IJDCST @Jan Issue- V-2, I-1, SW-2
ISSN-2320-7884 (Online) ISSN-2321-0257 (Print)

Dynamic Data Stream Compression Methods for Application-Level Semantics of Sensor Networks

[1]Alapati Harshavardhan,[2] D.Lokesh Sai Kumar
[1]M.Tech, PVPSIT,Kanuru,Vijayawada.
[2]Assistant Professor, PVPSIT,Kanuru,Vijayawada.

Abstract: Advances in location-acquisition technologies like global positioning systems (GPSs) and wireless sensor networks (WSNs) etc fostered many innovative applications like object tracking, environmental monitoring, and location-dependent services. These applications typically generate a large amount of location data, and thus, leading to transmission and storage challenges, especially in resource constrained environments like WSNs. Previously an implementation that can discover aggregated group movement patterns data and compress them rather than every data tuples originating from WSNs was developed. Driven by Cluster Ensemble approach along with a 2P2D algorithm for compression aspects it is able to deliver satisfactory results. Clustering ensembles combine multiple partitions of the given data into a single clustering solution of better quality by averaging multiple solutions obtained from an ensemble. Works efficiently well to categorize static data sets of voluminous information. Not efficient to dynamic data streams which is usually the chosen mode of data transfer in WSN. So we propose to replace only the clustering aspect of the above approach with any one of the Data Stream Clustering techniques for performance improvement. Using this dynamic data stream clustering provided with the above data compression aspects we can achieve an optimal system that that can discover and compress aggregated group movement patterns data efficiently and a practical implementation validates the claim.

Index Terms: Global Positioning Systems, wireless sensor networks, Trajectory Compression, Cluster Ensemble approaches.

I. INTRODUCTION

Sensory data comes from multiple sensors of different modalities in distributed locations. Recent location- acquisition technologies such as global positioning system and wireless sensor networks are have often in object tracking, environmental monitoring and location-dependent services.

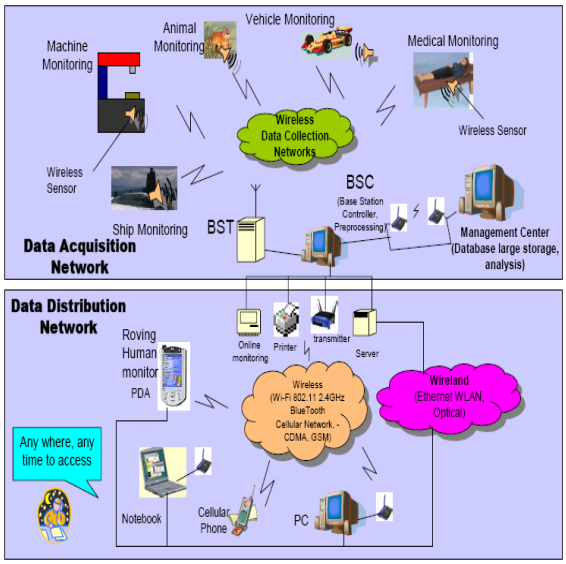

Figure 1: Data acquisitioning in wireless sensor networks for editing data analysis.

In object tracking applications many technical applications are developed in the sequential object often in some degree of regularity in their movements of positions. QoS can be specified in terms of

message delay, message due dates, bit error rates, packet loss, economic cost of transmission, transmission power, etc. Depending on QoS, the installation environment, economic considerations, and the application, one of several basic network topologies may be used. 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, leading to transmission and storage challenges, especially in resource constrained environments like WSNs. To implement the moving object clustering problem, we propose an efficient distributed mining algorithms called GMP Mine and Cluster Ensemble(CE) algorithm (predict next) to minimize the number of groups such that members in each of the discovered groups are highly related by their movement patterns. Then we propose a novel compression algorithm called 2P2D in which the initial phase is to compress (Merge Algorithm) the location data of a group of moving objects with or without loss of information.

Prior approaches used Cluster Ensemble approaches to minimize the number of groups such that members in each of the discovered groups are highly related by their movement patterns. Due to the inefficiency single clustering approach different clustering ensemble approach (predict next, Homogeneous ensembles, Random-k, Data subspace/sampling, Heterogeneous ensembles, Mixed heuristics) to obtain data clusters were developed and used. Clustering ensembles combine multiple partitions of the given data into a single clustering solution of better quality by averaging multiple solutions obtained from an ensemble.

II. RELATED WORK

Movement Pattern Mining

Agrawal and Srikant 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 Free Span, which is an FP-growth-based algorithm. 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 TMPMine algorithm for discovering the temporal movement patterns. However, the above Apriori-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.

Clustering

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 Apriori-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.

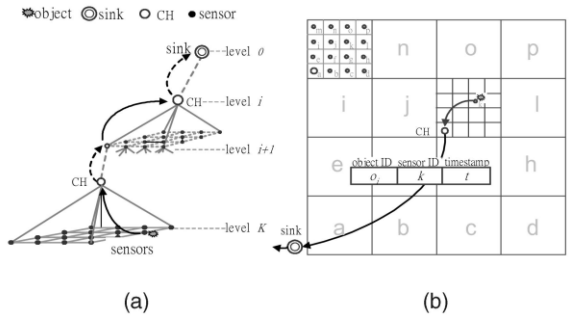

Figure 2: (a) The hierarchical- and cluster-based network structure and the data flow of an update-based tracking network. (b) A flat view of a two layer network structure with 16 clusters.

