Discovering Data Storage by Task Scheduling and Network Construction for Data Compression in Global Position System

K Adinarayana 1 K Lakshmaiah 2
Madanapalli Institute of Technology and Sciences, Angallu, India,
1sree.adinarayana@gmail.com, 2klakshmaiah78@gmail.com

Abstract: Natural trend show number of animals forms large social groups and move in regular patterns. We formulate a moving object clustering problem that jointly identifies a group of objects and discovers their movement patterns. Data storage and transmission, task scheduling, and network construction used for that. We have proposed a clustering algorithm to find the group relationships for query and data aggregation efficiency. 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 tolerable 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. 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) algorithm. 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.

Keywords: Data Compression; GMPMine; Distributed Clustering; Object Tracking.

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 resource constrained environments like WSNs. To reduce the data volume, various algorithms have been proposed for data compression and data aggregation. 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. 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. However, few of existing approaches consider these issues at once. 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. 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 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. We develop a novel two-phase and two dimensional algorithms, called 2P2D, which utilizes the discovered group movement patterns shared by the transmitting node and the receiving node to compress data. The constrained resource of WSNs should also be considered in approaching the moving object clustering problem. This is 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.

2. ASSOCIATED WORKS

A. Movement Pattern Mining

First defined the sequential pattern mining problem and proposed an Apriori-like Algorithm to find the frequent sequential patterns by the Agrawal and Srikant [1]. Pattern Projection method in mining sequential patterns consider by the Han et al and proposed FreeSpan [2], which describes an FP-growth-based algorithm. Chen et al. [3] discover path traversal patterns in a Web environment, while Peng and Chen [4] mine user moving patterns incrementally in a mobile computing system. However, sequential patterns and its variations like [3], do not provide sufficient information for location prediction or clustering. First, they did not offer perfect information for location prediction when time is concerned because they carry no time information between consecutive items second, they consider all objects characteristics, which make the meaningful movement characteristics of individual objects or a group of moving objects inconspicuous and ignored. Third, because a sequential pattern lacks information about its significance regarding to each individual trajectory, they are not fully representative to individual trajectories. In the meantime, Giannotti et al. [5] extract T-patterns from spatiotemporal data sets to provide concise descriptions of frequent movements, and Tseng and Lin [6] proposed the TMPMine algorithm for discovering the temporal movement patterns.

B. Clustering

Wang et al. [7] 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-growth-based algorithms that suffer from high computing cost and memory demand. Nanni and Pedreschi [8] proposed a density-based clustering algorithm, which makes use of an optimal time interval and finding the average euclidean distance between each point of two trajectories, for trajectory clustering problem. To cluster sequences, Yang and Wang proposed CLUSEQ [9], which iteratively classify a sequence to a learned model; however the created clusters may lie on top which differentiates their problem from ours.

C. Data Compression

Data compression can reduce the data space or transmission capacity for resource-constrained applications. Slepian-Wolf[10] distributed source coding uses joint entropy to encode two nodes’ data individually without sharing any data between them; yet, it requires prior awareness of cross correlations of sources. Related works, such as [11], [12], combine data compression with routing by exploiting cross correlations between sensor nodes to reduce the data size. In [13], 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 [14]. The above works do not address correlations of a group of moving objects, which we exploit to increase the compressibility.

3. PRELIMINARIES

A. Network and Location Models

A hierarchical architecture provides better coverage and scalability, and also extends the network lifetime of WSNs [15], [16]. A high-end sophisticated sensor node, such as Intel Stargate [17], is assigned as a
cluster head (CH) to perform high complexity tasks; while a resource-constrained sensor node, such as Mica2 mote [18], performs the sensing and low complexity tasks.

For the above work, we adopt a hierarchical network structure with \( K \) layers, as shown in Fig. 1a, where sensor nodes are grouped 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 \( n \times n \) sensor nodes handled by a CH and communicate with each other by using multihop routing [19]. We assume that each node in a sensor cluster has a locally unique ID and denote the sensor IDs by an alphabet \( \sum \). Fig. 1b shows an example of a two-layer tracking network, where each sensor cluster contains 16 nodes identified by \( \sum = \{a, b, \ldots, p\} \). Object tracking is defined as a task of detecting a moving object’s location and reporting the location data to the sink periodically at a time interval.

**B. Variable Length Markov Model (VMM) and Probabilistic Suffix Tree (PST)**

With the help Variable Length Markov Model (VMM), we model the regularity. The objects next location can be predicted based on its preceding locations, if and only if the movements of an object are regular. To learn the significant movement patterns, we adapt

Probabilistic Suffix Tree (PST) [20] for it has the lowest storage requirement among many VMM implementations [21]. PST’s low complexity, i.e., \( O(n) \) in both time and space [22], also makes it more attractive especially for streaming or resource-constrained environments [23].

**C. Problem Description**

We devise the problem of this paper as discovering the group movement patterns to compress the location sequences of a group of moving objects for transmission efficiency. Consider a set of moving objects \( O = \{O_1, O_2, \ldots, O_n\} \) and their associated location sequence data set \( S = \{S_1, S_2, \ldots, S_n\} \).

