Expert Systems With Applications 43 (2016) 93–108

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Capacity constrained maximizing bichromatic reverse nearest neighbor search

Fangshu Chen, Huaizhong Lin*, Yunjun Gao, Dongming Lu

College of Computer Science and Technology, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China

ARTICLE INFO

Keywords:
MaxBRNN query
Capacity constraint
Optimal region

ABSTRACT

When planning a new development (facility/service site), location decisions are always the major issue. In this paper we introduce a novel query capacity constraint MaxBRNN, which can solve the facility location selection problem efficiently.

The MaxBRNN (maximizing BRNN) query is based on bichromatic reverse nearest neighbor (BRNN) query which uses the number of reverse nearest customers to model the influence of a facility location. The MaxBRNN query has been appreciated extensively in spatial database studies because of its great potential in real life applications, such as, markets decision, sensor network clustering and the design of GSM (global system for mobile communication). The existing researches mostly suppose that the service facility's capacity is unlimited. However, in real cases, facilities are inevitably constrained by designed capacities. For example, if the government wants to select a new place to set up an emergency center to share the existing centers' patients, they need to know the current emergency centers' capacity so that they can estimate the new center's scale. Thus, the capacity constrained MaxBRNN query is significantly important in planning a new development. As far as we know, the capacity constrained MaxBRNN query has not been studied yet, so, we formulate this problem, propose a basic solution and develop some efficient algorithms for the query.

Our major contributions are as follows: (1) we propose a novel query capacity constraint MaxBRNN which can solve the facility location selection problem effectively and efficiently; (2) we develop a basic algorithm CCMB and two improved algorithms which can find out the optimal region in terms of building a new facility, maximize its impact and deal with the complicated reassignment when adding new facilities into the dataset; (3) we prove the algorithms' effectiveness and efficiency by extensive experiments using both real and synthetic data sets.

1. Introduction

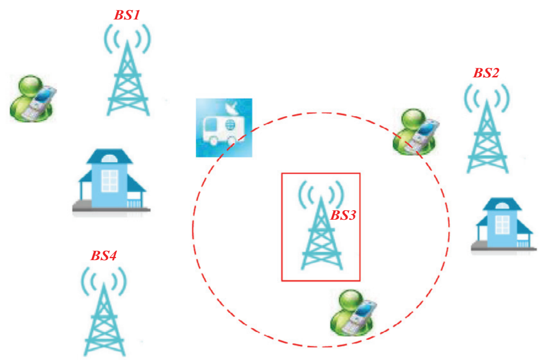
Given two data sets: the service (i.e., facility) point set $\mathcal{P}$ and the customer point set $\mathcal{O}$. For a service point $p \in \mathcal{P}$, a BRNN query finds all the points $o \in \mathcal{O}$ whose nearest neighbor in $\mathcal{P}$ is p. The customer points o in $\mathcal{O}$ constitute the influence set of p and the influence value of p equals to the cardinality of the influence set. For example, in Fig. 2(a), for a service point p_2, its BRNN, i.e., the influence set, is $\{o_4, o_5\}$.

The MaxBRNN problem (Cabello, Díaz-Báñez, Langerman, Seara, & Ventura, 2005, 2010) aims to find the region S such that setting up a new service site within this region might attract the maximal number of customers, namely the cardinality of BRNN set of all points p in S is maximized in a space. For example, in Fig. 2(a), there are five customer points o_1 to o_5 and five service points p_1 to p_5. The MaxBRNN problem is to find the optimal region, where all the p points has the maximum number of BRNNs, in this figure, the circles are called NLC (nearest location circle) in Cabello et al. (2005) or NLR (nearest location region) in Lin, Chen, Gao, and Lu (2013), the center of the circle is a given customer point in $\mathcal{O}$ and the radius is the distance from the customer to its nearest neighbor in $\mathcal{P}$. To find out the problem solution is to find regions that are enclosed by the largest number of NLCs. For example in Fig. 2(c), the MaxBRNN region is the black shaded area which is the intersection of the three circles o_1, o_2, and o_3.

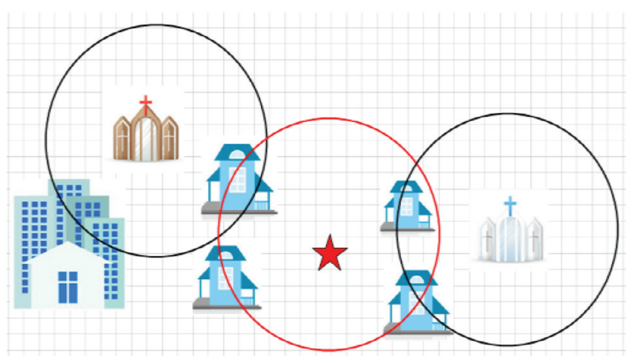
The MaxBRNN problem has many interesting real life applications in service location planning and emergency schedule such as the example in Fig. 1. MaxBRNN problem has been studied in Cabello et al. (2010), Liu, Wong, Wang, Li, and Chen (2012), Wong, Özsu, Yu, Fu, and Liu (2009), and Zhou, Wu, Li, Lee, and Hsu (2011), in which it is supposed that the service facility's capacity is unlimited. However in real cases, facilities are inevitably constrained by designed capacities, when the needs of service increase, facilities in those booming areas may be overloaded, and some facilities may run out of capacity.