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. Summarization of the original data by regression or linear modeling has been proposed for trajectory data compression.

III. EXISTING SYSTEM

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, leading to transmission and storage challenges, especially in resource constrained environments like WSNs.

However, the above works do not address application-level semantics, such as the group relationships and movement patterns, in the location data. That means 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 non overlapped 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.

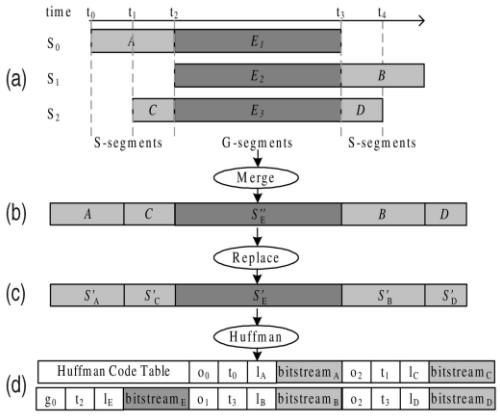

Figure 3: An example of constructing an update packet. (a) Three sequences aligned in time domain. (G-segments: E1, E2, and E3, Ssegments: A, B , C, and D.) (b) Combining G-segments by the Merge algorithm. (c) Replacing the

predictable items by the Replace algorithm. (d) Compressing and packing to generate the payload of the update packet.

A distributed mining algorithm, which consists of a local GM Mine 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.

IV. PROPOSED APPROACH

Prior approaches used Cluster Ensemble approaches to minimize the number of groups such that members in each of the discovered groups are highly related by their movement patterns. Due to the inefficiency single clustering approach different clustering ensemble approaches (predict next, Homogeneous ensembles, Random-k, Data subspace/sampling, Heterogeneous ensembles, Mixed heuristics) to obtain data clusters were developed and used. Clustering ensembles combine multiple partitions of the given data into a single clustering solution of better quality by averaging multiple solutions obtained from an ensemble. Works efficiently well to categorize static data sets of voluminous information. Not efficient to dynamic data streams which is usually the chosen mode of data transfer in WSN. So we propose to replace only the clustering aspect of the above approach with any one of the Data Stream Clustering techniques for performance improvement. Using this dynamic data stream clustering provided with the above data compression aspects we can achieve an optimal system that that can discover and compress aggregated group movement patterns data efficiently.

V. EXPERIMENTAL RESULTS

We study the effectiveness of the Replace algorithm by comparing the compression ratios of the Huffman encoding with and without our Replace algorithm. To the best of our knowledge, no research work has been dedicated to discovering application-level semantic for location data compression. We compare our batch-based approach with an online approach for the overall system performance evaluation and study the impact of the group size (n), as well as the group dispersion radius (GDR), the batch period (D), and the error bound of accuracy (eb). Our experiments are designed to demonstrate: 1) the performance gain of multiple random projections over a single random projection, and 2) that our proposed ensemble method outperforms PCA, a traditional approach to dimensionality reduction for clustering.

Evaluation Criteria

Evaluating clustering results is a nontrivial task. Because our method does not generate any model or description for the _nal clusters, internal criteria such as log-likelihood and scatter separability cann't be pplied.5 Because our data sets are labeled, we can assess the cluster quality by using measures such as conditional entropy and normalized mutual information. We chose to report results for both

criteria because as we explain below entropy is biased toward a large number of clusters and normalized mutual information under some conditions is biased toward solutions that have the same number of clusters as there are classes.

Analysis of Diversity for Cluster Ensembles:

For supervised ensemble approaches, diversity of the base-level classifiers has proven to be a key element in increasing classification performance. In comparing the diversity/quality results to the performance of the entire ensemble (indicated by the dotted line in each graph), we see evidence that for an ensemble of size thirty, high diversity leads to greater improvements in the ensemble quality. Specifically, we see the least improvement of the ensemble over a single run of RP+EM for the EOS data set, which has significantly lower diversity than the other two. On the other hand, less improvement is obtained for the HRCT data set in comparison with the CHART data set, which suggests that the quality of individual clustering solutions also limits the performance of a fixedsize ensemble. To gain further insight into these issues, we examined the impact of the ensemble size on performance.

VI. CONCLUSION

Driven by Cluster Ensemble approach along with a 2P2D algorithm for compression aspects it is able to deliver satisfactory results. Clustering ensembles combine multiple partitions of the given data into a single clustering solution of better quality by averaging multiple solutions obtained from an ensemble. Works efficiently well to categorize static data sets of voluminous information. Not efficient to dynamic data streams which is usually the chosen mode of data transfer in WSN. So we propose to replace only the clustering aspect of the above approach with any one of the Data Stream Clustering techniques for performance improvement. Using this dynamic data stream clustering provided with the above data compression aspects we can achieve an optimal system that that can discover and compress aggregated group movement patterns data efficiently and a practical implementation validates the claim.

VII. REFERENCES

[1] Achlioptas, D. (2001). Database-friendly random projections. Proceedings of the Twentieth ACM Symposium on Principles of Database Systems (pp. 274{ 281). ACM Press.

[2] Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications. Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data (pp. 94{105). ACM Press.

[3] Bingham, E., & Mannila, H. (2001). Random projection in dimensionality reduction: Applications to image and text data. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 245{250). ACM Press.

[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] Hsiao-Ping Tsai, "Exploring Application-Level Semantics for Data Compression", Ieee Transactions On Knowledge And Data Engineering, Vol. 23, No. 1, January 2011.