International Journal for Innovative Engineering and Management Research

A Peer Revieved Open Access International Journal

www.ijiemr.org

IJIEMR Transactions, online available on 20th July 2017. Link :

http://www.ijiemr.org/downloads.php?vol=Volume-6&issue=ISSUE-5

Title: Implementation of Application - Level Semantics In Data Compression.

Volume 06, Issue 05, Page No: 1945 – 1950.

Paper Authors

***T.NEEHARIKA, J.SARITHA.**

* Dept of CSE, CVRT.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IMPLEMENTATION OF APPLICATION - LEVEL SEMANTICS IN DATA COMPRESSION

***T.NEEHARIKA,**J.SARITHA**
*PG Scholar, Dept of CSE, CVRT, AP, India
**Assistant Professor,Dept of CSE, CVRT, AP, India

ABSTRACT

Natural phenomena show that many creatures form large social groups and move in regular patterns. However, previous works focus on finding the movement patterns of each single object or all objects. In this paper, we first propose an efficient distributed mining algorithm to jointly identify a group of moving objects and discover their movement patterns in wireless sensor networks. Afterward, we propose a compression algorithm, called 2P2D, which exploits the obtained group movement patterns to reduce the amount of delivered data.The compression algorithm includes a sequence merge and an entropy reduction phases. In the sequence merge phase, we propose a Merge algorithm to merge and compress the location data of a group of moving objects. In the entropy reduction phase, we formulate a Hit Item Replacement (HIR) problem and propose a Replace algorithm that obtains the optimal solution. Moreover, we devise three replacement rules and derive the maximum compression ratio. The experimental results show that the proposed compression algorithm leverages the group movement patterns to reduce the amount of delivered data effectively and efficiently.

1 INTRODUCTION

RECENT advances in location-acquisition technologies, such as global positioning systems (GPSs) and wireless sensor networks (WSNs), have fostered many novel applications like object tracking, environmental monitoring, and location-dependent service. These applications generate a large amount of location data, and thus, lead to transmission and storage challenges, especially in resourceconstrained environments like WSNs. To reduce the data volume, various algorithms have been proposed for data compression and data aggregation [1], [2], [3], [4], [5], [6]. However, the above works do not address application-level semantics, such as the group relationships and movement patterns, in the location data. In object tracking applications, many natural phenomena show that objects often exhibit some degree of regularity in their movements. For example, the famous annual wildebeest migration demonstrates that the movements of creatures are temporally and spatially correlated. Biologists also have found that many creatures, such as elephants, zebra, whales, and birds, form large social groups when migrating to find food, or for breeding or wintering. These characteristics indicate that the trajectory data of multiple objects may be correlated for biological applications. Moreover, some research domains, such as the study of animals' social behavior and wildlife migration [7], [8], are more concerned with the movement patterns of groups of animals, not individuals; hence, tracking each object is unnecessary in this case. This raises a new challenge of finding moving animals belonging to the same group and identifying their aggregated group movement patterns. Therefore, under the assumption that objects with similar movement patterns are regarded as a group, we define the moving object clustering problem as given the movement trajectories of objects, partitioning the objects into nonoverlapped groups such that the number of groups is minimized. Then, group movement pattern discovery is to find the most representative movement patterns regarding each group of objects, which are further utilized to compress location data. Discovering the group movement patterns is more difficult than finding the patterns of a single object or all objects, because we need to jointly identify a group of objects and discover their aggregated group movement patterns. The constrained resource of WSNs should also be considered in approaching the moving object clustering problem. However, few of existing approaches consider these issues simultaneously. On the one hand, the temporal-and-spatial correlations in the movements of moving objects are modeled as sequential patterns in data mining to discover the frequent movement patterns [9], [10], [11], [12]. However, sequential patterns 1) consider the characteristics of all objects, 2) lack information about a frequent pattern's significance regarding individual trajectories, and 3) carry no time information between consecutive items, which make them unsuitable for location prediction and similarity comparison. On the other hand, previous works, such as [13], [14], [15], measure the similarity among these entire trajectory sequences to group moving objects. Since objects may be close together in some types of terrain, such as gorges, and widely distributed in less rugged areas, their group relationships are distinct in some areas and vague in others.

