The Journal of Systems and Software 84 (2011) 2348–2362

Contents lists available at ScienceDirect

The Journal of Systems and Software

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

P2P-based multidimensional indexing methods: A survey

Chong Zhang*, Weidong Xiao, Daquan Tang, Jiuyang Tang

C4ISR Technology National Defense Science and Technology Key LAB, National University of Defense Technology, Changsha 410073, PR China

ARTICLE INFO

Article history:
Received 26 August 2010
Received in revised form 4 July 2011
Accepted 6 July 2011
Available online 20 July 2011

Keywords:
Multidimensional index
Peer-to-peer computing
Survey

ABSTRACT

P2P-based multidimensional index (MI) is a hotspot which absorbs many researchers to dedicate them into. However, no summarization or review on this technology has been made at present. To the best of our knowledge, this is the first work on reviewing P2P-based MI. This paper innovatively adopts visualization technique to show the research groups and then analyzes investigating style of research groups. Based on evolution of P2P-based MI inheriting from centralized MI and P2P, we divide P2P-based MI methods into 4 categories: extending centralized MI, extending P2P, combining centralized MI and P2P, and miscellaneous. For each category, the paper selects classical techniques and describes them in detail. This is the first time of doing the classification job over massive related works. Finally, load balancing and update strategies are described and discussed for they are important factors related to performance. We believe many researchers will get benefits from our work for further studies.

1. Introduction

Dealing with querying multidimensional data problems has attracted a lot of attention in the database community over the last decades (Gaede and Gunther, 1998; Bertino and Ooi, 1999). Effectively managing multidimensional data is very useful in many fields, such as geosciences, CAD, robotics, and environmental protection. Multidimensional indexing (MI) method is always a key technique to improve the querying efficiency on these data, and also is a challenging issue which interests many researchers to devote themselves into (Bertino et al., 1997; Samet, 1990a,b). Recent years, the researching hotspot on multidimensional indexing has been transferred from centralized paradigm to decentralized one, especially focused on P2P based MI methods (Mondal et al., 2004; Cai et al., 2003; Bharambe et al., 2004; Tanin et al., 2007).

Since the middle period of the last century to nowadays, MI technique has been developed rapidly. Gaede and Gunther (1998) elaborately reviewed MI methods. They summarized and categorized MI's applications, features and technical details. Simply speaking, MI aims to accelerate the querying efficiency on multidimensional data. The basic idea behind MI is to organize the underlying data from a global view, and all the dimensionality should be considered synthetically. Better indexing structures and searching algorithms are preferred to improve the querying procedure. The nature function of MI is pruning–cutting away a lot of useless searching paths.

Peer-to-Peer computing (P2P) (Lua et al., 2005; Milojicic and Kalogeraki, 2002) emerges as a whole new paradigm over the last decades. It emphasizes that participating peers are independent and autonomous, and they cooperatively accomplish the computations and form a self-organized and adaptive sharing community. We analyze the motivation of P2P based MI method as follows (Fig. 1):

(1) Centralized MI methods need to be decentralized to achieve high scalability (Mondal et al., 2004). Because more and more multidimensional data are being used, "single point" MI cannot scale well when the amount of load is highly increased. Bad performance, single-point-failure and bottle neck will cause the whole system running inefficiently. P2P is one of the best solutions to solve the problems brought by centralized mode of MI.
(2) P2P systems need to be equipped with multidimensional complex query processing capabilities (Cai et al., 2003; Bharambe et al., 2004; Tanin et al., 2007). In the area of P2P community, more and more users tend to issue complex queries to find objects which match the requirements on multidimensional perspective. For instances, in P2P multi-player games (Lee et al., 2004), P2P job-search networks (Tanin et al., 2007) or P2P auctions, people usually want to find all other players in a specified area (2-dimesional geometry or 3-dimensional one), all jobs which are suitable for multidimensional requirements, or the most satisfying product of both qualified and cheap. Traditional P2P techniques are able to provide exact-match querying capability but poor mechanism for multidimensional (or multi-

* Corresponding author. Tel.: +86 134 6906 7028.
E-mail addresses: leocheung8286@yahoo.com (C. Zhang), wilsonshaw@vip.sina.com (W. Xiao), dqtang@vip.sina.com (D. Tang), jiuyang_tang@sina.com (J. Tang).

