ISSN 2321-2926

Probabilistic Top-k Queries in Wireless Sensor Networks Sharing System

Mr. L. Sankararao[1] and Mr. N.K. Sumanth[2]

1. *Final Year M.Tech, Department of CSE, Baba Institute of Technology and Sciences, Visakhapatnam, AP, India.*
2. *Assistant Professor Department of CSE, Baba Institute of Technology and Sciences, Visakhapatnam, AP, India.*

Abstract: With the reference of the previous wireless sensor Network consists of spatially dispersed self-directed sensors to monitor the environment conditions. In probabilistic Top-k, a new approach such as Expected ranking method is used to get accurate result in finding Top-k results and as well as accurate probability. For inter cluster processing three algorithm is used namely Sufficient-set based, Necessary-set based and Boundary based algorithm. For reducing transmission cost adaptive algorithm is used and this gives the efficient result in the bounded rounds of communication. This gives least transmission cost and not exceeds two rounds of communication. In this paper, we propose the notion of sufficient set for distributed processing of probabilistic Top-k queries in cluster-based wireless sensor networks. Through the derivation of sufficient boundary, we show that data items ranked lower than sufficient boundary are not required for answering the probabilistic top-k queries, thus are subject to local pruning. Accordingly, we develop the sufficient set-based (SSB) algorithm for inter-cluster query processing. Experimental results show that the proposed algorithm reduces data transmissions significantly by finding the process sharing and method development.

Introduction:

A number of applications today need to manage data that is imprecise. For example imprecisions arise in fuzzy object matching across multiple databases, in data extracted automatically from unstructured text, in automatic schema alignments, in sensor data, in activity recognition data. In some cases it is possible to eliminate the imprecisions completely, but this is usually very costly, like manual removal of ambiguous matches in data cleaning; in other cases complete removal is not even possible. A recent approach to manage imprecisions is with a probabilistic database, which uses probabilities to represent the uncertainty about the data [5, 6, 7, 8, 21]. A simplistic definition is that every tuple belongs to the database with some probability, whose value is between 0 and 1, and, as a consequence, every tuple returned by a SQL query will have some probability, reflecting the systems' confidence in the answer. AMZN Reviews (asin, title, customer, rating, ...) AMZN Director (asin, director) AMZNActor(asin, actor) IMDB Movie (mid, movie Title, genre, did, year) IMDB Director (did, dirName) IMDB Cast (mid, aid) IMDB Actor (aid, actor Name) Title Matchp (asin, mid, p), Fragment of IMDB and Amazon schemas A major challenge in probabilistic databases is query evaluation. Dalvi and Suciu [6] have shown recently that most SQL queries are #P-complete, which rules out efficient evaluation algorithms. Previous approaches to query evaluation on probabilistic databases have either restricted the queries [2, 5, 8], or modified the semantics [17], or were not scalable [10]. In this paper we propose a new approach to query evaluation on probabilistic databases, by combining top-k style queries with approximation algorithms with provable guarantees. More precisely, we compute and rank the top k answers (in order of their probabilities) of a SQL query. We guarantee the correct ranking of the top k answers, but only approximate their probabilities to the extent needed to compute their ranking. Thus, the users specifies a SQL query and a number k, and the system returns the highest rankedk answers, which are guaranteed to be correct: the probabilities of these answers are reported too, but they may be approximate. When managing imprecisions in data the most meaningful information lies not in the exact values of the output probabilities but in the ranking of the queries' answers. Thus, we shift the focus from probabilities to ranks, and give a new, provably optimal algorithm for computing the top k answers.

A wireless sensor network (WSN) (sometimes called a wireless sensor and actor network (WSAN)[1]) of spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, pressure, etc. and to cooperatively pass their data through the network to a main location. The more modern networks are bi-directional, also enabling control of sensor activity. The development of wireless sensor networks was motivated by military applications such as battlefield surveillance; today such networks are used in many industrial and consumer applications, such as industrial process monitoring and control, machine health monitoring, and so on. The WSN is built of "nodes" – from a few to several hundreds or even thousands, where each node is connected to one (or sometimes several) sensors. Each such sensor network node has typically several parts: a radio transceiver with an internal antenna or connection to an external antenna, a microcontroller, an electronic circuit for interfacing with the sensors and an energy source, usually a battery or an embedded form of energy harvesting. A sensor node might vary in size from that of a shoebox down to the size of a grain of dust, although functioning "motes" of genuine microscopic dimensions have yet to be created. The cost of sensor nodes is similarly variable, ranging from a few to hundreds of dollars, depending on the complexity of the individual sensor nodes. Size and cost constraints on sensor nodes result in corresponding constraints on resources such as energy, memory, computational speed and communications bandwidth. The topology of the WSNs can vary from a simple star network to an advanced multi-hop wireless mesh network. The propagation technique between the hops of the network can be routing or flooding

SQL Queries:

The query i retrieves all directors that produced a lowly rated comedy and a highly rated drama less than five years apart. There are 1415 answers to this query (in decreasing order of their probabilities). Consider one such answer, say Stanley Donen. The system computed its probability 0.88 by considering all low ranking reviews that may match one of his comedies, and all high ranking reviews that may match one of his dramas, and accounting for the probabilities of these matches: the exact semantics is based on possible worlds, and is reviewed. Note that, unlike the threshold approach, a poor match between a movie by Donenand a review is not automatically discarded, butkept and used during query processing: its low probability however is taken into account when ranking Donen in the query's answer

Contemplate Process:

There are three proposed algorithms to minimize the transmission cost. We show the applicability of sufficient set and necessary set to wireless sensor networks with both two-tier hierarchical and tree-structured network topologies. There are several top-k query semantics and solutions proposed recently, including U-Topk and UkRanks in PT-Topk in PK-Topk in expected rank in and so on. A common way to process probabilistic top-k queries is to first sort all tuples based on the scoring attribute, and then process tuples in the sorted order to compute the final answer set. Nevertheless, while focusing on optimizing the transmission bandwidth, the proposed techniques require numerous iterations of computation and communication, introducing tremendous communication overhead and resulting in long latency. As argued in this is not desirable for many distributed applications, e.g., network monitoring, that require the queries to be answered in a good response time, with a minimized energy consumption. In this paper, we aim at developing energy efficient algorithms optimized for fixed rounds of communications.

Communicating channel:

Embodiments contemplate methods and systems for determining and communicating channel state information (CSI) for one or more transmission points (or CSI reference signal resources). Embodiments further contemplate determining transmission states may include applying at least one CSI process for channel state information (CSI) reporting. Embodiments also contemplate aperiodic and/or periodic reporting of one or more report types (e.g., rank indicator (RI)) of CSI, perhaps based on one or more reporting modes that may be configured for each of the one or more CSI process. A detailed description of illustrative embodiments will now be described with reference to the various Figures. Although this description provides a detailed example of possible implementations, it should be noted that the details are intended to be exemplary and in no way limit the scope of the application. As used herein, the article "a", absent further qualification or characterization, may be understood to mean "one or more" or "at least one", for example.Embodiments contemplate methods and systems for communicating transmission states or CSI processes. For example, a method for determining transmission states or CSI processes may include applying at least one transmission state or CSI process parameter to channel state information (CSI). The method may also include reporting CSI based on a transmission state or CSI process and at least one transmission state or CSI process parameter applied thereto, and applying a correction factor to the at least one transmission state or CSI process.

Detailed Description:

The communications system 100 may be a multiple access system that provides content, such as voice, data, video, messaging, broadcast, etc., to multiple wireless users. The communications system 100 may enable multiple wireless users to access such content through the sharing of system resources, including wireless bandwidth. For example, the communications systems 100 may employ one or more channel access methods, such as code division multiple access (CDMA), time division multiple access (TDMA), frequency division multiple access (FDMA), orthogonal FDMA (OFDMA), single- carrier FDMA (SC-FDMA), and the like. As shown in Figure 1A, the communications system 100 may include wireless transmit/receive units (WTRUs) 102a, 102b, 102c, 102d, a radio access network (RAN) 104, a core network 106, a public switched telephone network (PSTN) 108, the Internet 1 10, and other networks 1 12, though it will be appreciated that the disclosed embodiments contemplate any number of WTRUs, base stations, networks, and/or network elements. Each of the WTRUs 102a, 102b, 102c, 102d may be any type of device configured to operate and/or communicate in a wireless environment. By way of example, the WTRUs 102a, 102b, 102c, 102d may be configured to transmit and/or receive wireless signals and may include user equipment (UE), a mobile station, a fixed or mobile subscriber unit, a pager, a cellular telephone, a personal digital assistant (PDA), a smartphone, a laptop, a netbook, a personal computer, a wireless sensor, consumer electronics

Uses:

- Additionally, NSB and BB take advantage of the skewed necessary sets and necessary boundaries among local clusters to obtain their global boundaries, respectively, which are very effective for intercluster pruning.
- The transmission cost increases for all algorithms because the number of tuples needed for query processing is increased.

