DISCOVERING APPLIANCE-LEVEL SEMANTICS FOR INFORMATION COMPRESSION

Abdul Haleem [1], Sri Lavanya Sajja [2]

[1]MCA Student, Dept of MCA, DRK College of Engineering and Technology, Hyderabad, Andhra Pradesh, India

[2]Assistant Professor, Dept of CSE, DRK College of Engineering and Technology, Hyderabad, Andhra Pradesh, India

ABSTRACT:

Latest advances in location-acquisition technologies have promoted many new applications like object tracing, conservational monitoring, and location reliant service. These requests produce a large amount of location information, and thus, lead to broadcast and storage contests, especially in resource constrained Surroundings. Usual sensations show that many individuals form great social groups and move in systematic forms. However, earlier mechanisms have focused on discovery the undertaking patterns of each particular object or all objects. In this paper, to mutually identify a group of moving objects and determine their undertaking patterns in wireless sensor networks an efficient distributed mining algorithm is proposed. Subsequently, to diminish the amount of distributed data, a compression algorithm, called 2 phase and 2D algorithm (2P2D) is proposed to exploit the acquired group movement patterns. A sequence merge and an entropy reduction phases is included by the compression algorithm. A Merge algorithm is proposed to merge and compress the location data of a group of moving objects in the sequence merge phase. A Hit Item Replacement (HIR) problem is formulated and a Replace algorithm is proposed so that it obtains the optimal solution in the entropy reduction phase. The investigational results show that the proposed compression algorithm influences the group movement patterns to diminish the amount of delivered data effectually and proficiently.

Keywords: Information Compression, Movement Patterns, Hit Item Replacement, Merge Algorithm

1. INTRODUCTION:

Many new applications have been promoted by location acquisition technologies in recent years. Applications like object locating, environmentally friendly monitoring and position reliant service generates a large amount of position data and indicates to broadcast and storing experiments in resource- controlled surroundings like WSNs [1] [2]. Numerous algorithms have been proposed for data compression and data aggregation in order to reduce the data volume. In object tracing applications, various natural phenomena show that objects frequently unveil some degree of consistency in their activities. Therefore, under the supposition that objects with related movement patterns are considered as a group [3] [4], we describe the moving object clustering problem as given the movement routes of objects, separating the objects into non- overlapped groups such that the number of groups is decreased. Then, group movement pattern detection is to find the most illustrative movement patterns concerning each group of objects, which are further used to compress location data [5]. We need to mutually identify a group of objects and determine their accumulated group movement patterns as determining the group movement patterns is more challenging than finding the patterns of a single object or all objects. In approaching the moving object clustering problem the constrained resource of WSNs should also be considered [6]. The basic idea of data compression is

that we are removing redundancies in the original text. An ultimate compression system can't possibly exist for a very modest reason. In a perfect compression system, the compression function in mapping is from a greater set to a lesser set. That means that for numerous inputs, it must produce the similar output. The better the ideal compressor i.e. the lesser it manages to make its compressed output the additional inputs must be compressed to the same output. If numerous inputs compress to the
similar output, then we don't have an invertible compression function: we can't decompress your text which is shown in Fig1.

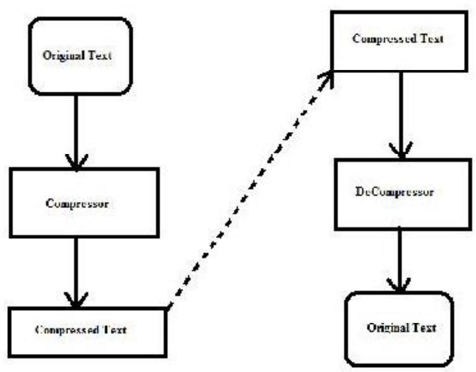

Fig 1. Data Compression Diagram

However, few current approaches consider these issues concurrently. On the one hand, to discover the frequent movement patterns the temporal-and-spatial correlations in the movements of moving objects are modeled as sequential patterns in data mining [7] [8]. However, sequential patterns deliberates the characteristics of all objects and lacks the information about a regular pattern's impact regarding individual routes and carry no time information between successive items, which make them inappropriate for location prediction and similarity comparison. To approach the moving object clustering problem and determine group movement patterns we propose a disseminated mining algorithm in this paper [9] [10]. To tackle the group data compression problem a novel compression algorithm is proposed which is based on the discovered group movement patterns. Our disseminated mining algorithm contains a Group Movement Pattern Mining (GMPMine) and a Cluster Ensembling (CE) [11] [12] algorithms which evades spreading redundant and dismissed data by transmitting only the local grouping results to a base station. We devise a novel twophase and 2D algorithm, called 2P2D that uses the exposed group movement patterns shared by the transmitting node and the receiving node to compress data. In addition to eliminate redundancy of data conferring to the associations within the data of each single object, the 2P2D algorithm further influences the associations of numerous objects and their movement patterns to improve the compressibility [13] [14]. Precisely, the 2P2D algorithm contains a sequence merge and an entropy lessening 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 articulate a Hit Item Replacement (HIR) problem to diminish the entropy of the combined data and propose a Replace algorithm to attain the optimum solution.