0164-1212/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jss.2011.07.027

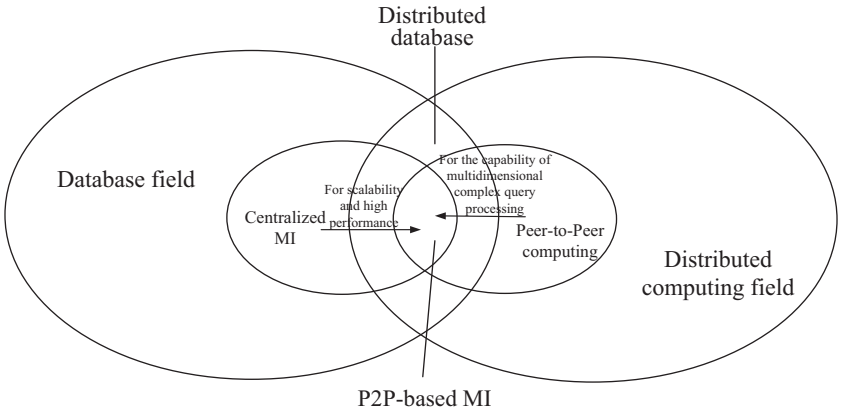

Fig. 1. The motivation of P2P-based MI. Centralized MI and P2P try to develop towards each other.

Table 1
Comparison of centralized MI and P2P-based MI.

	MI in central mode	MI in P2P mode
Node	A section of memory or disk	An autonomous computer (a peer)
Pruning	Reducing I/O times	Reducing communication cost
How to prune	Partitioning the space, designing suitable indexing structure and centralized searching algorithm	Partitioning the space, designing scalable overlay network and distributed searching algorithm

attribute) range search and similarity search. So, P2P should be equipped with MI components.

Although P2P based MI is a kind of distributed computing, the basic principle behind the technique is still pruning. In centralized mode, pruning almost means reducing I/O bound, while in P2P mode, pruning mostly means reducing communication cost. Table 1 shows the comparison of MI in centralized mode and that in P2P mode.

This paper reviews the P2P-based MI methods. To the best of our knowledge, this is the first work to summarize and review the P2P-based MI methods. Besides, in this paper, we apply a new survey to our survey to make readers more intuitionistic to the state-of-art – visualization. The main contribution of this paper is listed as follows:

(1) For the first time, we make a survey of P2P-based MI methods which are new for the database community and P2P community.
(2) We adopt a visualization method in this survey to show the researching groups and analyze the researching style of groups. We believe this is a new concept compared to other field's surveys.
(3) We make a summary of load balancing algorithms and update strategies in the field of P2P-based MI, which are 2 important factors related to system performance. Besides, this will be helpful for designing optimal indexing structures.

The rest of this paper is organized as follow: Section 2 describes querying types in P2P-based MI and makes a classification. In Section 3, we summarize the evolution of P2P-based MI methods. Section 4 gives our new concept – visualizing the researching groups and shows our analysis on researching style. Section 5 summarizes and compares the P2P-based MI methods. From Sections 6–9, we describe every structure in detail by classification. In Section 10, we summarize load balance and update strategies in P2P-based MI methods. We draw a conclusion in Section 11.

2. Query types and classification

2.1. Query types

Gaede and Gunther (1998) listed 9 kinds of multidimensional querying types. We have scanned state-of-the-art for P2P-based MI methods, and the following query types are usually discussed:

(1) Window Query (WQ): Given a d-dimensional interval $I^d = [l_1, u_1] \times [l_2, u_2] \times \cdots \times [l_d, u_d]$, return all objects o intersecting with I^d in all peers: $WQ(I^d) = \{o | I^d \cap o \neq \phi\}$
(2) Range Query (RQ): This kind of query is usually used in similarity search, given a query object q and the range δ, from all peers, return all objects o, from which to q the distance is not more than δ, $RQ(q, \delta) = \{o \,|\, ||o - q|| \leq \delta\}$, here $||.||$ means a similarity distance.
(3) k Nearest Neighbor Query (kNNQ): Given a query object q and the integer k, from all peers, return the k-ary $(o_1, o_2, \ldots, o_k)$, in which o_i is ith nearest neighbor of q:

2.2. Classifications

P2P-based MI methods can be classified by different criteria.

(1) From the view of evolution, P2P-based MI methods can be classified into extending centralized MI, extending P2P, combining centralized MI, P2P, etc. This kind of classification will be explained in Section 3;
(2) From the view of application, P2P-based MI methods can be classified into P2P-based spatial index (especially for GIS), P2P-based multi-attribute index (usually for database), and P2P-based CBIR-oriented index (usually for multimedia applications);
(3) From the view of query types, P2P-based MI can be classified into P2P-based spatial (or multi-attribute) range search index, and P2P-based similarity search (including RQ and kNNQ) index.

3. Deriving and evolutions

P2P-based MI inherits basic idea of traditional central indexing technique or P2P technique or both of them. We illustrate the evolution map of P2P-based MI in Fig. 2 from which some knowledge can be found: