidate nodes apply R^t -tree filtering algorithm. In the verification step we verify the subscriptions in the leaf candidate nodes generated in the filter step.

F. Review Analysis using Nave bayes Algorithm:

After filtering and verification Decision tree: In classification task, feature selection aims to select a subset of features that can lead to a highest performance. There are a no. of data mining and statistical methods such as decision tree, correlation analysis etc. have widely been used to remove unnecessary, redundant attributes. There are various feature selection methods but while using it practically its computational cost and complexity is calculated. A decision tree algorithm is a tree like structure where all attributes are assumed to be relevant. For implementation of decision tree algorithm TDM is created [11]. TDM is term document matrix where rows denote terms and columns denote sentences.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 9, September 2017

G. Matching of Reviews and tips:

Three types of similarity are performed for matching re-views and tips. Syntactic similarity if a sentence share same important keyword then it is called as syntactic similarity. The vector space model is a well established model for keyword similarity. Semantic similarity, if a sentence and tip defines the same concept; even if the same keyword is not used then it is called as semantic similarity. Sentiment similarity when tips and opinions share the same sentiments i.e. positive and negative then it is called as sentiment similarity.

IV. EXPERIMENTAL SETUP

A. Software Requirement

For the experimentation work JSP (Net Beans IDE 8.0) is used with Processor Pentium IV, RAM 1 GB & operating system Windows 7. MySQL server is used for storing review dataset, information about publishers and subscribers.

V. SYSTEM ANALYSIS

The results of the proposed method have been quantitatively evaluated using accuracy, precision and recall. By applying *R-tree* we increase efficiency without reducing quality. Nave bayes algorithm classify manipulated and no manipulated reviews. Figure 2 & 3 shows result of proposed method.

TABLE I
COMPARISON R^t AND R^t WITH NAIVE BAYES

Algorithm	Accuracy(%)
<i>R^t-tree</i>	84
<i>R^t-tree + Nave bayes</i>	93

TABLE II
COMPARISON R^t AND R^t WITH NAIVE BAYES

Algorithm	Precision	Recall
<i>R^t-tree</i>	0.6	0.7
<i>R^t-tree+Nave bayes</i>	0.78	0.9

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 9, September 2017

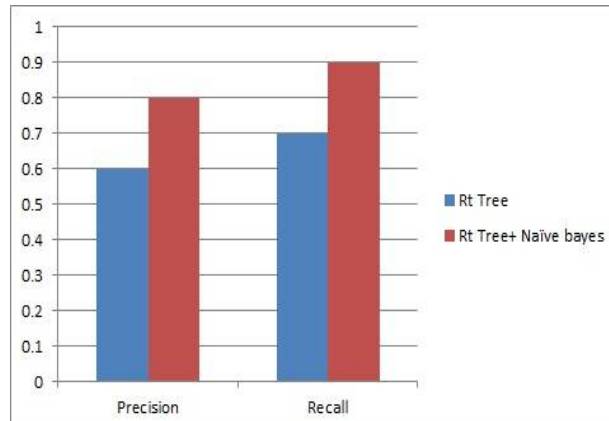


Fig. 2. Precision & Recall graph

VI. CONCLUSION

In proposed method, a R-tree is generated from subscriber and publisher data. R tree reduces number of tokens by using filtering and verification method, hence it decrease index size and the efficiency get increased. Proposed method gives better performance and increase the efficiency without increase in Error Rate. Ranking of results using reviews improve the accuracy and helps to identified authorized publishers.

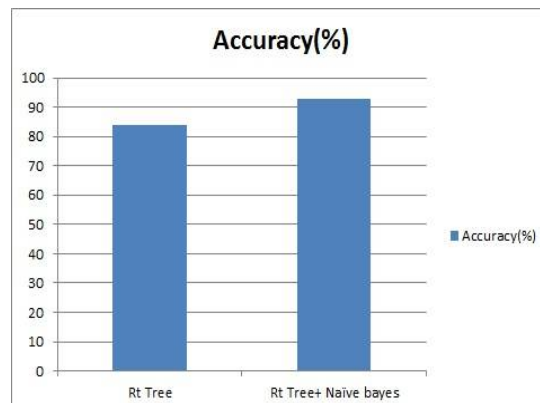


Fig. 3. Accuracy Graph

Nave bayes algorithm helps to identify fake reviews and increase customer trust towards enterprise by detecting untrue reviews. This system can be extended by using filters at testing and matching phase.

REFERENCES

- [1] R. Gruber, B. Krishnamurthy, and E. Panagos, "The architecture of theREADY event notation service", In Proceedings of the 19th MiddlewareWorkshop, 1999.
- [2] R. Strom, G. Banavar, T. Ch, M. Kaplan, K. Miller, B. Mukherjee, D. Sturman, and M. Ward. Gryphonn, "information flow based approach to message brokering", In Proc. of the International Symposium on Software



ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 9, September 2017

Reliability Engineering, 1998.

- [3] G. Cugola, E. D. Nitto, and A. Fuggetta. "The JEDI Event-based Infrastructure and its Application to the Development of the OPSS WFMS", IEEE Transactions on Software Engineering, 2001.
- [4] P. Costa and G. P. Picco, Semi-probabilistic content-based publishsubscribe, in Proc. 25thIEEE Int. Conf. Distrib. Comput. Syst., 2005, pp. 575585
- [5] R. Shah, R. Jain, and F. Anjum. "Efficient Dissemination of Personalized Information Using Content-Based Multicast", In Proceedings of IEEE Infocom, 2002.
- [6] I. D. Felipe, V. Hristidis, and N. Risse, Keyword search on spatial databases, in Proc. Int. Conf. Data Eng., 2008, pp. 656665.
- [7] X. Cao, G. Cong, and C. S. Jensen, "Retrieving top-k prestige- based relevant spatial web objects," Proc. VLDB Endowment, vol. 3, no. 1, pp. 373384, 2010.
- [8] J. H. Samet, "Foundations of Multidimensional and Metric Data Structures", San Mateo, CA, USA: Morgan Kaufmann, 2006.
- [9] M. Sharifzadeh and C. Shahabi, "Approximate voronoi cell computation on spatial data streams," VLDB J., vol. 18, no. 1, pp. 5775, 2009.
- [10] D. Wu, M. L. Yiu, C. S. Jensen, and G. Cong, "Efficient continuously moving top-k spatial keyword query processing", in Proc. Int. Conf. Data Eng., 2011, pp. 541552.
- [11] Long- Sheng Chen, Jui-Yu Lin, "A study on Review Manipulation Classification using Decision Tree", IEEE conference publication, 2013
- [12] L. Li, G. Xu, Z. Yang, P. Dolog, Y. Zhang, and M. Kitsuregawa, An efficient approach to suggesting topically related web queries using hidden topic model, WWW, pp. 273297, 2013.