* Corresponding author. Tel.: +86 13018939989.
E-mail addresses: youyou_chen@foxmail.com (F. Chen), linhz@zju.edu.cn (H. Lin), gaoyj@zju.edu.cn (Y. Gao), ldm@zju.edu.cn (D. Lu).

http://dx.doi.org/10.1016/j.eswa.2015.08.051
0957-4174/



(a) Adjust the BS's channel bandwidth when necessary



(b) Location selection for a new emergency center

Fig. 1. Practical examples.

In this paper, we introduce a novel query, the capacity constrained MaxBRNN query which aims at finding a region which can attract the largest number of customers when building a new service facility under the capacity constraint of the existing service sites.

This problem has many real life applications. A common problem many companies are facing with is to establish new facilities in most suitable places to maximize the new site's economic benefits and share the booming area's overload. For example in Fig. 1(a), suppose that at first, only two mobile signal base stations ($\boldsymbol{BS_1}$ and $\boldsymbol{BS_2}$) are working, and with the increase of the volume of customers, $\boldsymbol{BS_1}$ and $\boldsymbol{BS_2}$ become overloaded. The network operators need to open up a new facility to share the existing facilities' workload maximally, From Fig. 1(a), we can see the best choice is to open up $\boldsymbol{BS_3}$ in Fig. 1(a). Another example is shown in Fig. 1(b), if the government wants to select a new place to set up an emergency center to share the existing centers' patients, they need to know the current emergency centers' capacity so that they can estimate the new center's scale and maximize the new center's influence, in Fig. 1(b), it is obvious that the best choice for a new facility is the red star in the circle.

The above examples both face a main issue: how to choose the best location and find it out efficiently, and this is what we are going to talk about in this paper.

When solving the problem, we can make the following assumptions:

(i) A customer seeks service from its nearest facility which has remaining capacity, and if the facility has no remaining capacity, the customer will go to the second nearest facility and so forth.
(ii) A service site prefers providing service to clients nearer to it.
(iii) The capacity of the new site in our returned region R is unlimited since our goal is to maximize the new service site's economic benefits, and it has little impact on the algorithms' performance if we set the new site's capacity to a given constant.

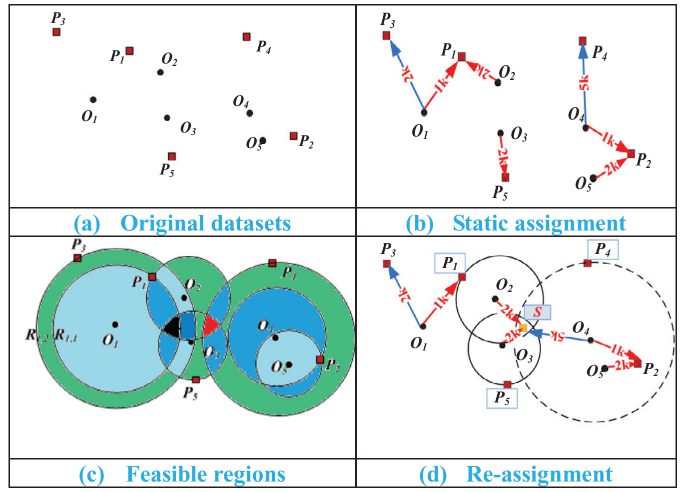


Fig. 2. An example of capacity constrained MaxBRNN. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

Table 1
Weight of dataset and the assignment.

(a)

$\mathcal{O}$	Weight	$\mathcal{P}$	Capacity
O_1	3k	P_1	3k
O_2	2k	P_2	3k
O_3	2k	P_3	3k
O_4	6k	P_4	5k
O_5	2k	P_5	3k

(b)

Assignment	Weight
o_1,p_1	1k
o_1,p_3	2k
o_2,p_1	2k
o_3,p_5	2k
o_4,p_2	1k
o_4,p_4	5k
o_5,p_2	2k

Fig. 2 shows an example of the capacity constraint MaxBRNN problem, there are five customer points O_1 to O_5 and five service points P_1 to P_5 (Fig. 2(a)), the customer points' weights are given in Table 1(a), which can be regarded as the population in the site, similarly the service points' capacities are given in Table 1(b).

If we assign each customer to his nearest neighbor according to above assumptions, we can get a static assignment strategy, as shown in Fig. 2(b), since we assume that a service site prefers providing service to clients nearer to it, we can find that customer points O_1 and O_4 are poorly served since they are out of the capacities of their nearest service sites (P_1, P_2), and have to go to a further site. Next, we draw an NLR (Lin et al., 2013) circle (the center of the circle is a given customer point in $\mathcal{O}$ and the radius is the distance from the customer to its nearest neighbor in $\mathcal{P}$) for each customer, we can get Fig. 2(c). We call the colored areas in Fig. 2(c) *feasible regions – the smallest possible region to locate a new service site*, we will discuss the definition of it in Section 3.

Follow the algorithms in Liu et al. (2012); which do not consider the capacity of service site, we can find the optimal region is the black feasible region in Fig. 2(c), since this region is covered by the largest number of NLRs, the region's weight is 7 ($w_{o2,p1} + w_{o3,p5} + w_{o1,p3} + (w_{o1} - w_{o1,p3})$). (Notice that the weight of circle O_1 equals to the total weight of point O_1 (i.e., w_{o1}) minus the weight that has been assigned to P_3 (i.e., $w_{o1,p3}$).). However, if we sum up the weight of the red feasible region in Fig. 2(c), it is 9 ($w_{o2,p1} + w_{o3,p5} + w_{o4,p4}$). So the red feasible region is the actual