Thus, approaches that perform clustering among entire trajectories may not be able to identify the local group relationships. In addition, most of the above works are centralized algorithms [9], [10], [11], [12], [13], [14], [15], which need to collect all data to a server before processing. Thus, unnecessary and redundant data may be delivered, leading to much more power consumption because data transmission needs more power than data processing in WSNs [5]. In [16], we have proposed a clustering algorithm to find the group relationships for query and data aggregation efficiency. The differences of [16] and this work are as follows: First, since the clustering algorithm itself is a centralized algorithm, in this work, we further consider systematically combining multiple local clustering results into a consensus to improve the clustering quality and for use in the update-based tracking network. Second, when a delay is tolerant in the tracking application, a new data management approach is required to offer transmission efficiency, which also motivates this study. We thus define the problem of compressing the location data of a group of moving objects as the group data compression problem. Therefore, in this paper, we first introduce our distributed mining algorithm to approach the moving object clustering problem and discover group movement patterns. Then, based on the discovered group movement patterns, we propose a novel compression algorithm to tackle the group data compression problem. Our distributed mining algorithm comprises a Group Movement Pattern Mining (GMPMine) and a Cluster Ensembling (CE) algorithms.

It avoids transmitting unnecessary and redundant data by transmitting only the local grouping results to a base station (the sink), instead of all of the moving objects' location data. Specifically, the GMPMine algorithm discovers the local group movement patterns by using a novel similarity measure, while the CE algorithm combines the local grouping results to remove inconsistency and improve the grouping quality by using the information theory. Different from previous compression techniques that remove redundancy of data according to the regularity within the data, we devise a novel two-phase and 2D algorithm, called 2P2D, which utilizes the discovered group movement patterns shared by the transmitting node and the receiving node to compress data. In addition to remove redundancy of data according to the correlations within the data of each single object, the 2P2D algorithm further leverages the correlations of multiple objects and their movement patterns to enhance the compressibility. Specifically, the 2P2D algorithm comprises a sequence merge and an entropy reduction phases. In the sequence merge phase, we propose a Merge algorithm to merge and compress the location data of a group of objects. In the entropy reduction phase, we formulate a Hit Item Replacement (HIR) problem to minimize the entropy of the merged data and propose a Replace algorithm to obtain the optimal solution.

The Replace algorithm finds the optimal solution of the HIR problem based on Shannon's theorem [17] and guarantees the reduction of entropy, which is conventionally viewed as an optimization bound of compression performance. As a result, our approach reduces the amount of delivered data and, by extension, the energy consumption in WSNs. Our contributions are threefold: . Different from previous works, we formulate a moving object clustering problem that jointly identifies a group of objects and discovers their movement patterns. The application-level semantics are useful for various applications, such as data storage and transmission, task scheduling, and network construction. . To approach the moving object clustering problem, we propose an efficient distributed mining algorithm to minimize the number of groups such that members in each of the discovered groups are highly related by their movement patterns. . We propose a novel compression algorithm to compress the location data of a group of moving objects with or without loss of information. We formulate the HIR problem to minimize the entropy of location data and explore the Shannon's theorem to solve the HIR problem. We also prove that the proposed compression algorithm obtains the optimal solution of the HIR problem efficiently. The remainder of the paper is organized as follows: In Section 2, we review related works. In Section 3, we provide an overview of the network, location, and movement models and formulate our problem. In Section 4, we describe the distributed mining algorithm. In Section 5, we formulate the compression problems and propose our compression algorithm. Section 6 details our experimental results. Finally, we summarize our conclusions in Section 7.

2 RELATED WORK

2.1 Movement Pattern Mining

Agrawal and Srikant [18] first defined the sequential pattern mining problem and proposed an Apriori-like algorithm to find the frequent sequential patterns. Han et al. consider the pattern projection method in mining sequential patterns and proposed FreeSpan [19], which is an FP-growth-based algorithm. Yang and Hu [9] developed a new match measure for imprecise trajectory data and proposed TrajPattern to mine sequential patterns. Many variations derived from sequential patterns are used in various applications, e.g., Chen et al. [20] discover path traversal patterns in a Web environment, while Peng and Chen [21] mine user moving patterns incrementally in a mobile computing system. However, sequential patterns and its variations like [20], [21] do not provide sufficient information for location prediction or clustering. First, they carry no time information between consecutive items, so they cannot provide accurate information for location prediction when time is concerned.

Second, they consider the characteristics of all objects, which make the meaningful movement characteristics of individual objects or a group of moving objects inconspicuous and ignored. Third, because sequential pattern lacks information about its significance regarding to each individual trajectory, they are not fully representative to individual trajectories. To

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

discover significant patterns for location prediction, Morzy mines frequent trajectories whose consecutive items are also adjacent in the original trajectory data [10], [22]. Meanwhile, Giannotti et al. [11] extract T-patterns from spatiotemporal data sets to provide concise descriptions of frequent movements, and Tseng and Lin [12] proposed the TMPMine algorithm for discovering the temporal movement patterns. However, the above Apriori-like or FP-growthbased algorithms still focus on discovering frequent patterns of all objects and may suffer from computing efficiency or memory problems, which make them unsuitable for use in resource-constrained environments.

