==
Engineering & Technology in India www.engineeringandtechnologyinindia.com
Vol. 1:2 March 2016
==

Application Level Semantics for Compressing Group Movement Patterns in Wireless Sensor Networks

J. Biju, M.E.
==

Abstract

This paper proposes an efficient distributed mining algorithm to jointly identify a group of moving objects and discover their movement patterns in wireless sensor networks. Then propose a compression algorithm, called two–phase and two-dimensional algorithm, which exploits the obtained group movement patterns to reduce the amount of delivered data. The compression algorithm includes a sequence merge phase and an entropy reduction phase. In the sequence merge phase, propose a merge algorithm to merge and compress the location data of a group of moving objects. In the entropy reduction phase, formulate a Hit Item Replacement (HIR) problem and propose a replace algorithm that obtains the optimal solution. Then 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.

Key words: mining algorithm, wireless sensor, compression leverages

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 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 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 GMP Mine 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, 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.

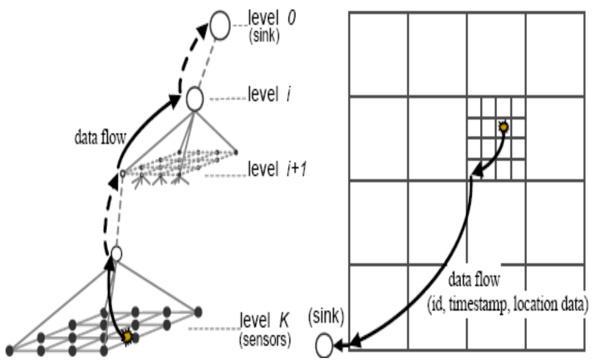

Figure 1.1: Network and location models

2. Proposed Compression Method

Consider the pattern projection method in mining sequential patterns and proposed FreeSpan, which is an FP-growth-based algorithm. Yang and Hu developed a new match measure for imprecise trajectory data and proposed Traj Pattern to mine sequential patterns. Many variations derived from sequential patterns are used in various applications, e.g., Chen et al. discover path traversal patterns in a Web environment, while Peng and Chen mine user moving patterns incrementally in a mobile computing system. However, sequential patterns and its variations 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 a sequential pattern lacks information about its significance regarding to each individual trajectory, they are not fully representative to individual trajectories.

To discover significant patterns for location prediction, Morzy mines frequent trajectories whose consecutive items are also adjacent in the original trajectory data. Meanwhile, Giannotti, et al. extract T-patterns from spatiotemporal data sets to provide concise descriptions of frequent movements, and Tseng and Lin proposed the GMP Mine algorithm for discovering the temporal movement patterns. However, the above A priori-like or FP-growth based 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.

For each of their significant movement patterns, the new similarity measure considers not merely two probability distributions but also two weight factors.

The similarity score simp. of oi and oj based on their respective PSTs, Ti and Tj, is defined as follows**:**

$$sim_p = -log\frac{\sum_{s\in\tilde{S}}\sqrt{\sum_{\sigma\in\Sigma}\left(P^{T_1}(s\sigma) - P^{T_2}(s\sigma)\right)^2}}{2L_{max}+\sqrt{2}}$$

The similarity score simp includes the distance associated with a pattern s, defined as*,*

$$d(s) = \sqrt{\sum_{\sigma\in\Sigma}\left(P^{T_i}(s\sigma) - P^{T_j}(s\sigma)\right)^2}$$
$$= \sqrt{\sum_{\sigma\in\Sigma}\left(P^{T_i}(s) \times P^{T_i}(\sigma|s) - P^{T_j}(s) \times P^{T_j}(\sigma|s)\right)^2},$$

where d(s) is the Euclidean distance associated with a pattern s over Ti and Tj.

A. Clustering

Recently, clustering based on objects' movement behavior has attracted more attention. Wang et al. 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 A priori-like or FP-growth based algorithms that suffer from high computing cost and memory demand. Nanni and Pedreschi 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 , but the above dynamic programming approaches suffer from scalability problem . To provide scalability, approximation or summarization techniques are used to represent original data. Guralnik and Karypis 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, which iteratively identifies a sequence to a learned model, yet the generated clusters may overlap which differentiates their problem from ours.

B. Data Compression

Data compression can reduce the storage and energy consumption for resource-constrained applications. In 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 include such as combine data compression with routing by exploiting cross correlations between sensor nodes to reduce the data size. In 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. 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.

