International Journal of Engineering & Technology, 7 (1.3) (2018) 4-8

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Recommender system for business development using poi

R. Divya [1]*, S. Angel Latha Mary [2]

[1] *Department of Computer Science and Engineering, Karpagam College of Engineering, Coimbatore, Tamilnadu, India*
[2] *Department of Computer Science and Engineering, Karpagam College of Engineering, Coimbatore, Tamilnadu, India*
**Corresponding author E-mail: todivyasmailbox@gmail.com*

Abstract

In the technology era, internet access is unavoidable to find something or to share something with the world. Because of Globalization, the need for withstanding in the competitive business market is a challenging task. If we are to start a new business, ideas are needed. It is a tough task to decide where to start a new business with a scope of good profit as well as what type of business to start. If the business is a existing system, then how to increase the profit will be a challenging task. Unfortunately, no existing system is available for business recommendation. The proposed system is a step to solve this problem. Our paper deals with this challenge.Initially, the user's interest is obtained to fix a location as well as preferences were obtained for profit maximization. Then, the nearest neighbor is calculated from user's preferences, checks for available businesses and recommends the business type to the user. In this paper, the step to calculate nearest neighbor for the preferences that the user gives are elaborated.

Keywords: *KCG Finder, Gap Preservance, Spatial Inverted Index, User Preferences, Business Recommender System*

1. Introduction

The idea of collaborative spatial computing has been widely used in various domains including location based social networks , geo-crowd sourcing ,recommender system, activity planning , group decision making and disaster rescue . One of the most important applications of collaborative spatial computing in the database field is geo-social queries, which are attracting increasing interest from both industrial and academic communities. Recommender system is now popularly getting noticed by research people among other widely used domains .Business recommender system is a new upcoming research areas.

Technology's rapid development shares business based and location based data about a person. Grouping of both the data's will yield a new data called economy-spatial data. This data can be used in disaster rescue, activity planning, geo-crowd sourcing, spatial task outsourcing ,business plans recommendation and travel package recommendation. Though there were many areas in which economy-spatial data has its roots, business plans recommendation is an area which not only benefits an individual but also the society.

The Economy-spatial data can be used for improving business in a location. This in turn, increases the social development. Business is a competitive market where each follows some strategy to improve their income. All companies have their own department for not only improving their business, but also to analyze their business growth. But, for an individual either by own, thinks about an idea for business and starts collecting information from various resources or gets idea from family, neighbors ,well wishers and starts collecting the information.

After collecting the required information, analyzes the feasibility. If it is feasible to carry out , starts planning and works out to make that idea into a realistic one. Ir respective of whether the idea is own or from social circle, the individual has to do some case study in the idea what they have. Then they have to analyze the constraints, possibilities, permissions to obtain from government, competitors, success of business in the location where they wish to start and this to be factors increases largely by depending on an individual's thinking. To collect resources and to start analyzing feasibility, it consumes more time. There are lot of factors which are to be considered before making a decision to start a business in a particular area.

The proposed system solves this issue. It is to decrease the time consumption for feasibility analysis and to provide required guidelines to obtain necessary information for starting the business. Our major contributions are summarized below, a)The proposed system uses the Inverted spatial index structure which out performs other data processing index structure like SaR tree and B-Tree. b)For efficient processing of large amount of data, the query space is limited with the user's preference. c)The user's business idea were also considered to suggest a business .

The rest of the paper are organized as follows. Section 2 formally formulates the problem. Section 3 presents the related works and section 4 states about the proposed system architecture . Finally, we conclude the paper with the future work in Section 5..

2. Problem definition

The most widely used search engine Google's products such as Google Earth and Google Maps , as well as other geographic applications, returns locations as objects as a search result. They return such results by querying the spatial databases.Thus spatial queries had became predominant in recent years.

Many queries exist and some of them include pure spatial queries such as range queries, mck queries, RTknn queries, nearest neighbor queries , range queries and spatial joins Queries on spatial objects associated with textual data are represented by set of keywords get more interest from the spatial database research community and the industry.

Consider the Fig.1 below. Spatial databases returns a set of results for a search query. When we give the keyword for searching, the engine takes the keyword and it in turn produces three sets of data and locates them on Google map by pinpointing as a search result. The set of data it produces are,a) Name of the object.b)Review of the object.c) Address of the object.