Wireless Sensor Networks Sharing System:

In this thesis we present and operating system and three generations of a hardware platform designed to address the needs of wireless sensor networks. Our operating system, called TinyOS uses an event based execution model to provide support for fine-grained concurrency and incorporates a highly efficient component model. TinyOS enables us to use a hardware architecture that has a single processor time shared between both application and protocol processing. We show how a virtual partitioning of computational resources not only leads to efficient resource utilization but allows for a rich interface between application and protocol processing. This rich interface, in turn, allows developers to exploit application specific communication protocols that significantly improve system performance.

The hardware platforms we develop are used to validate a generalized architecture that is technology independent. Our general architecture contains a single central controller that performs both application and protocol-level processing. For flexibility, this controller is directly connected to the RF transceiver. For efficiency, the controller is supported by a collection of hardware accelerators that provide basic communication primitives that can be flexibility composed into application specific protocols.

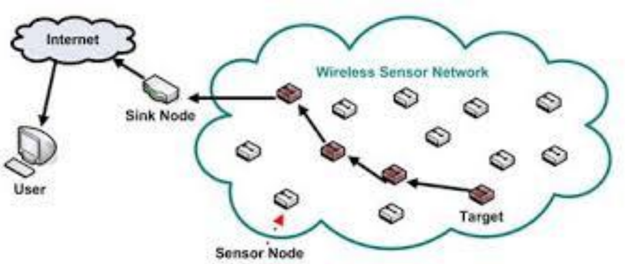
The three hardware platforms we present are instances of this general architecture with varying degrees of hardware sophistication. The Rene platform serves as a baseline and does not contain any hardware accelerators. It allows us to develop the TinyOS operating system concepts and refine its concurrency mechanisms. The Mica node incorporates hardware accelerators that improve communication rates and synchronization accuracy within the constraints of current microcontrollers. As an approximation of our general architecture, we use Mica to validate the underlying architectural principles. The Mica platform has become the foundation for hundreds of wireless sensor network research efforts around the world. It has been sold to more than 250 organizations.

Spec is the most advanced node presented and represents the full realization of our general architecture. It is a 2.5 mm x 2.5 mm CMOS chip that includes processing, storage, wireless communications and hardware accelerators. We show how the careful selection of the correct accelerators can lead to orders-of-magnitude improvements in efficiency without sacrificing flexibility. In addition to performing a theoretical analysis on the strengths of our architecture, we demonstrate its capabilities through a collection of real-world application deployments.

Top-k Queries:

Due to the existence of uncertain data in a wide spectrum of real applications, uncertain query processing has become increasingly important, which dramatically differs from handling certain data in a traditional database. In this paper, we formulate and tackle an important query, namely probabilistic top-k dominating (PTD) query, in the uncertain database. In particular, a PTD query retrieves k uncertain objects that are expected to dynamically dominate the largest number of uncertain objects. We propose an effective pruning approach to reduce the PTD search space, and present an efficient query procedure to answer PTD queries. Furthermore, approximate PTD query processing and the case where the PTD query is issued from an uncertain query object are also discussed. Extensive experiments have demonstrated the efficiency and effectiveness of our proposed PTD query processing approaches.

In proposed approach, expected ranking is estimated after aggregate probability is calculated. Then sufficient set and necessary set is calculated to obtain Top-k result. Then the three algorithms are used to send those set to base station for finding Top-k. A. Sensor Network A wireless sensor network that consists of a large number of sensor nodes deployed in a geographical region. Feature readings (e.g., moisture levels or speed of wind gust) are collected from these distributed sensor nodes. In this network, sensor nodes are grouped into clusters, within each of which one of sensors is selected as the cluster head for performing localized data processing. By using statistic methods a cluster head may generate a set of data tuples for each zone within its monitored region. Each tuple is comprised of tupleid, zone, a derived possible attribute value, along with a confidence that serves as a measurement of data uncertainty. ThePT-Topk queries in a centralized uncertain database, which provides a good background for the targeted distributed processing problem. The query answer can be obtained by examining the tuples in descending ranking order from the sorted table (which is still denoted as T for simplicity). We can easily determine that the highest ranked k tuples are definitely in the answer set as long as their confidences are greater than p since their qualifications as PT-Topk answers are not dependent on the existence of any other tuples



Modules:

1. PT-Topk Query Processing
2. Sensor Networks
3. Data pruning
4. Structured network topology
5. Data transmission
6. Performance evaluation

PT-Topk Query Processing:

