Expert Systems with Applications 41 (2014) 6890–6903

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Dynamic tolerant skyline operation for decision making

Junyi Chai [a,*], Eric W.T. Ngai [a], James N.K. Liu [b]

[a] Department of Management and Marketing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region
[b] Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region

ARTICLE INFO

Article history:
Available online 6 May 2014

Keywords:
Decision making
Skyline
Preference analysis
Personnel selection

ABSTRACT

Skyline operation is typical multicriteria decision making well documented in data engineering. The assumption of skyline operation is settled human preference, which may be subject to huge challenges in practical decision-making applications because it simplifies preference scenarios that are usually dynamic. This study establishes the mathematical formulation of dynamic preference in real settings. A decision approach called tolerant skyline operation (T-skyline) is completely developed, including its conceptual modeling, computation methods, and a skyline maintenance mechanism on a database. The method is established and its computation mechanism is designed, and both are evaluated through an empirical study of personnel selection and evaluation. We also analyze computation efficiency and system stability. The decision targets are fully achieved, the computation results are satisfactory, and the computation efficiency is rational. The effectiveness and advantages of the approach are significant, as illustrated in different real-world settings. Experiments facilitated the examination of the design and development of T-skyline operation by adopting real and public datasets to evaluate players in the National Basketball Association in the United States. The experiment results validate the practical viability of our decision model, which can inspire discussions in sport industries. The methodology used in this study is valuable for further academic research, particularly for the interdisciplinary investigation of decision making and data engineering.

1. Introduction

Multicriteria decision making (MCDM) aims to provide decision makers (DMs) a knowledge recommendation amid a finite number of objects (also known as alternatives, actions, or candidates) evaluated from multiple viewpoints called criteria (also known as dimensions, attributes, or features). MCDM covers four issues: criteria analysis, sorting, ranking, and choice (Figueira, Greco, & Ehrgott, 2005). We are particularly interested in studies of preference MCDM, an issue well documented in various research fields. Representative research topics include *preference learning* in machine learning (e.g., Hüllermeier, Fürnkranz, Cheng, & Brinker, 2008), *preference relations* in cognitive science (e.g., Herrera-Viedma, Herrera, Chiclana, & Luque, 2004; Xu, 2007), *preference query* in data engineering (e.g., Stefanidis, Koutrika, & Pitoura, 2011), and *preference programming or modeling* in decision making (e.g., Greco, Matarazzo, & Slowinski, 2001; Salo & Punkka, 2011). By contrast, interdisciplinary studies of preference MCDM are limited despite the significance of its complementary advantages. In the current study, we examine a data-engineering preference query technique for solving MCDM problems, namely, skyline operation.

1.1. Skyline operation

Preference query (Adomavicius & Tuzhilin, 2005; Stefanidis et al., 2011) aims to retrieve several objects from a database in which all outputted objects must fulfill one or several preferences preset by DMs. This issue has been studied from two aspects: top-*K* operation (Mamoulis, Yiu, Cheng, & Cheung, 2007) and skyline operation (Borzsonyi, Kossmann, & Stocker, 2001). Skyline operation aims to retrieve a set of objects from multidimensional datasets in which all outputted objects conform to the dominance principle: with objects A and B in a multidimensional dataset, A dominates B if A's values are not inferior to B's values in all dimensions and superior to B's values in at least one dimension. Skyline is an elementary set in which objects cannot be dominated by other objects of the dataset. In other words, skyline operation is the acquisition of an object subset in which chosen objects fulfill all preference requirements and no object can dominate a skylining object under defined preferences.

---

* Corresponding author. Tel.: +852 5138 0601.

E-mail addresses: don.chai@polyu.edu.hk, donchaiam@gmail.com (J.Y. Chai), eric.ngai@polyu.edu.hk (E.W.T. Ngai), james.liu@polyu.edu.hk (J.N.K. Liu).

http://dx.doi.org/10.1016/j.eswa.2014.04.041
0957-4174/

We illustrate skyline operation through a simple example of personnel selection (Robertson & Smith, 2001). National Basketball Association (NBA) technical statistics of the 2008–2009 regular season (retrieved from http://espn.go.com/nba/) involving 10 basketball players are shown in Table 1. The dimensions of the dataset include the index, the last name of the players, the affiliated team, and several key technical criteria, such as games played, minutes per game, number of assists (A), number of turnovers (T), and number of steals (S). In the database context, skyline operation accepts two user preferences, GAIN (large values are preferred) and COST (small values are preferred). If the human preference is preset to A and S with the GAIN type and to T with the COST type, the skyline of Table 1 can be calculated according to the above definition as object set {P1, P2, P3, P4, P5, P10}. Thus, the performance of the six skylining players will be recognized because any other object of this dataset cannot be superior to any of them. Meanwhile, the skyline objects are non-comparable with each other and are therefore deemed to have equal performance. The computed skyline may vary with the settings on the preferences. Once it considers only A and S with the GAIN type, for example, the corresponding skyline shall be {P1, P2, P3}.