2.2 Clustering

Recently, clustering based on objects' movement behavior has attracted more attention. Wang et al. [14] transform the location sequences into a transaction-like data on users and based on which to obtain a valid group, but the proposed AGP and VG growth are still Apriori-like or FP-growthbased algorithms that suffer from high computing cost and memory demand. Nanni and Pedreschi [15] proposed a density-based clustering algorithm, which makes use of an optimal time interval and the average euclidean distance between each point of two trajectories, to approach the trajectory clustering problem. However, the above works discover global group relationships based on the proportion of the time a group of users stay close together to the whole time duration or the average euclidean distance of the entire trajectories. Thus, they may not be able to reveal the local group relationships, which are required for many applications.

In addition, though computing the average euclidean distance of two geometric trajectories is simple and useful, the geometric coordinates are expensive and not always available. Approaches, such as EDR, LCSS, and DTW, are widely used to compute the similarity of symbolic trajectory sequences [13], but the above dynamic programming approaches suffer from scalability problem [23]. To provide scalability, approximation or summarization techniques are used to represent original data. Guralnik and Karypis [23] project each sequence into a vector space of sequential patterns and use a vector-based K-means algorithm to cluster objects. However, the importance of a sequential pattern regarding individual sequences can be very different, which is not considered in this work. To cluster sequences, Yang and Wang proposed CLUSEQ [24], which iteratively identifies a sequence to a learned model, yet the generated clusters may overlap which differentiates their problem from ours.

2.3 Data Compression

Data compression can reduce the storage and energy consumption for resource-constrained applications. In [1], distributed source (Slepian-Wolf) coding uses joint entropy to encode two nodes' data individually without sharing any data between them; however, it requires prior knowledge of cross correlations of sources. Other works, such as [2], [4],

combine data compression with routing by exploiting cross correlations between sensor nodes to reduce the data size. In [5], a tailed LZW has been proposed to address the memory constraint of a sensor device. Summarization of the original data by regression or linear modeling has been proposed for trajectory data compression [3], [6]. However, the above works do not address application-level semantics in data, such as the correlations of a group of moving objects, which we exploit to enhance the compressibility.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

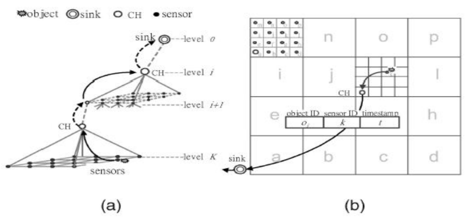

Fig. 1. (a) The hierarchical- and cluster-based network structure and the data flow of an update-based tracking network. (b) A flat view of a two-layer network structure with 16 clusters.

3 PRELIMINARIES

3.1 Network and Location Models

Many researchers believe that a hierarchical architecture provides better coverage and scalability, and also extends the network lifetime of WSNs [25], [26]. In a hierarchical WSN, such as that proposed in [27], the energy, computing, and storage capacity of sensor nodes are heterogeneous. A high-end sophisticated sensor node, such as Intel Stargate [28], is assigned as a cluster head (CH) to perform high complexity tasks; while a resource-constrained sensor node, such as Mica2 mote [29], performs the sensing and low complexity tasks. In this work, we adopt a hierarchical network structure with K layers, as shown in Fig. 1a, where sensor nodes are clustered in each level and collaboratively gather or relay remote information to a base station called a sink. A sensor cluster is a mesh network of n n sensor nodes handled by a CH and communicate with each other by using multihop routing [30]. We assume that each node in a sensor cluster has a locally unique ID and denote the sensor IDs by an alphabet . Fig. 1b shows an example of a two-layer tracking network, where each sensor cluster contains 16 nodes identified by ¼ fa, b; ... ; pg. In this work, an object is defined as a target, such as an animal or a bird, that is recognizable and trackable by the tracking network. To represent the location of an object, geometric models and symbolic models are widely used [31]. A geometric location denotes precise two-dimension or three-dimension coordinates; while a symbolic location represents an area, such as the sensing area of a sensor node or a cluster of sensor nodes, defined by the application. Since the accurate geometric location is not easy to obtain and techniques like the Received Signal Strength (RSS) [32] can simply estimate an object's location based on the ID of the sensor node with the strongest signal, we employ a symbolic model and describe the location of an object by using the ID of a nearby sensor node. Object tracking is defined as a task of detecting a moving object's location and reporting the location data to the sink