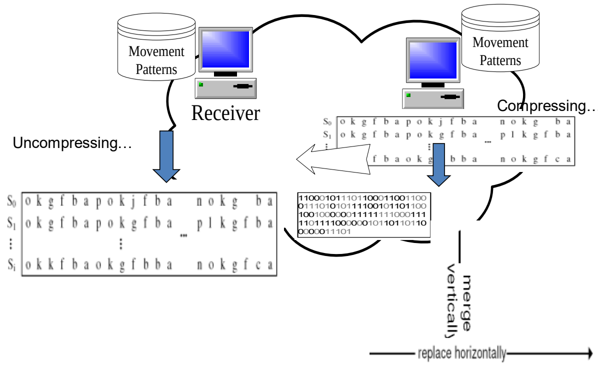

Figure 1.5 Design of a Compression Algorithm

C. Sequence Merge Phase

In the application of tracking wild animals, multiple moving objects may have group relationships and share similar trajectories. In this case, transmitting their location data separately leads to redundancy. Therefore, in this section, we concentrate on the problem of compressing multiple similar sequences of a group of moving objects.

D. Entropy Reduction Phase

In the entropy reduction phase, we propose the Replace algorithm to minimize the entropy of the merged sequence obtained in the sequence merge phase. Since data with lower entropy require fewer bits for storage and transmission, we replace some items to reduce the entropy without loss of information. The object movement patterns discovered by our distributed mining algorithm enable us to find the replaceable items and facilitate the selection of items in our compression algorithm. In this section, we first introduce and define the HIR problem, and then, explore the properties of Shannon's entropy to solve the HIR problem. We extend the concentration property for entropy reduction and discuss the benefits of replacing multiple symbols simultaneously. We derive three replacement rules for the HIR problem and prove that the entropy of the obtained solution is minimized.

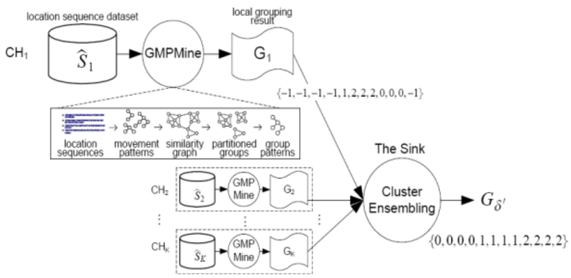

Figure3.2:Group movement pattern mining algorithm

E. Segmentation, Alignment, and Packaging

In an online update approach, sensor nodes are assigned a tracking task to update the sink with the location of moving objects at every tracking interval. In contrast to the online approach, the CHs in our batch-based approach accumulate a large volume of location data for a batch period before compressing and transmitting it to the sink; and the location update process repeats from batch to batch. In real-world tracking scenarios, slight irregularities of the movements of a group of moving objects may exist in the microcosmic view. Specifically, a group of objects may enter a sensor cluster at slightly different times and stay in a sensor cluster for slightly different periods, which lead to the alignment problem among the location sequences. Moreover, since the trajectories of moving objects may span multiple sensor clusters, and the objects may enter and leave a cluster. multiple times during a batch period, a location sequence may comprise multiple segments, each of which is a trajectory that is continuous in time domain. To deal with the alignment and segmentation problems, we partition location sequences into segments, and then, compress and package them into one update packet

3. Conclusion

The compression algorithm effectively reduces the amount of delivered data and enhances compressibility by extension, reduces the energy consumption expense for data transmission in Wireless Sensor Networks by exploiting the characteristics of group movement patterns to discover the information about groups of moving objects in tracking applications. The distributed mining algorithm consists of a local GMP Mine algorithm and a CE Algorithm to discover the group movement patterns.

With the discovered information ,devise the Two-Phase and Two–Dimensional algorithm ,which comprises a sequence merge phase and an entropy reduction phase. In sequence merge phase, 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 formulate the HIR problem and propose a replace algorithm to tackle the HIR problem. In addition devise and prove three replacement rules, with which the Replace algorithm obtains the optimal solution of HIR effectively.

==

References

[1] Hsiao-Ping Tsai,IEEE,De-Nian Yang,and Ming-Syan Chen,Fellow,"Exploring Application Level Semantics For Data Compression",IEEE Transactions on Knowledge And Data Engineering,vol 23,No.1,Jan 2011.

[2] 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.

[3] 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.

[4] N. Meratnia and R.A. de By, "A New Perspective on Trajectory Compression Techniques," Proc. ISPRS Commission II and IV, WGII/5, II/6, IV/1 and IV/2 Joint Workshop Spatial, Temporal and Multi-Dimensional Data Modelling and Analysis, Oct. 2003.

[5] C.M. Sadler and M. Martonosi, "Data Compression Algorithms for Energy-Constrained Devices in Delay Tolerant Networks".[6]R. Agrawal and R. Srikant, "Mining Sequential Patterns".

J. Biju, M.E.
Assistant Professor
Department of Computer Science
Theni Kammavar Sangam College of Technology

Theni 625 534
Tamilnadu
India
jbijuinfo@gmail.com