1.2. Related works

We review literature on skyline operation from two aspects: the concepts and the computation methods. The initial concept of skyline operation was proposed by Borzsonyi et al. (2001). Since this primary study, its conceptual extensions have been comprehensively studied. Representative works include subspace skyline operation (Pei et al., 2006), R-tree-based skyline operation (Papadias, Tao, Fu, & Seeger, 2003), and constrained skyline operation (Lu, Jensen, & Zhang, 2011). Studies since 2010 have tended to the extension of skyline-based applications, such as using skyline operation as an aggregation function to build data cubes for fast online analytical processing (Yiu, Lo, & Yung, 2012) and extending skyline operation processing in peer-to-peer systems (Hose & Vlachou, 2011).

As to the computation methods of skyline, Borzsonyi et al. (2001) pioneered two baseline algorithms, block–nest–loop (BNL) and divide-and-conquer (D&C). BNL compares every object with other objects and identifies its skyline membership if it cannot be dominated. D&C retrieves partial skylines from several subsets of datasets and merges all obtained partial skylines into a final result. Chomicki, Godfrey, Gryz, and Liang (2003) proposed the sort–filter skyline (SFS), which developed BNL by first sorting objects by a monotone function. Godfrey, Shipley, and Gryz (2005) proposed the linear elimination sort for skyline (LESS), which improved SFS by removing a part of objects in the sorting process. Papadias, Tao, Fu, and Seeger (2005) developed progressive skyline computation. Zhang and Chomicki (2011) developed a framework for skyline preference queries in which preferences are set over the whole profile of the dataset instead of just over the dimensions.

Table 1
The running case: an extracted technical statistics of NBA.

No.	LN	TM	G	MPG	A	T	S
P1	Nash	Sun	68	35.3	786	264	56
P2	Williams	Jazz	71	37.5	672	218	78
P3	Kidd	Nets	71	36.9	645	188	117
P4	Paul	Hornets	57	36.5	495	143	109
P5	Davis	Warriors	55	35.5	451	170	114
P6	Ford	Raptors	67	30.4	535	215	93
P7	Miller	76ers	71	36.9	562	200	99
P8	Wade	Heat	46	38.9	362	193	96
P9	Iverson	Nuggets	57	42.7	417	238	112
P10	Billups	Pistons	64	36.7	460	132	78

The development of skyline operation since 2013 has shown wider perspective. Huang, Jiang, Pei, Chen, and Tang (2013) examined the skyline distance to measure the minimum cost of upgrading a querying skyline point to the skyline. Hu, Sheng, Tao, Yang, and Zhou (2013) examined time efficiency in external memory to retrieve the skyline of *N* points in multidimensional space. Trimponias, Bartolini, Papadias, and Yang (2013) studied skyline query processing when a dataset was vertically decomposed into different servers. Zhang, Li, Hassan, Rajasekaran, and Das (2014) investigated a new problem of searching the skyline group.

1.3. Research motivations

Despite being well documented, existing studies are based on data engineering rather than on practical decision making. Preference query processing can be regarded as a classical MCDM problem (Figueira et al., 2005). Specifically, top-*K* operations aim to generate preference-ordered *K* objects and are therefore multicriteria ranking approaches, whereas skyline operations aim to find an object subset that consists of all superior objects and are therefore multicriteria sorting approaches (Chai & Liu, 2014). Studying skyline operation in the MCDM paradigm is important because of the following reasons. In decision-making fields, problems can be addressed on flexible assumptions for practical purposes, but this approach is usually ineffective for many objects (Slowinski, Greco, & Matarazzo, 2009), unless particular decision models (Chai & Liu, 2010; Chai, Liu, & Ngai, 2013) or specialized decision support systems (e.g., Ngai, Peng, Alexander, & Moon, 2014; Ngai et al., 2011) are used. On the other hand, studying skyline operation in data engineering can facilitate the processing of large datasets through database technologies, albeit under restrictive assumptions or conditions. Few studies have conducted an interdisciplinary investigation into skyline operation to fully utilize the complementary advantages of different fields, although Huang et al. (2013) have emphasized the competence of skyline operation in MCDM applications. In the current study, we develop a skyline operation applicable to a database environment but still capable of solving practical decision problems.

Through a conventional mechanism, the outputted size of skyline operation is completely settled with respect to a dataset once query preferences are identified. The operation faces huge challenges in practical decision making because of the following scenarios: (1) the skyline size might be undesirable (e.g., too large or too small) and therefore requires mechanisms to control the outputs, and (2) the preference of DMs might be imperfect and therefore requires approaches for meeting dynamic predefined settings and adjustable outputs. Several studies have resolved related issues. Some studies (Lin, Yuan, Zhang, & Zhang, 2007; Lu et al., 2011; Tao, Ding, Lin, & Pei, 2009) have provided methods to find a representative subset of skylines. These methods can control the skyline size and remain valid when this size is larger than desired. For an outputted skyline that is too little to fulfill the DM's needs, Chai, Liu, and Li (2013) designed mechanisms that accept marginal objects hierarchically. However, these mechanisms still failed to adjust the skyline along with dynamic settings of preference. Wong, Pei, Fu, and Wang (2009) argued that the most straightforward solution is to transfer dynamic skylines to conventional skylines, thus making existing skyline operations feasible. However, this solution requires the full materialization of datasets and time-consuming preprocessing and is thus prohibitive, providing a semi-materialization method via consideration of an implicit preference rather than of ordered values. Jiang, Pei, Lin, Cheung, and Han (2008) studied preference relations in a specific problem domain. Yiu, Lu, Mamoulis, and Vaitis (2011) provided preference query techniques for a spatial database. These studies partly refer to the dynamic preference of skyline