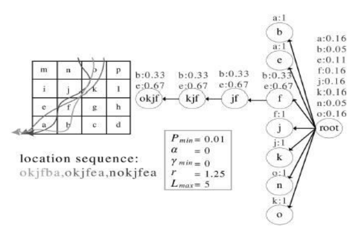

Fig. 2. An example of an object's moving trajectory, the obtained location sequence, and the generated PST T.

periodically at a time interval. Hence, an observation on an object is defined by the obtained location data. We assume that sensor nodes wake up periodically to detect objects. When a sensor node wakes up on its duty cycle and detects an object of interest, it transmits the location data of the object to its CH. Here, the location data include a times tamp, the ID of an object, and its location. We also assume that the targeted applications are delay-tolerant. Thus, instead of forwarding the data upward immediately, the CH compresses the data accumulated for a batch period and sends it to the CH of the upper layer. The process is repeated until the sink receives the location data. Consequently, the trajectory of a moving object is thus modeled as a series of observations and expressed by a location sequence, i.e., a sequence of sensor IDs visited by the object. We denote a location sequence by S ¼ s0s1 . . . sL_1, where each item si is a symbol in _ and L is the sequence length. An example of an object's trajectory and the obtained location sequence is shown in the left of Fig. 2. The tracking network tracks moving objects for a period and generates a location sequence data set, based on which we discover the group relationships of the moving objects.

7 CONCLUSIONS

In this work, we exploit the characteristics of group movements to discover the information about groups of moving objects in tracking applications. We propose a distributed mining algorithm, which consists of a local GMPMine algorithm and a CE algorithm, to discover group movement patterns. With the discovered information, we devise the 2P2D algorithm, which comprises a sequence merge phase and an entropy reduction phase. In the sequence merge phase, we propose the Merge algorithm to merge the location sequences of a group of moving objects with the goal of reducing the overall sequence length. In the entropy reduction phase, we formulate the HIR problem and propose a Replace algorithm to tackle the HIR problem. In addition, we devise and prove three replacement rules, with which the Replace algorithm obtains the optimal solution of HIR efficiently. Our experimental results show that the proposed compression algorithm effectively reduces the amount of delivered data and enhances compressibility and,by extension, reduces the energy consumption expense for data transmission in WSNs.

REFERENCES

[1] S.S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed Compression in a Dense Microsensor Network," IEEE Signal Processing Magazine, vol. 19, no. 2, pp. 51-60, Mar. 2002.
[2] A. Scaglione and S.D. Servetto, "On the Interdependence of Routing and Data Compression in Multi-Hop Sensor Networks," Proc. Eighth Ann. Int'l Conf. Mobile Computing and Networking, pp. 140-147, 2002.
[3] N. Meratnia and R.A. de By, "A New Perspective on Trajectory Compression Techniques," Proc. ISPRS Commission II and IV, WG II/5, II/6, IV/1 and IV/2 Joint Workshop Spatial, Temporal and Multi-Dimensional Data Modelling and Analysis, Oct. 2003.
[4] S. Baek, G. de Veciana, and X. Su, "Minimizing Energy Consumption in Large-Scale Sensor Networks through Distributed Data Compression and Hierarchical Aggregation," IEEE J. Selected Areas in Comm., vol. 22, no. 6, pp. 1130-1140, Aug. 2004.
[5] C.M. Sadler and M. Martonosi, "Data Compression Algorithms for Energy-Constrained Devices in Delay Tolerant Networks," Proc. ACM Conf. Embedded Networked Sensor Systems, Nov. 2006.
[6] Y. Xu and W.-C. Lee, "Compressing Moving Object Trajectory in Wireless Sensor Networks," Int'l J. Distributed Sensor Networks, vol. 3, no. 2, pp. 151-174, Apr. 2007.
[7] G. Shannon, B. Page, K. Duffy, and R. Slotow, "African Elephant Home Range and Habitat Selection in Pongola Game Reserve, South Africa," African Zoology, vol. 41, no. 1, pp. 37-44, Apr. 2006.
[8] C. Roux and R.T.F. Bernard, "Home Range Size, Spatial Distribution and Habitat Use of Elephants in Two Enclosed Game Reserves in the Eastern Cape Province, South Africa," African J. Ecology, vol. 47, no. 2, pp. 146-153, June 2009.
[9] J. Yang and M. Hu, "Trajpattern: Mining Sequential Patterns from Imprecise Trajectories of Mobile Objects," Proc. 10th Int'l Conf. Extending Database Technology, pp. 664-681, Mar. 2006.