

Title	Research on Continuous Preference Query Processing in Distributed Environments
Author(s)	Udomlamlert, Kamalas
Citation	大阪大学, 2017, 博士論文
Version Type	VoR
URL	<a href="https://doi.org/10.18910/61865">https://doi.org/10.18910/61865</a>
rights	
Note	

*Osaka University Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

Osaka University

## 論文内容の要旨

氏名 ( UDOMLAMLERT KAMALAS )	
論文題名	Research on Continuous Preference Query Processing in Distributed Environments (分散環境における連続型の嗜好問合せ処理に関する研究)
論文内容の要旨	
<p>We are witnessing the era that many things around us (for example, smartphones, watches, cars and cameras) start generating data likewise sensors in traditional sensor network, and especially these are mostly generated in a distributed manner. Many of instruments continuously create data in real-time and become vast streams of data. Monitoring on these streams of data enables many promising and useful applications such as environmental monitoring, smart cities and industrial monitoring. However, it is no use exploring everything in a huge flood of data. On the contrary, end-users are more interested in and want to get informed only a few of most important data that match to their preference (objectives). Therefore, continuous preference query processing does the job to return only a small number of potential data from immense data space. Although, the straightforward implementations of continuous query processing can be achieved by repeatedly executing snapshot query processing of such queries which have been extensively studied, most of them do not perfectly fit because they cannot reduce the critical costs that matter to distributed environments.</p> <p>In this thesis, we focus on monitoring of two fundamental preference queries including top-k and skyline queries. We begin with the proposal of subscription-based top-k query processing in 2-tier distributed systems, as often seen in traditional sensor networks and peer-to-peer systems (P2Ps). Even though there are existing solutions for this problem, they fail to efficiently reduce the number of sent data or yield too much overhead to support users in a large scale. Our proposed method not only keeps the number of sent data low by using subscriptions but also carefully minimizes the number of subscriptions to be maintained. As a result, the total communication cost, which is a severe factor in sensor networks, can be largely reduced. The results through simulation experiments confirm the advantages of our proposed method over the existing methods.</p> <p>In skyline processing, a single data update may totally change the skyline set (result set); therefore, for monitoring purposes, it needs to be re-computed periodically to detect changes of the skyline set. However, skyline query processing is a computationally intensive task. Therefore, in the second part of the thesis, we propose an efficient algorithm to process skyline monitoring queries on frequent updates. Minimum bounding rectangles (MBRs) are used to summarize data movement in each snapshot. This approach enables us to identify a smaller set of candidates for skyline computation. Through the experiments, our proposed method shows the better results in terms of total execution time (computation cost) compared with other methods.</p> <p>Regarding scalability, the way we manage, collect and utilize data recently has been altered to cloud (a specialized form of distributed computing). Everything in the cloud has been abstracted and provided as convenient services on a pay-as-you-go basis, and data access is not an exception. Running continuous preference queries that need to access a lot of data for comparisons in such a system is high-priced (expense or monetary cost). In the last part of the thesis, we describe cost-minimizing methods for monitoring queries in cloud (i.e., sensor cloud) where the cost is denoted by the expense of data access. Instead of requesting all the latest data in each timestamp, we present a cost-minimizing framework for top-k monitoring - a novel e-top-k query processing delivering approximate top-k answers with a probabilistic guarantee. The extensive experiments on the real-world datasets demonstrate that our approach can reduce the expense by more than half with desirable accuracy. In addition, we show that the underlying concepts of this method can be easily extended to be used in skyline monitoring, which also gives the same promising results as that of top-k monitoring.</p>	

## 論文審査の結果の要旨及び担当者

氏 名 ( UDOMLAMLERT KAMALAS )		
	(職)	氏 名
論文審査担当者	主 査	教授 原 隆浩
	副 査	教授 鬼塚 真
	副 査	教授 松下 康之
	副 査	教授 藤原 融
	副 査	教授 下條 真司
<b>論文審査の結果の要旨</b>		
<p>近年、スマートフォン、時計、車、カメラといった身の回りの多くのものが従来のセンサネットワークのようにセンシングデータを、分散的、継続的、および大量に生成し始めている。このようなストリーム環境におけるデータのモニタリングは、環境モニタリングやスマートシティなどへの応用の基盤技術である。また、生成される全てのデータを処理することは非実用的であるため、ユーザの嗜好・目的にマッチするデータのみを検索・モニタリングする技術が求められる。本論文では、嗜好問合せとしてTop-k検索とスカイライン検索に取り組んでおり、通信量の削減、計算の高速化、およびクラウド環境におけるデータ取得コストの削減を目標としている。本論文の主要な研究成果を要約すると次の通りである。</p>		
<p>(1) 2階層型分散システムにおけるサブスクリプションに基づくTop-k検索処理アルゴリズムを提案している。本問題には幾つかの既存研究が存在するが、送信データの効率的な削減について大きな改善の余地があり、大量のユーザをサポートするような大規模システムへの拡張が可能でない。提案アルゴリズムでは、ユーザが指定したクエリの集合およびTop-k検索の結果から、今後生成されるデータがTop-kデータになり得る領域を求め、その結果を幾つかのサブスクリプションとしてデータ送信のフィルタリングに利用する。これにより、データの送信の数を少なく抑えており、Top-kデータモニタリングに必要な通信量を劇的に削減している。シミュレーション実験の結果から、既存アルゴリズムに対する優位性を示している。</p>		
<p>(2) スカイライン検索では、たった1つのデータの更新が検索結果を劇的に変える可能性がある。そのため、データの更新が頻繁に起こる環境では、スカイラインとなるデータをリアルタイムにモニタリングするための効率的なアルゴリズムが必要となる。提案アルゴリズムでは、データの更新を最小矩形領域 (Minimum bounding rectangle, MBR) を用いて要約し、データの更新が起きた場合に計算が必要となるデータの数を削減することで計算時間を短縮している。</p>		
<p>(3) クラウドシステムの普及により、全てのデータをクラウド上で管理し、要求されたデータだけを提供する pay-as-you-go サービスが浸透している。データの更新が起こる環境では、一旦取得したデータは時間経過に伴って更新される。そのため、検索結果の更新の確認のためにデータを再取得しなければならないが、に支払いコストが非常に大きくなってしまふ。この問題を解決するため、データアクセス回数 (つまり支払いコスト) を削減するアルゴリズムを提案している。提案アルゴリズムは、正規分布を想定し、データの更新を推定することで、ユーザが望む近似率を確率的に保証する近似解を提供する。実データを用いた実験から、各タイムスタンプにおける最新のデータを全て取得するアプローチに対して、データアクセス回数を半分以上に削減していることを確認している。</p>		
<p>以上のように、本論文は分散環境における連続型の嗜好問合せ処理に関する先駆的な研究として、情報科学に寄与するところが大きい。よって本論文は博士 (情報科学) の学位論文として価値のあるものと認める。</p>		