So, it produces the search result objects as marking for all the resultant set objects. With this information, we get plenty of choices among which we have to select what we require. Since they were not listed in order, it gives a clumsy picture of non linear type. With this kind of search results, when we try to use them for obtaining some information through searching ,it provides a wide set of results. It takes more time to select what we require from them.It may even misguide the user to select a wrong option. Thus, it acts as the base problem that is to be treated.

Fig.1,is a screenshot that describes the above problem. The query gaved to the engine is "Restaurant near Gandhipuram with Pizza as menu". It returned a list of 10 restaurant names with their reviews, address and was pinpointed in Google Map. But it is clumsy and requires more time to select the right option without any business recommendation. In case, if the restaurant has more items offered ,we cannot taste each item and it is quiet expensive. When we use this system it would eliminate this problem.

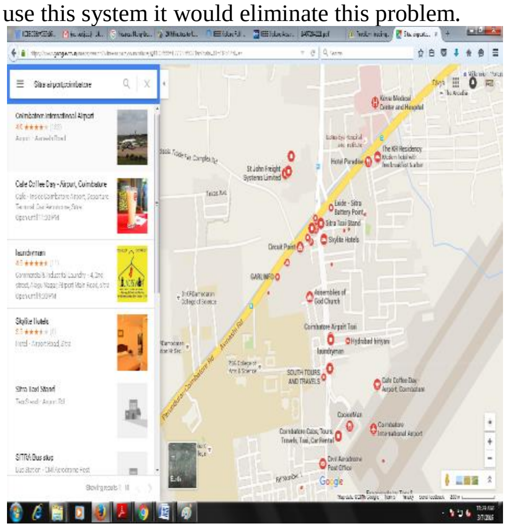


Fig 1.Screenshot depicting the business locations.

Google Maps is a product of Google which is used world wide. Around 1000 000 000 worldwide use Google Maps daily. There are roughly 55 Million unique visitors each month in the US. Google Earth has been downloaded over 500 000 000 times as on July,2015. But no system provides ranking on objects and recommendation based on it. In India, no product is specifically used for business recommendation in location basis. Google Maps , a product of Google which is used world wide is used here as a way for finding locations. But it does not provides ranking on objects and recommendation based on it. So, the system what we develop will be the first system for location aware business recommendation

3. Related Works

Some of the queries that were relevant to the problem is as follows. Nearest Neighborhood queries is a Query in which given a set W of points, a query point q, and a radius circle r, the NNH query returns the nearest position of r to q such that there are at least *m* points enclosed by r depicted as Circle C.[1].

It uses a Brute force approach to search the entire list for finding every nearest neighbor. Hence it consumes more time and cannot be used for practical applications.

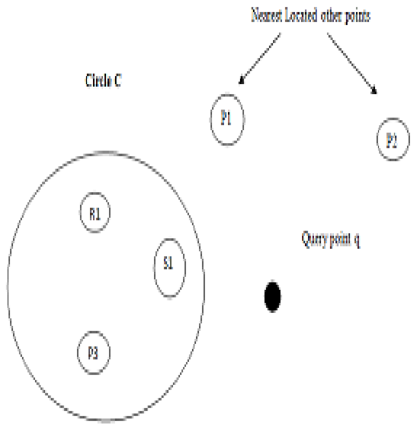


Fig.2 Nearest Neighbourhood Queries

Top-k local user search (TkLUS for short) from tweets with geo-tags. Given a location q , a distance r , and a set of keywords W , the TkLUS query finds the top-k users who have posted tweets relevant to the desired keywords in W at a place within the distance r from q . TkLUS queries are useful in many application scenarios such as friend recommendation, spatial decision, etc. The user set U and the post set P are very large on many social media platforms like Twitter. It is definitely inefficient to check the sets iteratively. Also, it is not straightforward to measure the relevance of a user to a given query.[2]

Circle of Friend Query (CoFQ) is another type of query which works by finding a group of friends in a Geo-Social network whose members are close to each other both socially and geographically. More specifically, the members in the group have tight social relationships with each other and they are constrained in a small region in the geospatial space as measured by a "diameter" that integrates the two aspects. Here "circle of friends" refers to the group of k + 1 users including q with small pairwise social distances in a social network. And "spatial proximity" is to constrain the diameter of the point set formed by the locations of the k +1users.