**Definition 1.** Two objects are similar to each other if their movement patterns are similar.

**Definition 2.** A set of objects is recognized as a group if they are highly similar to one another.

4. **MINING OF GROUP MOVEMENT PATTERNS**

We recommend a distributed mining algorithm for moving object clustering problem, which include the CE algorithms and GMPMine.

**A. The Group Movement Pattern Mining (GMPMine) Algorithm**

Since the optimization of the graph separation problem is difficult in general, we split the similarity graph in the following way. We influence the HCS cluster algorithm [24] to separation the graph and get the location group information. GMPMine include important steps. First, extract the moving patterns from the location sequences by knowledge PST for each object. Second, our algorithm constructs an undirected, unweighted similarity graph where similar objects share an edge between each other. Third, we select the ensembling result.
5. DESIGN OF A COMPRESSION ALGORITHM WITH GROUP MOVEMENT PATTERNS

A WSN is self-possessed of a large number of tiny sensor nodes that are install in a remote area for different tasks, some of them are environmental monitoring or wildlife tracking. Such sensor nodes are regularly battery-powered and recharging a large number of them is difficult. Therefore, energy protection is vital among all design issues in WSNs [25], [26]. Because the target objects are moving, preserving energy in WSNs for tracking moving objects is more difficult than in WSNs that monitor motionless phenomena, such as humidity or vibrations. Hence, previous works, such as [27], especially consider movement characteristics of moving objects in their designs to track objects efficiently. And the most energy costly tasks in WSNs is transmission of data, data compression is utilized to reduce the amount of delivered data. To reduce the amount of delivered data, this paper suggests the 2P2D algorithm which leverages the group movement patterns to compress the location sequences of moving objects richly. The algorithm includes the sequence merge phase and the entropy reduction phase to compress location sequences vertically and horizontally.

A. Sequence Merge Phase

The following wild animals, several moving objects may have cluster relationships and share alike trajectories. Transmitting their location data separately leads to redundancy. Therefore, in this section, we focus on the problem of compressing multiple comparable sequences of a group of moving objects.

![Fig. 2. An example of the Merge algorithm. (a) Three sequences with high similarity. (b) The merged sequence S”.](image)

Particularly, specified a collection of n sequences, the items of an S-column are replaced by a single symbol, whereas the items of a D-column are enfolded up between two ‘/ ’ symbols. Our algorithm generates a merged sequence containing the same information of the original sequences. In decompressing from the merged sequence, while symbol ‘/ ’ is encountered, the items after it are output until the next ‘/ ’ symbol. Otherwise, for each item, we repeat it n times to generate the original sequences. Fig. 2b shows the merged sequence S” whose length is decreased from 60 items to 48 items such that 12 items are sealed.

**Algorithm:** Sequence Merge

**Input:** a group of sequence \{S_i | 0 ≤ i < n\} with length L an error bound eb

**Output:** a merged sequence S”

0. Initialize ps, S”
1. dc_start = 0
2. for 0 ≤ j < L
3. \( σ = \text{null} \)
4. if is_S-column (S_i[j], 0 ≤ i < n) then
5. \( σ = S_0[j] \)
6. else if eb > 0 then
7. \( σ = \text{getRS} \{S_i[j], 0 ≤ i < n\}, eb \)
8. if σ = = null then
9. if dc_start = = 0 then
10. append (S”,”/’)
11. dc_start = 1
12. for 0 ≤ i < n
13. append (S”S_i[j])
14. else
15. if dc_start = = 1 then
16. append (S”,”/’)
17. dc_start = 0
18. append (S”, $\alpha$)
19. return S”

Table 1: Description of the Notations

<table>
<thead>
<tr>
<th>Description</th>
<th>Notations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Sigma$</td>
<td>The alphabet of Sensor IDs.</td>
</tr>
<tr>
<td>S</td>
<td>A location sequence.</td>
</tr>
<tr>
<td>‘/’</td>
<td>The hit symbol that is used to replace a predictable item.</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>A symbol in $\Sigma$.</td>
</tr>
<tr>
<td>$S_{ij}$</td>
<td>$j^{th}$ item of $S_i$.</td>
</tr>
</tbody>
</table>

Since data with lower entropy require fewer bits for storage and transmission, we replace some items to reduce the entropy without loss of information. We first introduce and define the Hit Item Replacement problem, and then, explore the properties of Shannon’s entropy to solve the HIR problem.

6. EXPERIMENTS AND ANALYSIS

We implement this paper in c sharp and dot net. To the best of our knowledge, no research work has been committed to discovering data storage, task scheduling and network construction for data compression. Moreover, we use the amount of data in kilobyte (KB) and compression ratio (r) as the evaluation metric, where the compression ratio is defined as the ratio between the uncompressed data size and the compressed data size, i.e., $r = \frac{\text{uncompressed data size}}{\text{compressed data size}}$.

7. CONCLUSION

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. 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


International Journal on Advanced Computer Engineering and Communication Technology Vol-I Issue:1 :ISSN 2278 – 5140


Discovering Data Storage by Task Scheduling and Network Construction for Data Compression in Global Position System