The notion of sufficient set and necessary set for distributed processing of probabilistic top-k queries in cluster-based wireless sensor networks. These two concepts have very nice properties that can facilitate localized data pruning in clusters. Accordingly, we develop a suite of algorithms, namely, sufficient set-based (SSB), necessary set-based (NSB), and boundary-based (BB), for intercluster query processing with bounded rounds of communications. Moreover, in responding to dynamic changes of data distribution in the network, we develop an adaptive algorithm that dynamically switches among the three proposed algorithms to minimize the transmission cost. We show the applicability of sufficient set and necessary set to wireless sensor networks with both two-tier hierarchical and tree-structured network topologies. Experimental results show that the proposed algorithms reduce data transmissions significantly and incur only small constant rounds of data communications. The experimental results also demonstrate the superiority of the adaptive algorithm, which achieves a near-optimal performance under various conditions.

Sensor Networks

The extensive number of research work in this area has appeared in the literature. Due to the limited energy budget available at sensor nodes, the primary issue is how to develop energy-efficient techniques to reduce communication and energy costs in the networks. Approximate-based data aggregation techniques have also been proposed. The idea is to tradeoff some data quality for improved energy efficiency. Silberstein et al. develop a sampling-based approach to evaluate approximate top-k queries in wireless sensor networks. Based on statistical modeling techniques, a model-driven approach was proposed in to balance the confidence of the query answer against the communication cost in the network. Moreover, continuous top-k queries for sensor networks have been studied in and . In addition, a distributed threshold join algorithm has been developed for top-k queries. These studies, considering no uncertain data, have a different focus from our study.

Data pruning:

The cluster heads are responsible for generating uncertain data tuples from the collected raw sensor readings within their clusters. To answer a query, it's natural for the cluster heads to prune redundant uncertain data tuples before delivery to the base station in order to reduce communication and energy cost. The key issue here is how to derive a compact set of tuples essential for the base station to answer the probabilistic top-k queries.

Structured Network Topology:

To perform in-network query processing, a routing tree is often formed among sensor nodes and the base station. A query is issued at the root of the routing tree and propagated along the tree to all sensor nodes. Although the concepts of sufficient set and necessary set introduced earlier are based on two-tier hierarchical sensor networks, they are applicable to tree-structured sensor network.

Data transmission:

The total amount of data transmission as the performance metrics. Notice that, response time is another important metrics to evaluate query processing algorithms in wireless sensor networks. All of those three algorithms, i.e., SSB, NSB, and BB, perform at most two rounds of message exchange there is not much difference among SSB, NSB, and BB in terms of query response time, thus we focus on the data transmission costin the evaluation. Finally, we also conduct experiments to evaluate algorithms, SSB-T, NSB-T, and NSB-T-Opt under the tree-structured network topology.

Performance evaluation:

The performance evaluation on the distributed algorithms for processing PT-top k queries in two-tier hierarchical cluster based wireless sensor monitoring system. As discussed, limited energy budget is a critical issue for wireless sensor network and radio transmission is the most dominate source of energy consumption. Thus, we measure the total amount of data transmission as the performance metrics. Notice that, response time is another important metrics to evaluate query processing algorithms in wireless sensor networks.

System Artech Center: We describe the motivation and design of a novel embedded systems architecture for large networks of small devices, tha canonical example being wireless sensor networks. The architecture differs from previous work in being based explicitly on a hardware/software co-design approach centred around the deployment of novel programming language constructs directly onto hardware in order to improve optimisation and expressibility. The programming interface enables the dynamic download and execution of domain-specific code to facilitate the development of context aware pervasive computing systems whose behaviour must adapt to their changing environment. To this end, the architecture implements a virtual machine operating environment based on Scheme andμClinux that encapsulates a CPU core, digital logic, generic I/O, network interfaces and domain-specific programming language composition.

References:

[1] Distributed Processing of Probabilistic Top-kQueries in Wireless Sensor Networks Mao Ye, Wang-Chien Lee, Member, IEEE, Dik Lun Lee, Member, IEEE, and Xingjie Liu, Student Member, IEEE
http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6175892

[2] K. Yi, F. Li, G. Kollios, and D. Srivastava, "Efficient Processing of Top-k Queries in Uncertain Databases with X-Relations," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 12, pp. 1669-1682, Dec. 2008.

[3] J. Li, B. Saha, and A. Deshpande, "A Unified Approach to Ranking in Probabilistic Databases," Proc. Int'l Conf. Very Large Data Bases (VLDB), vol. 2, no. 1, pp. 502-513, 2009.