There are many challenges in answering k-CoFQ in a large GSN. The major challenge is ,The k-CoFQ requires to find a subgraph with the smallest diameter. The NP-hard max-clique problem can be reduced to the k-CoFQ problem which Makes k-CoFQ NP-hard and computational expensive. With spatial prox-imity considered, it is even more challenging to design a scalable algorithm.

Considering that k -Nearest Neighbor (k -NN) search is one of the most popular web interfaces in accessing spatial data on the web, The problem of retrieving geospatial data from the web for a given spatial range query using only k -NN searches were studied. Besides web data integration and search applications, the solution to this more general problem is beneficial for other application domains in the areas of sensor networks, online maps and ad-hoc mobile networks where their applications require great deal of data integration for data analysis and decision support queries. For example, the McDonalds web site provides a restaurant locator service through which one can ask for the five closest restaurants from a given location. This type of web interface to search for a number of "nearest neighbors" from a given geographical point (e.g., a mailing address) or an area (e.g., a zip code) is very popular for accessing geospatial data on the web. It nicely serves the intended goal of quick and convenient dissemination of business location information to potential customers. However, if the consumer of the data is a computer program, as in the case Geo informatics of web data integration utilities (e.g., wrappers) and search programs (e.g., crawlers),such an interface may be a very inefficient way of retrieving all data in a given query region.[5]

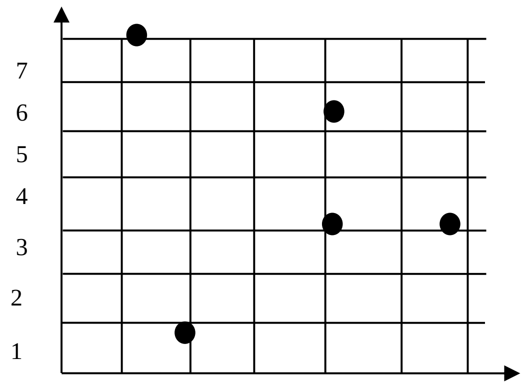
Nearest Neighbor queries are a frequently encountered type of Query in Geographic Information Systems and it is used to find the k-Nearest Neighbour objects to given a given point in space. Processing such queries requires substantially different search algorithms than those for location or range queries.[6]

Given a set of persons D connected in a social network SN with information abouttheir current location, a geo-social skyline query reports for a given user U D and a given location P (not necessarily the location of the user) the pareto-optimal set of persons who are close to P and closely connected to U in SN.They exclude the query node from the result set: In the foreseeable applications including oneself in the result gives no additional information gain, but may actually prune and therefore exclude other nodes from the result. This design decision is without loss of generality; depending on application details the query node itself may be allowed to be a valid result as well. Yet, They haven't specified the distance measures involved in the GSSQ.[4]

GSKCG Queries work as follows. Given a set of spatial query points and an underlying social network, a GSKCG query finds a minimum user group in which the members satisfy certain social relationship and their associated regions can jointly cover all the query points. For the spatial factor, instead of finding a group of users close to the query points (e.g.,Spatial task sites or a rally point), a GSKCG query finds a user group whose associated regions (e.g,Service regions or familiar regions) jointly cover a set of query points; for the social factor, we employ the more reasonable k-core notion to measure the intensity of the relationships of users in the selected group, for example, each user should be familiar with at least k other users. For this reason, the techniques developed for previous geo-social queries cannot be directly applied to our problem.[7]

We use the spatial data, basic facilities and landmarks to suggest a suitable business idea in the preferred location. In addition, we provide the competitors information based on Inverted index tree. It considers all the factors that were most basic and provides the suggestion. The data what they provide will be an updated one. Thus eliminates the time consumed for an individual to carry out the case study in the particular area.

There are different spatial data structures that can be used to solve our problem. Among them inverted spatial index is as follows.



Where,
○ - Query point
- Multi dimensional points

Fig.2(a): Locations of points

Let P be a set of multidimensional points. As we combine keyword search with location and textual information on facilities such as restaurants, hotels, etc.,. Here, we mainly focus on dimensionality 2 by considering the points in P with integer co ordinates[0,t],where t is a large integer. The valued coordinates represented in 2D is still finite and enumerable. So with proper scaling we proceed with such consideration.

Each character represented in the graph is an object in P with its textual data as a document is represented by W_p W_p .For example, if alphabets stands for restaurant, W_p W_p can be its menu. In addition, it may also have different useful informations.

