

Exploring Application-Level Semantics for Data Compression

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.

Existing System:

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 However, sequential patterns

- 1) Consider the characteristics of all objects,
- 2) Lack information about a frequent pattern's significance regarding individual trajectories,

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 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 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 Wireless Sensor Networks (WSNs).

Proposed System:

We have proposed a clustering algorithm to find the group relationships for query and data aggregation efficiency. The differences of 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. 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.

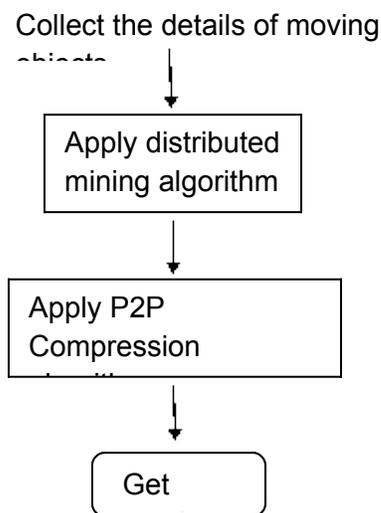
Our distributed mining algorithm comprises a Group Movement Pattern Mining (GMPMine) and a Cluster Ensembling (CE) algorithm. It avoids transmitting unnecessary and redundant data by transmitting only the local grouping results to a base station (the sink), instead of all of the moving objects' location 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.

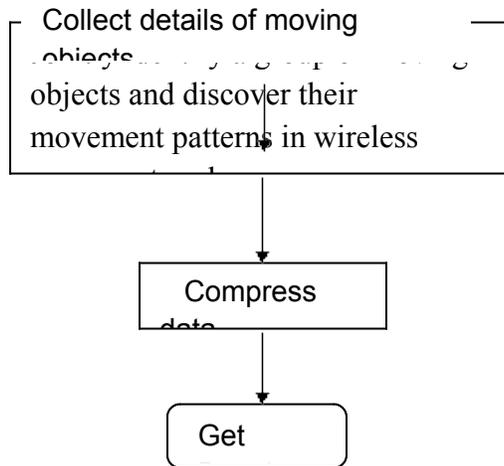
Different from previous compression techniques that remove redundancy of data according to the regularity within the data, we devise a novel two-phase and 2D algorithm, called 2P2D, which utilizes the discovered group movement patterns shared by the transmitting node and the receiving node to compress 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.

Architecture Diagrams:

1. Techniques used:



2. Process:



Module Description:

1) Input data

We 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 animal's social behavior and wildlife migration, are more concerned with the movement patterns of groups of animals. These details are given as an input data.

2) Apply mining technique

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.

3) Apply compression technique

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.

4) View result

View the data result that the result contains the mined and compressed data. 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.

SYSTEM SPECIFICATIONS:

HARDWARE REQUIREMENTS:

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 1.44 Mb.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.

- Ram : 512 MB.

SOFTWARE REQUIREMENTS:

- Operating system : Windows XP Professional.
- Coding Language : ASP .Net,C#
- Database : Sql Server 2005.

REFERENCE:

Hsiao-Ping Tsai, De-Nian Yang, and Ming-Syan Chen, “Exploring Application-Level Semantics for Data compression”, **IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No.1, January 2011.**