Information Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys

# Reverse nearest neighbor search with a non-spatial aspect

Jeong-hoon Park, Chin-Wan Chung *, U Kang

*Department of Computer Science, Korea Advanced Institute of Science and Technology, Republic of Korea*

ARTICLE INFO

*Article history:*
Received 16 December 2014
Received in revised form
10 May 2015
Accepted 26 June 2015

*Keywords:*
Reverse nearest neighbor
Spatial
Non-spatial
RNN
LBS

ABSTRACT

With the recent surge in the use of the location-based service (LBS), the importance of spatial database queries has increased. The reverse nearest neighbor (RNN) search is one of the most popular spatial database queries. In most previous studies, the spatial distance is used for measuring the distance between objects. However, as the demands of users of the LBSs are becoming more complex, considering only the spatial factor as a distance measure is not sufficient. For example, through a hotel finding service, users want to choose a hotel considering not only the spatial distance, but also the non-spatial aspect of the hotel such as the quality which can be represented by the number of stars. Therefore, services that consider both spatial and non-spatial factors in measuring the distance are more useful for users. In such a case, techniques proposed in the previous studies cannot be used since the distance measure is different. In this paper, we propose an efficient method for the RNN search in which a distance measure involves both the spatial distance and the non-spatial aspect of an object. We conduct extensive experiments on a large dataset to evaluate the efficiency of the proposed method. The experimental results show that the proposed method is significantly efficient and scalable.





## 1. Introduction

In recent years, location-based service (LBS) becomes one of the most popular trends in Web-based services. Because LBSs utilize spatial position information, diverse techniques in the spatial database area are employed. Among many spatial database techniques supporting LBSs, the reverse nearest neighbor (RNN) search and its variations are broadly used. Given a query object, the RNN search finds the set of objects that consider the query object as their nearest neighbor. For example, the advertisement target finding is an application using the RNN search. For a given cafe, the service finds the homes for which the cafe is closer than all the other cafes by using the RNN search. Since the homes in the result are considered as the potential customers, the cafe can send the brochures and coupons to the homes.

Lots of researches on the RNN search [1–11] have been conducted over the last decade. In [1–6], methods for finding the reverse nearest neighbor from a snapshot of a dataset are devised. The authors in [7–11] propose methods for the continuous nearest neighbor search which find the reverse nearest neighbor and continuously update the result as the objects change their locations. All the researches use only the spatial distance for measuring the distance between objects. However, in many real-life LBSs, we are given more information than just the physical locations of objects. Recent services

* Corresponding author.
*E-mail addresses:* jhpark@islab.kaist.ac.kr (J.-h. Park), chungcw@kaist.edu (C.-W. Chung), ukang@cs.kaist.ac.kr (U. Kang).

http://dx.doi.org/10.1016/j.is.2015.06.010
0306-4379/

such as Google Maps,[1] Facebook[2] and Groupon[3] provide users with rating scores of products. Such information is valuable in defining a new distance measure in the RNN search. For example, when choosing a premium-grade steakhouse for dinner, we generally consider not only the spatial proximities of steakhouses, but also the quality of the steakhouse based on items such as the food, atmosphere, price and service. Therefore, in order to choose the better steakhouse, the spatial proximity and the quality of the steakhouse are necessary to be comprehensively considered. In such a case, the traditional distance measure based solely on the spatial distance cannot be used. Consequently, the distance measure based on both the spatial proximity and the quality of the item is more useful and realistic than the traditional measure.

In this paper, given a set of items $\hat{I}$ and a set of users $U$, we introduce a new distance called the IU (Item-User) distance which is a distance measure between an item and a user considering the spatial distance and the non-spatial aspect of the item. Then, we propose an efficient method for the problem of the reverse nearest neighbor search with the non-spatial score. Specifically, given a query item, the problem is to find the users to which the query item is the nearest item based on the IU distance. Our problem supports updates of users and items, and the values of parameters for the IU distance can be given at the query time instead of being pre-determined. Since the distance measure is different from the previous researches, traditional approaches cannot be used in our case.

In the proposed method, we define the domination relationship among items, by using the properties of the domination relationship, we devise an efficient algorithm, given a query item $I_q$, to find the items having the domination relationships with $I_q$. Since an item dominated by another item cannot be a candidate of the nearest item for all the users w.r.t. the IU distance, the dominated items can be pruned without considering users. Then, we propose an efficient algorithm for the RNN search in which a 2-layered structure is devised to avoid redundant visits in synchronously traversing 2 R-tree indexes, one for users and the other for items. In addition, the three pruning techniques are developed and used in the algorithm utilizing the threshold of the IU distance and the spatial distance. These techniques incrementally prune the items and users considering each other.

The applications of our problem include a more practical marketing support system. People want to find a gas station such that the total cost for visiting the gas station and filling the gas is minimum based on their locations. Therefore, the marketing targets of a gas station are promising buyers for which visiting the gas station is more economical than visiting other gas stations. The method for our problem can find such promising buyers by considering the spatial distance and the price as the non-spatial aspect.

To the best of our knowledge, this is the first work that addresses this problem. Note that our problem cannot be a special case of the reverse spatial and textual RNN search [12]. It is because the existing spatial and textual RNN search problem is a monochromatic search which considers single type objects while our problem is a bichromatic search that considers two types of objects.

Our contributions are as follows:

- (*Introduction of the bichromatic RNN search with a non-spatial aspect*) We firstly address the problem of the bichromatic reverse nearest neighbor search with a non spatial aspect. In our reverse nearest neighbor search problem, instead of the traditional distance measure, we employ a new distance measure named the IU(Item-User) distance that considers both the spatial proximity and the quality of the item.
- (*Item pruning method*) We propose an effective method to filter out items before performing the RNN search. We define the domination relationship among items. Then, only the domination relationships of items with a given query item are used for filtering items. As a result, the proposed method can reduce the search space of items without considering the location of users. In addition, we statistically analyze the performance of the item domination.
- (*Novel search algorithms*) We propose a novel search algorithm for the RNN search with a non-spatial aspect, based on a 2-layered structure for maintaining the contexts of search between users and items.By the 2-layered structure, the algorithm avoids redundant computations. In addition, three pruning techniques are devised and used in the algorithm for incrementally reducing the search spaces for users and items.
- (*Experiments on a large data set*) We conduct extensive experiments for evaluating the efficiency of the proposed method using synthetic datasets and real datasets. The experimental results show that the proposed method is at least 4 times more efficient than an adapted version of an existing method

This paper is organized as follows. Section 2 reviews related works on the reverse nearest neighbor search. The problem is formally defined in Section 3. In Section 4, we present the index strategy. Section 5 describes the item domination and a method for the item pruning using the item domination relationship. An efficient algorithm for the RNN search is proposed in Section 6. In Section 7, we present experimental results. Finally, conclusions are made in Section 8.

[1] https://maps.google.com
[2] http://facebook.com
[3] http://www.groupon.com



Please cite this article as: J.-h. Park, et al., Reverse nearest neighbor search with a non-spatial aspect, Information Systems (2015), http://dx.doi.org/10.1016/j.is.2015.06.010