

## Accepted Manuscript

Top k probabilistic skyline queries on uncertain data

Zhibang Yang, Kenli Li, Xu Zhou, Jing Mei, Yunjun Gao

PII: S0925-2312(18)30440-5  
DOI: <https://doi.org/10.1016/j.neucom.2018.03.052>  
Reference: NEUCOM 19473

To appear in: *Neurocomputing*

Received date: 13 August 2017  
Revised date: 7 February 2018  
Accepted date: 19 March 2018

Please cite this article as: Zhibang Yang, Kenli Li, Xu Zhou, Jing Mei, Yunjun Gao, Top k probabilistic skyline queries on uncertain data, *Neurocomputing* (2018), doi: <https://doi.org/10.1016/j.neucom.2018.03.052>



This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Top $k$ probabilistic skyline queries on uncertain data

Zhibang Yang <sup>1</sup>, Kenli Li <sup>2</sup>, Xu Zhou <sup>2</sup>, Jing Mei <sup>3</sup>, Yunjun Gao <sup>4</sup>

<sup>1</sup> College of Computer Engineering and Applied Mathematics, Changsha University, Hunan 410003, China

<sup>2</sup> College of Information Science and Engineering, Hunan University, Hunan 410082, China

<sup>3</sup> College of Mathematics and Computer Science, Hunan Normal University, Changsha, Hunan 410082, China

<sup>4</sup> College of Computer Science, Zhejiang University, Hangzhou 310027, China

---

## Abstract

Uncertainty of data is inherent in many applications, and query processing over uncertain data has gained widespread attention. The probabilistic skyline query is a powerful tool for managing uncertain data. However, the famous probabilistic skyline query, called  $p$ -skyline query, is likely to return unattractive objects which have no advantage in either their attributes or skyline probabilities with comparing to other query results. Moreover, it may return too many objects to offer any meaningful insight for customers. In this paper, we first propose a modified  $p$ -skyline (MPS) query based on a strong dominance operator to identify truly attractive results. Then we formulate a top  $k$  MPS (TkMPS) query on the basis of a new ranking criterion. We present effective approaches for processing the MPS query, and extend these approaches to process the TkMPS query. To improve the query performance, the reuse technique is adopted. Extensive experiments verify that the proposed algorithms for the MPS and TkMPS queries are efficient and effective, our MPS query can filter out 34.44% unattractive objects from the  $p$ -skyline query results at most, and although in some cases the results of the MPS and the  $p$ -skyline queries are just the same, our MPS query needs much less CPU, I/O, and memory costs.

*Keywords:* Skyline query, Query processing, Uncertain data.

---

## 1. Introduction

The skyline operator was first introduced to the database community in [1]. Thereafter, it has drawn great attention due to it is widely used in applications related to multi-criteria decision making [2, 3]. Given a dataset, the skyline

---

*Email addresses:* yangzhibang2006@126.com (Zhibang Yang <sup>1</sup>), lk1@hnu.edu.cn (Kenli Li <sup>2</sup>), zhouxu2006@126.com (Xu Zhou <sup>2</sup>), jingmei1988@163.com (Jing Mei <sup>3</sup>), gaoyj@zju.edu.cn (Yunjun Gao <sup>4</sup>)

Download English Version:

<https://daneshyari.com/en/article/10151107>

Download Persian Version:

<https://daneshyari.com/article/10151107>

[Daneshyari.com](https://daneshyari.com)