A Nearest neighbour (NN query) gives a point q and a set W_q W_q of keywords(W_q W_q as the document of the query).It returns the point in P_q P_q that is the nearest to q, where P_q P_q is defined as,

$$P_q = \{character \in P | W_q \subseteq W_p\}.$$
$$P_q = \{character \in P | W_q \subseteq W_p\}. \quad \ldots \quad (1)$$

In simple, P_q P_q is the set of objects in P whose documents contain all the keywords in W_q. W_q. If the P_q P_q returns is empty, the query returns nothing. This problem as an overall, can be considered as *k* nearest neighbour (*k* NN) search, which finds the *k* points, the entire P_q P_q should be returned.

For example, assume that P consists of 6 points whose locations and a query point q are given as black dots and white dots respectively in the figure 2(a). Table 1 describes the associated text for Characters as below.

Table 1: Associated text for Characters.

P	W_p
A	{a,b}
B	{c}
C	{d,e}
D	{f,g,h}
E	{a,f,g}
F	{c,g,h}

Consider the query point q at the white dot of fig.2(a) with the set of keywords W_q W_q ={f,g}.Nearest neighbour finds D as the nearest neighbour as F misses {f} in it. If k=2,In addition, E is also returned. So the result set contains two character points namely {D,E}.The result set remains unchanged for k=3 or higher values as they were the only two objects that have both the keywords{f,g}. Inverted Indexes(I-index) have proved to be an effective access method for keyword-based document retrieval. Consider the table 2 below,

Table 2: Example of an inverted index(I-Tree)

Word	**Inverted list**
a	A,E
b	A
c	B,F
d	C
e	C
f	D,E
g	E,F
h	D,F

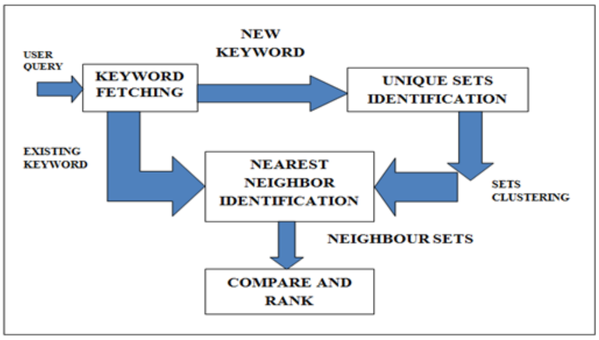
It contains the index for the data set of Table 1.Each word in the vocabulary has an inverted list by pinpointing the ids of the points that have the word in their documents. The list of each word maintains a sorted order of point ids. Thus, it provides considerable convenience in query processing by allowing a merge step.

Given a nearest neighbour query q with the keyword set W_q W_q ,the query algorithm of I-Index first retrieves the set P_q P_q of the points that have all the keywords of W_q W_q .Then it ranks them based on the distance from the centre point. Then, our system analyses the menu items which were unique with the menu items of the resultant restaurant's list. Then it considers as how many menu items were repeated and recommends the items that are to be added for our business.

In case, if no restaurant is available with such queried menu items, then it recommends the new such menu item if we opt to start restaurant. Thus, it eliminates the time consumed for gathering information about how to develop business. Thus, the proposed system helps the individual as well as society's development by reducing time and saving cost.

4. Proposed system

Architecture of the proposed system to solve the problem is as below,



The modules that have been identified to solve the problem is as below,

Module 1-Keyword fetching.

This module is used to get the user query and fragment into pieces of keywords. These keywords act as the preferences and the user's idea for starting a new business. These keywords will be used in upcoming modules for business recommendation. It uses "sub string matching" as the key technique.

Module 2-Unique sets identification.

This module is used to identify the keywords as "existing" or "new". Depending on the keyword the processing starts. Whether the keyword is "new" or "existing" ,the system selects the preference data set given by the user. These two keywords act as the major words that decides flow of execution in the system. It uses "sub string matching" as the key technique. This module is going to use Inverted Index spatial data structure.

Module 3-Nearest Neighbor identification.

This module is used to identify the nearest neighbour satisfying all the keywords. If the business that the user wishes to start is a new business, then suitable areas for starting the business would be identified and given to the user. This module uses the IR^2 IR^2 tree algorithm to compute the nearest neighbour for the preferences that the user has given.