2. OVERVIEW OF CLUSTERING AND DATA COMPRESSION:

Lately, clustering based on objects' movement performance has attracted more attention. To approach the trajectory clustering problem, a density-based clustering algorithm is proposed such that it makes use of a best time interval and the average Euclidean distance between each point of two routes. However, the above works determines the universal group relationships based on the amount of the time a group of users stay close together to the whole time duration or the average Euclidean distance of the complete routes. Thus, they may not be able to disclose the local group associations, which are essential for various requests. In addition, though calculating the average Euclidean distance of two geometric routes is simple and beneficial, the geometric coordinates are exclusive and not always accessible. Data compression can lessen the storage and energy consumption for resource constrained requests. The disseminated source coding uses combined entropy to encrypt two nodules' data independently without distributing any data between them; however, it involves earlier information of cross associations of sources. Other works, such as associated data compression with routing by using cross associations between sensor nodes to lessen the data size. For route data compression, the summarization of the inventive data by reversion or linear modeling has been proposed. However, the above works do not address application level semantics in data, such as the associations of a set of moving objects, which we exploit to improve the compressibility.

3. NETWORK AND LOCATION SIMULATIONS:

Several investigators believe that a categorized architecture provides better exposure and scalability, and also outspreads the network lifetime of WSNs. In a categorized WSN, the energy, calculating, and storing capacity of sensor nodes are varied. A high-end refined sensor node is allocated as a cluster head (CH) to accomplish high difficulty tasks; while a resource-constrained sensor node achieves the sensing and low difficulty tasks. In this paper, we adopt a classified network structure with N layers where sensor nodes are gathered 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 x n sensor nodes controlled by a CH and connect with each other by using multi hop routing. We assume that each node in a sensor cluster has a nearby distinctive ID and denote the sensor IDs by an alphabet. In this paper, an object is well-defined as a target so that it can be identifiable and controllable by the tracking network. Geometric models and symbolic models are widely used to denote the location of an object. A geometric location signifies exact two-dimension or three-dimension coordinates; while a symbolic location represents an area defined by the request. Since the exact geometric location is not easy to find and methods like the Received Signal Strength (RSS) can simply estimate an object's location based on the ID of the sensor node with the toughest signal, we employ a symbolic model and designate the location of an object by using the ID of a nearby sensor node.

4. RESULT:

We implement an event-driven simulator to evaluate the performance of our design. A location-dependent mobility model is used to simulate the roaming behavior of a group leader. We utilize the group dispersion radius (GDR) to control the dispersion degree of the objects. We use the amount of data and compression ratio as the estimation metric, where the compression ratio is

defined as the ratio between the uncompressed data size and the compressed data size. The amount of data of our batch based approach is relatively low and stable as the GDR increases.

5. CONCLUSION:

In this work, we exploit the features of group activities to determine the information about groups of moving objects in tracing requests. To determine group movement patterns, a dispersed mining algorithm is proposed so that it contains of a local GMP Mine algorithm and a CE algorithm. A Merge algorithm to merge the location sequences of a group of moving objects with the goal of reducing the overall sequence length in the sequence merge phase. The HIR problem is formulated and a Replace algorithm is proposed to tackle the HIR problem in the entropy reduction phase. In addition, we develop and demonstrate three standby rules, with which the Replace algorithm attains the ideal solution of HIR proficiently. The anticipated compression algorithm efficiently diminishes the amount of delivered data and improves compressibility which is demonstrated in our results.

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

ISSN-2320-7884 (ONLINE)
ISSN-2321-0257 (PRINT)

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.
[10] M. Morzy, "Mining Frequent Trajectories of Moving Objects for Location Prediction," Proc. Fifth Int'l Conf. Machine Learning and Data Mining in Pattern Recognition, pp. 667-680, July 2007.
[11] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory Pattern Mining," Proc. ACM SIGKDD, pp. 330-339, 2007.
[12] V.S. Tseng and K.W. Lin, "Energy Efficient Strategies for Object Tracking in Sensor Networks: A Data Mining Approach," J. Systems and Software, vol. 80, no. 10, pp. 1678-1698, 2007.
[13] L. Chen, M. Tamer O ¨ zsu, and V. Oria, "Robust and Fast Similarity Search for Moving Object Trajectories," Proc. ACM SIGMOD, pp. 491-502, 2005.
[14] Y. Wang, E.-P. Lim, and S.-Y. Hwang, "Efficient Mining of Group Patterns from User Movement Data," Data Knowledge Eng., vol. 57, no. 3, pp. 240-282, 2006.

BIOGRAPHY:

Abdul Haleem has completed B.SC (Computers) from Manjeera Degree College, Patancheru, Medak, Andhra Pradesh, and pursuing MCA from DRK College of Engineering and Technology, JNTUH, Hyderabad, Andhra Pradesh, India. His main research interest includes Databases and Data Mining.

Sri Lavanya Sajja is working as an Assistant Professor in DRK College of Engineering and Technology, JNTUH, Hyderabad, Andhra Pradesh, India. She has received M.Tech degree in Computer Science and Engineering from JNTUH along with an M.Tech degree in IT. Her main research interest includes Data Mining and Networking.