[4] E. Elnahrawy and B. Nath, "Poster Abstract: Online Data Cleaning in Wireless Sensor Networks," Proc. First Int'l Conf. Embedded Networked Sensor Systems (SenSys '03), pp. 294-295, 2003.

[5] A. Deshpande, C. Guestrin, S.R. Madden, J.M. Hellerstein, and W. Hong, "Model-Driven Data Acquisition in Sensor Networks," Proc. 13th Int'l Conf. Very Large Data Bases (VLDB '04), pp. 588-599, 2004.

[6] R. Cheng, D.V. Kalashnikov, and S. Prabhakar, "Evaluating Probabilistic Queries over Imprecise Data," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '03), pp. 551-562, 2003.

[7] S. Abiteboul, P. Kanellakis, and G. Grahne, "On the Representation and Querying of Sets of Possible Worlds," Proc. ACM SIGMOD Int'l Conf.

Management of Data (SIGMOD '87), pp. 34-48, 1987.

[8] N. Dalvi and D. Suciu, "Efficient Query Evaluation on Probabilistic Databases," Proc. 30th Int'l Conf. Very Large Data Bases (VLDB '04), pp. 864-875, 2004.

[9] A.D. Sarma, O. Benjelloun, A. Halevy, and J. Widom, "Working Models for Uncertain Data," Proc. 22nd Int'l Conf. Data Eng. (ICDE '06), p. 7, 2006.

[10] R. Cheng, Y. Xia, S. Prabhakar, R. Shah, and J.S. Vitter, "Efficient Indexing Methods for Probabilistic Threshold Queries over Uncertain Data," Proc. 30th Int'l Conf. Very Large Data Bases (VLDB '04), pp. 876-887, 2004.

[11] Y. Tao, R. Cheng, X. Xiao, W.K. Ngai, B. Kao, and S. Prabhakar, "Indexing Multi-Dimensional Uncertain Data with Arbitrary Probability Density Functions," Proc. 31st Int'l Conf. Very Large Data Bases (VLDB '05), pp. 922-933, 2005.

[12] C. Re, N. Dalvi, and D. Suciu, "Efficient Top-k Query Evaluation on Probabilistic Data," Proc. Int'l Conf. Data Eng. (ICDE '07), pp. 896-905, 2007.

[13] M. Hua, J. Pei, W. Zhang, and X. Lin, "Ranking Queries on Uncertain Data: A Probabilistic Threshold Approach," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '08), 2008.

[14] F. Li, K. Yi, and J. Jestes, "Ranking Distributed Probabilistic Data," Proc. 35th SIGMOD Int'l Conf. Management of Data (SIGMOD '09), 2009.

[15] H.W. Rabiner, C. Anantha, and B. Hari, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," Proc. 33rd Hawaii Int'l Conf. System Sciences (HICSS '00), 2000.

[16] Y. Diao, D. Ganesan, G. Mathur, and P.J. Shenoy, "Rethinking Data Management for Storage-Centric Sensor Networks," Proc. Conf. Innovative Data Systems Research (CIDR '07), pp. 22-31, 2007.

[17] M.A. Soliman, I.F. Ilyas, and K.C. Chang, "Top-k Query Processing in Uncertain Databases," Proc. Int'l Conf. Data Eng. (ICDE '07), 2007.

[18] C. Jin, K. Yi, L. Chen, J.X. Yu, and X. Lin, "Sliding-Window Top-k Queries on Uncertain Streams," Proc. Int'l Conf. Very Large Data Bases (VLDB '08), 2008.

[19] G. Cormode, F. Li, and K. Yi, "Semantics of Ranking Queries for Probabilistic Data and Expected Ranks," Proc. IEEE Int'l Conf. Data Eng. (ICDE '09), 2009.

[20] P. Cao and Z. Wang, "Efficient Top-k Query Calculation in Distributed Networks," Proc. 23rd Ann. ACM Symp.Principles of Distributed Computing (PODC), pp. 206-215, 2004.

[21] S. Madden, M.J. Franklin, J. Hellerstein, and W. Hong, "TAG: A Tiny AGgregation Service for Ad-Hoc Sensor Networks," Proc. Fifth Symp. Operating Systems Design and Implementation (OSDI '02), 2002.

[22] M. Ye, X. Liu, W.-C. Lee, and D.L. Lee, "Probabilistic Top-k Query Processing in Distributed Sensor Networks," Proc. IEEE Int'l Conf. Data Eng. (ICDE '10), 2010.