Computers and Electrical Engineering 40 (2014) 350–366

Contents lists available at SciVerse ScienceDirect

Computers and Electrical Engineering

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

# Energy-efficient filtering for skyline queries in cluster-based sensor networks ☆

Bo Yin [a], Yaping Lin [a,*], Jianping Yu [b], Qing Luo [a]

[a] *College of Information Science and Engineering, Hunan University, Hunan, Changsha 410082, China*
[b] *College of Mathematics and Computer Science, Hunan Normal University, Hunan, Changsha 410082, China*

ARTICLE INFO

*Article history:*
Received 19 March 2012
Received in revised form 27 March 2013
Accepted 27 March 2013
Available online 28 April 2013

ABSTRACT

Filtering is a generic technique for skyline retrieval in sensor networks, for the purpose of reducing the communication cost, the dominant part of energy consumption. The vast majority of existing filtering approaches are suitable for uniform and correlated datasets, whereas in many applications the data distribution is clustered or anti-correlated. The only work considering anti-correlated dataset requires significant energy for filtering construction, and it is hard to be efficiently adapted to clustered databases. In this paper, we propose a new filtering algorithm, which settles the problem by utilizing individual node characteristics and generating personalized filters. Given a fraction $k$, a personalized filter prunes at least $k$ percent of points on assigned nodes. A novel scheme for data cluster representation and a sampling method are then proposed to reduce the filtering cost and maximize the benefit of filtering. Extensive simulation results show the superiority of our approach over existing techniques.



## 1. Introduction

Wireless sensor networks (WSNs) have been widely deployed in many monitoring and control systems, where methods and algorithms are developed to solve various query problems, such as range queries, top-$k$ queries, and nearest neighbor queries. With the development of sensing devices and wireless communication technologies, it is becoming urgent to support more complicated queries [1,2]. The powerful capability of retrieving interesting points from a large multi-dimensional data set has made Pareto-optimal (or skyline) queries well suitable for many sensing applications. Given a set of $d$-dimensional points, a skyline query [3] returns the subset of points that are not dominated by any other point. A point $p$ dominates another point $q$, if $p$ is no worse than $q$ in any dimension but better in at least one dimension.

Imagine the following application scenario. To ensure safe working conditions in coal mines, sensors can be deployed to monitor underground conditions including the amount of gas, oxygen, water. Gas leakage could cause explosions if a certain district of gas accumulates to critical explosive density. High oxygen density creates healthy environmental conditions for human beings. Coal mine tunnel surfaces may be corroded and the structural integrity may be threatened if water seepage brings large areas of water into the tunnels. In this case, areas of low oxygen density, high gas density, or high concentration of water (the skyline results) are dangerous zones that need to be carefully searched and examined.

Due to the limited resources of sensor nodes, methods and algorithms designed for WSNs should be of high energy efficiency [4,5]. Since network communication is the main energy consumer, the skyline queries in WSNs raise up a challenge on how to minimize the communication cost. The filtering technique has been exploited to solve this problem, because it can

☆ Reviews processed and approved for publication by Editor-in-Chief by Dr. Manu Malek.

* Corresponding author.
*E-mail addresses:* yinbo@hnu.edu.cn (B. Yin), yplin@hnu.edu.cn (Y. Lin), jianpinghn@163.com (J. Yu), luoqing@hnu.edu.cn (Q. Luo).

0045-7906/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compeleceng.2013.03.021

help identify unqualified points that otherwise would be reported, thus reducing the communication cost over the network [6–10]. Specifically, values, tuples, and grids have been used as filters, and different criterions to measure their filtering capabilities are adopted. For instance, with the assumption that the skyline is computed with respect to the minimum conditions on all dimensions, one [6] chooses the minimum of the maximum values of points (MinMax), while another two measure the quality of a filtering tuple based on its distance to the origin (the shorter the better), which we call here "Min-Distance criterion" [7,8]. In addition to using the filtering value as in [6], one [9] employs another filtering tuple with the maximum volume of dominance region (which we call "Max-Region criterion"). These filtering criterions are also adopted for skyline queries in other distributed environments, such as mobile ad hoc network [11], peer-to-peer systems [12,13], and more generic distributed systems [14–17].