The IR^2 IR^2 Tree is a combination of an R -Tree and signature files. In particular, each node of an IR^2 IR^2 Tree contains both spatial and keyword information; the former in the form of a minimum bounding area and the latter in the form of a signature. An IR^2 IR^2 Tree facilitates both top -k spatial queries and top -k spatial keyword queries.

Module 4-Compare and rank.

This module is used to obtain search results and display them as ranking for user's convenience. This rank will be further used to recommend suitable business for the user.

Detailed submission guidelines can be found on the journal web pages. All authors are responsible for understanding these guidelines before submitting their manuscript.

5. Conclusion and future work

Spatial databases though was limitedly used ,it can also be used to extend in information processing systems like business recommender system, travel recommender system, crowd sourcing, task out sourcing and other applications. Data structure utilization for such a product decides its performance and data set type decides its usability level from the user's point of view. Achieving 100% accuracy in such a system is practically impossible for such a system. It is due to the different views that a user in selecting the business area and the importance level that they give to the selection criteria. But, this system is an attempt to minimize the work load of a person to start a new business as well as to improve his business. As a future work, the system is planned to be extended with more real world data sets and to give more accurate results for existing users.

References

[1] Nearest Neighborhood Search in Spatial Databases, IEEE International Conference on Data Engineering,(ICDE),2015.
[2] Finding top-k local users in geo-tagged social media data ,Published in Data Engineering (ICDE), M.Sozio,2015 IEEE 31st International Conference on April 2015.
[3] Circle of friend query in geo-social networks,Proceeding in the International Conference on Database Systems for Advanced Applications,B.Gedik,C.Masolo,Volume-Part II,2012.
[4] Geo-Social skyline Queries,Proceedings in the International Conference, Q.Shan,Y.Tao,DASFAA 2014, 2014.
[5] Web data retrieval: solving spatial range queries using ***k***-nearest neighbor searches,T.Lappas,Springer,DOI-10.1007/s10707-008-0055-2,2009-12.
[6] Nearest Neighbour Queries,M.Zavollini,ACM SIGMOD International Conference on management of data ,1995.
[7] Geo-Social K-Cover Group Queries for Collaborative Spatial Computing, I.V.Hicks,IEEE Transaction, IEEE Transactions on Knowledge and Data Engineering,October 2015.
[8] Nearest neighbor search, W.C.Lee,C.Chen ,SODA '00 Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms,2000.
[9] Software fault prediction using Quad-tree based K-Means Clustering Algorithm, Q.Zhu,H.Hu, IEEE Transactions on Knowledge and Data Engineering,Volume4,June 2012.
[10] A R-Tree based Fine Directional Query Filtering, Internet Computing for Science and Engineering (ICICSE),Y.Li,D.Yu, 2009 Fourth International Conference on , 2009.
[11] Fast Nearest Neighbor Search with Keywords,Yufei Tao and Cheng Sheng,IEEETransactions on Knowledge and Data Engineering,Vol.26,No.4,April 2014.
[12] INSPIRE:A framework for incremental spatial prefix query relaxation, YuxinZheng, Zhifeng Bao,(et.al.,),IEEE Transactions on Knowledge and data Engineering,Vol.27,No.7,July 2015.
[13] Nearest neighbor queries,N.Roussopopulous,S kelley and F.Vincent., In Proceedings of ACM SIGMOD International Conference on Management of Data(SIGMOD),1995.
[14] Towards an Analysis of Range query Performance in Spatial Data Structures., In Proceedings of the 20th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database systems(PODS),2000.
[15] Location-Aware Type Ahead Search on Spatial Databases:Semantics and Efficiency., In Proceedings of ACM SIGMOD International Conference on Management of Data(SIGMOD),1995.
[16] The R*-tree:An Efficient and Robust Access Method for Points and Rectangles,A.Land and A.Doig Proc.ACM SIGMOD Int'l Conf.Management of Data,1990.

[17] Collective Spatial Keyword Querying,L.Kazemi,C.Shahbi, Proc.ACM SIGMOD Int'l Conf.Management of Data,2011.
[18] Efficient Query Processing in Geographic web Search Engines, W.Liu,W.Sun,Proc.ACM SIGMOD Int'l Conf.Management of ata,2006.
[19] Approximate string search in spatial data-bases,Y.Zhou,X.Xie,Y.Gong Proc.Int.Conf.Data Eng.,2010.