Information Systems 62 (2016) 136–154

Contents lists available at ScienceDirect

Information Systems

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

Efficient multiple bichromatic mutual nearest neighbor query processing

Marta Fort *, J. Antoni Sellarès

Departament d'Informàtica, Matemàtica Aplicada i Estadística, Universitat de Girona, Spain

A R T I C L E I N F O

Article history:
Received 25 February 2016
Received in revised form
30 June 2016
Accepted 7 July 2016
Recommended by: D. Shasha
Available online 16 July 2016

Keywords:
Computer science
Decision-making support system
Bichromatic mutual nearest neighbor query
Graphics Processing Unit (GPU)

A B S T R A C T

In this paper we propose, motivate and solve multiple bichromatic mutual nearest neighbor queries in the plane considering multiplicative weighted Euclidean distances. Given two sets of facilities of different types, a multiple bichromatic mutual (k, k')-nearest neighbor query finds pairs of points, one of each set, such that the point of the first set is a k-nearest neighbor of the point of the second set and, at the same time, the point of the second set is a k'-nearest neighbor of the point of the first set. These queries find applications in collaborative marketing and prospective data analysis, where facilities of one type cooperate with facilities of the other type to obtain reciprocal benefits. We present a sequential and a parallel algorithm, to be run on the CPU and on a Graphics Processing Unit, respectively, for solving multiple bichromatic mutual nearest neighbor queries. We also present the time and space complexity analysis of both algorithms, together with their theoretical comparison. Finally, we provide and discuss experimental results obtained with the implementation of the proposed sequential and a parallel algorithm.

1. Introduction

In this paper we study multiple bichromatic mutual nearest neighbor queries between two sets of facilities of different types. These queries are used in applications where facilities of one type cooperate with facilities of the other type in order to obtain reciprocal benefits.

The two main elements of a mutual nearest neighbor query are facilities and locations. Facilities provide goods or services (hotels, cinemas, museums, manufacturer factories, supplier warehouses) and locations are spatial positions where facilities are located. Throughout the paper we represent locations by points on the Euclidean plane and use interchangeably the terms facility and location of the facility.

We assign a weight to each facility to take into account its importance for the facilities of the other type which would potentially cooperate with it. In practice, experts take available information of the facilities (prestige, magnitude, services, etc.) and then aggregate these factors to obtain their weights [11,12]. To reflect that influence between facilities depends on distance and importance, according to the model presented in [5] which is widely used in location analysis [14,16,35], we associate to each facility a multiplicative weighted Euclidean distance defined by the product of the Euclidean distance and

* Corresponding author.
E-mail addresses: mfort@imae.udg.edu (M. Fort), sellares@ima.udg.edu (J.A. Sellarès).

http://dx.doi.org/10.1016/j.is.2016.07.003
0306-4379/

the inverse of the facility weight. The multiplicative weighted Euclidean distance is not symmetric and does not satisfy the triangle inequality.

Even though the multiplicative weighted Euclidean distance of a facility to the other facilities is an indicator of its potential impact, we also want to know its capacity to influence the other facilities with which it interfaces and how much it can be influenced by them. In fact, the points that have highest influence on a query point are the answer of a k-nearest neighbor query. Given two sets of facilities of different types and one of the points of the first set as a query point, the k-nearest neighbor query retrieves the k-nearest neighbor points to the query point among the points of the second set. Observe that a k-nearest neighbor query is asymmetric because it only considers the proximity of the answer points to the query point, but not the proximity of the query point to the answer points. In fact, the proximity of the answer points to the query point is studied by the reverse k-nearest neighbor queries which retrieve the subset of points that have the query point as one of their k-nearest neighbors. However, the reverse problem is again asymmetric.

This asymmetry is fixed in the (k, k')-mutual nearest neighbors query as we will see next. Given two sets of facilities of different types and a query point of the first set, a bichromatic (k, k')-mutual nearest neighbor query finds the subset of k'-nearest neighbors of the query point in the second set that have the query as one of the k-nearest neighbors in the first set. Mutual nearest neighbor queries model how a facility influences and it is influenced by its neighboring facilities in a natural way. Mutual nearest-neighbor retrieval is symmetric because it considers not only the proximity of the answer points to the query point but also the proximity of the query point to the answer points. Thus, mutual nearest neighbor queries are more suitable, compared with nearest neighbor queries, for those applications involving symmetric nearest neighbor relationships, including data clustering [4,22], outlier detection [4,25], decision making [18,19,41] or pattern recognition [23].

Given two sets of facilities of different types together with a non-empty subset of each such set, a multiple bichromatic mutual (k, k')-nearest neighbor query finds all pairs of points, one on each subset, such that the point of the first subset has the point of the second subset as a k-nearest neighbor in the second set and, at the same time, the point of the second subset has the point of the first subset as a k'-nearest neighbor in the first set. To the best of our knowledge, multiple bichromatic mutual nearest neighbor queries considering multiplicative weighted Euclidean distances have not received attention so far, although they have practical applications as we will see in the next examples.

1.1. Motivational examples

Nowadays, there is an increasing interest in inter-businesses relationships with the objective of building successful collaborative strategies that can provide competitive advantages to the firms involved. This is because it has become difficult for businesses to stay competitive without allying themselves with other businesses. Bearing this in mind, we provide two motivational examples for the problem we tackle in this paper, the first related to collaborative marketing and the second to prospective data analysis.

Collaborative marketing is accomplished by companies working together in cooperative activities, such as product promotion or marketing communication, in order to create synergies and achieve a superior market position for their products and services. Let us see the following example.

Hotels and recreational centers: Consider the set of hotels and the set of cultural and recreational centers (museums, theaters, zoos, thematic parks, etc.) of a given region, together with a subset of each of them, for instance, the subset of hotels of a chain and the subset of those centers that are members of a consortium. An advertisement agency can use the results of a multiple bichromatic mutual nearest neighbor query for the hotels of the chain and the centers of the consortium, to make directly collaborative advertisements with more guarantees of success; e.g., hotels of the chain offer 5% discount for people going to one of the centers of the consortium and these centers offer 10% discount to hotel customers who visit them.

Prospective data analysis could be applied to discover opportunities for cooperation between suppliers and manufacturers in different industries. The solutions of multiple bichromatic mutual nearest neighbor queries, for example, can assist analysts to develop strategic collaborative opportunities:

Manufacturers and suppliers: Vehicle manufacturing businesses, for instance, and auxiliary car companies (electronic, textile, metal, glass) frequently tend to be located in proximity, because there are several advantages related to the proximity of suppliers, for example, it makes the organization of the production easier and more flexible and reduces the transportation costs. When doing a prospective analysis, a multiple bichromatic mutual nearest neighbor query for a subset of auxiliary car companies captures the vehicle manufacturing firms with which each auxiliary company could potentially collaborate.

1.2. Our contributions

In this paper, we propose and motivate multiple bichromatic mutual nearest neighbor queries in the plane considering multiplicative weighted Euclidean distances. Efficiently solving the problem is challenging because multiplicative weighted Euclidean distances do not satisfy the triangle inequality and, consequently, index structures associated with a partition of the plane that heavily rely on this inequality cannot be used to filter out irrelevant points during the search. However, since we can process in parallel many bichromatic mutual nearest neighbor queries and compute many distances between points