Another important issue that needs to be considered for skyline retrieval is the data distribution. Fig. 1 illustrates four typical distributions: independent, correlated, anti-correlated, and clustered [3,18]. The common feature of the first two datasets is that some tuples lie very close to the origin. These tuples dominate most non-skyline tuples and they are certainly good filter candidates with an obvious filtration effect. All aforementioned criterionts try to pick out such tuples (or correlated values). In the other two datasets, however, there is a low chance that a tuple lies close to the origin. Thus, filters that are computed according to criterions [6–9] only dominate a part of non-skyline tuples with inferior effect. Consider the example in Fig. 2, which shows an anti-correlated data set. The aforementioned criterions advocate filtering tuples *a*, *b*, and *c*, respectively. Clearly, a great number of non-skyline tuples (marked by gray solid circles) can not be pruned by these filtering tuples.

Therefore, in this work, we explore how skyline queries can be computed energy-efficiently for anti-correlated (and clustered) databases in sensor networks. To the best of our knowledge, so far there is only a single piece of work [10] that takes into account the data distributions while devising filters. Specifically, for the anti-correlated dataset, a grid filter is proposed to use a regular grid to capture the coarse-grained data distribution and find out the distributing regions (i.e., the cells in the grid) whose tuples cannot belong to final results. The work of [10], however, needs a great effort to form this regular grid that may have a large size and broadcast it to the entire network. This consumes a significant energy. In fact, the grid filter may be useless or overkill for some nodes, since data space can be different from node to node in practice. Furthermore, it is hard to efficiently adopt this filtering strategy in a clustered database, due to the special characteristics of the clustered database.

Motivated by these observations, we propose a new filtering approach called FSKY. The key idea is to generate personalized filters for individual nodes in the network. Specifically, given a fraction *k*, a personalized filter can prune at least *k* percent of data points on assigned node. This principle is adopted to enhance the overall filtering efficiency and guarantee the benefit of filtering. FSKY essentially allocates different filters to different subsets in the database, thereby adapting to the

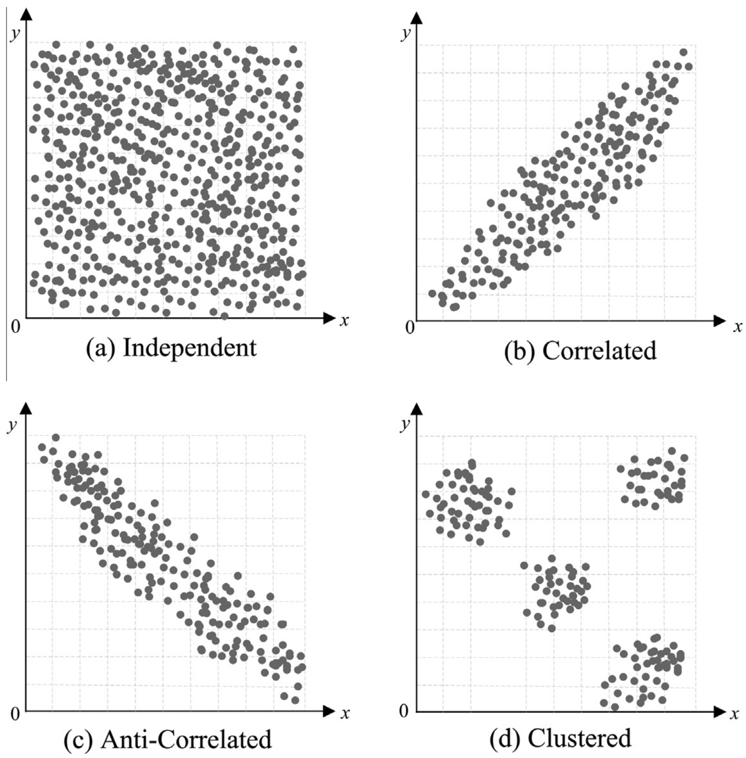


**Fig. 1.** Four common data distributions.