Data Transfer Model - Tracking and Identification of Data Files Using Clustering Algorithms

Ramkumar.S[1], Elakkiya.A[2], Emayavaramban.G[3]

[1]*Research Scholar, Karpagam University, Coimbatore, India.*
[2]*Research Scholar, Karpagam University, Coimbatore, India.*
[3]*Research Scholar, Karpagam University, Coimbatore, India.*

***Abstract*—This paper investigates and utilized the characteristic of the group movement of objects to explore the group relationship and tracking them. The goal is to efficiently mine the group movement activity using clustering and sequential pattern mining. Clustering was applied to find both groups of similar teams and similar individual members. Sequential pattern mining was used to extract sequences of frequent events. To enable the continuous monitoring the group object movement, the system introduces a special technique called minor clustering and Cluster Ensembling algorithm. Several solutions on route were implemented, but those methods were energy consumed. In order to reduce the energy the proposed system used data mining methods to effectively handle the group movement of objects.**

Keywords*— Cluster Ensembling, Wireless sensor Network.Minor clustering, Cluster Head, Energy Efficient Object Tracking Sensor Network.Heterogeneous Tracking Model.

I.INTRODUCTION

In object tracking applications, many natural phenomena show that moving objects often exhibit some degree of regularity in their movements. 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 large amounts of location data, and many approaches focus on compiling the collected data to identify the repeating movement patterns of objects of interest. The objective is to facilitate the analysis of past movements and estimate future movements, as well as support approximate queries on the original data. Existing object tracking applications focus on finding the moving patterns of a single object or all objects. In contrast, the proposed distributed mining algorithm that identifies a group of objects with similar movement patterns. This information is important in some biological research domains, such as the study of animals' social behavior and wildlife migration. The proposed algorithm comprises a local mining phase and a cluster ensembling phase. In the local mining phase, the algorithm finds movement patterns based on local trajectories. Then, based on the derived patterns, the proposed new similarity measure to compute the similarity of moving objects and identify the local group relationships. To address the energy conservation issue in resource-constrained environments, the algorithm only transmits the local grouping results to the sink node for further ensembling. In the cluster ensembling phase, our algorithm combines the local grouping results to derive the group relationships from a global view. The proposed model further influences the mining results to track moving objects efficiently.

In our study we are identify and track the group of objects with past movements and estimate the future movements. Energy based enhancements also considered as the main goal. It aims to provide the effective tracking mechanism in a distributed clustering environment. To address the group object tracking problems in the distributed environment, the proposed system has discovered minor clustering with CE algorithm. To generate a little research has been dedicated to discovering a group of objects with similar movement patterns for data aggregation purposes. Moreover, the group information enables to adaptively adjust the range in which a group of objects is monitored and, thereby, limit the overhead due to flooding within that range. Therefore, the proposed work aimed to reduce long-distance transmissions and the amount of data to be transmitted. It also reduces in-network traffic and prevents hot spots around the cluster heads (CHs).

Literature review details are explained in section 2, Data Mining and its types are explained in section 3, while section 4 details the methodology techniques. The experimental results are presented and discussed in section 5 and section 6 provides the conclusion and future work.

II. BACKGROUND

Research on data tracking model has been undertaken in the last decade, some of the prominent studies are given below. Wen-ChihPeng et al, Proposes a heterogeneous tracking model, referred to as HTM, to efficiently mine object moving patterns and track objects. Specifically, use a variable memory Markov model to exploit the dependencies among object movements. Furthermore, due to the hierarchical nature of HTM, multi-resolution object moving patterns are provided. The proposed HTM is able to accurately predict the movements of objects and thus reduces the energy consumption for object tracking. Simulation results show that HTM not only able to effectively mine object moving patterns but also save energy in tracking objects [1]. Jorge Huere Peña, and Maribel Yasmina Santos gave idea about moving objects have been collected in huge amounts due to the proliferation of mobile devices, which capture the position of objects over time. Studies about moving objects have been developed as a specific research area of Geographic Information Systems. Those systems are designed to process traditional, static or slowly changing, geospatial data. However, moving objects have inherent a dynamism that requires different approaches to data storage and analysis [2]. Jennifer Yick, Biswanath Mukherjee, and DipakGhosa developed target tracking in wireless sensor networks requires efficient coordination among sensor nodes. Existing methods have focused on tree-based collaboration, selective activation, and group clustering. This paper presents a prediction-based Adaptive algorithm for tracking mobile targets. They use adaptive Kalman filtering to predict the future location and velocity of the target. This location prediction is used to determine the active tracking region which corresponds to the set of sensors that needs to be "lighted". The velocity prediction is used to adaptively determine the size of the active tracking region and to modulate the sampling rate as well [4].Jiong Yang and Meng Hu introduce mobile objects have become ubiquitous in our everyday lives, ranging from cellular phones to sensors; therefore, analyzing and mining mobile data becomes an interesting problem with great practical importance. For instance, by finding trajectory patterns of the mobile clients, the mobile communication network can allocate resources more efficiently. However, due to the limited power of the mobile devices, we are only able to obtain the imprecise location of a mobile object at a given time. Sequential patterns are a popular data mining model. By applying the sequential pattern model on the set of imprecise trajectories of the mobile objects, they have uncover important information or further our understanding of the inherent characteristics of the mobile objects, e.g., constructing a classifier based on the discovered patterns or using the patterns to improve the accuracy of location prediction [5]. Yida Wang et al proposed A new approach to derive groupings of mobile users based on their movement data. They assume that the user movement data are collected by logging location data emitted from mobile devices tracking users. Formally define group pattern as a group of users that are within a distance threshold from one another for at least a minimum duration. To mine group patterns, first propose two algorithms, namely AGP and VG-growth. In our first set of experiments, it is shown when both the number of users and logging duration are large; AGP and VG-growth are inefficient for the mining group patterns of size two. Therefore propose a framework that summarizes user movement data before group pattern mining. In the second series of experiments, we show that the methods using location summarization reduce the mining overheads for group patterns of size two significantly. Then conclude that the cuboids based summarization methods give better performance when the summarized database size is small compared to the original movement database. In addition, we also evaluate the impact of parameters on the mining overhead [6].

III.DATA MINING AND TECHNIQUE

Data mining, *the extraction of hidden predictive information from large databases*, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by presentation tools typical of decision support systems.

A. Frequent Pattern Mining

Possible types of patterns: item sets, sequences, trees, and graphs.

• A core ingredient of the search is a canonical form of the type of pattern.

◦ Purpose: ensure that each possible pattern is processed at most once.

(Discard non-canonical code words, process only canonical ones.)

◦ It is desirable that the canonical form possesses the prefix property.

◦ Except for general graphs there exist canonical extension rules.

◦ For general graphs, restricted extensions allow to reduce the number of actual canonical form tests considerably.

• Frequent pattern mining algorithms prune with the Apriori property:

$\forall P : \forall S \supset P : s_D(P) < s_{min} \rightarrow s_D(S) < s_{min}$.That is: No super-pattern of an infrequent pattern is frequent.

• Additional filtering is important to single out the relevant patterns.

Moreover, a number of research works have been elaborated upon mining traversal patterns for various applications. For example, the FS and SS algorithms for mining path traversal patterns in a Web environment and an incremental algorithm to mine user moving patterns for data allocation in a mobile computing system. However, sequential patterns or path traversal patterns do not provide sufficient information for location prediction or clustering. The reasons are as follows: First, for sequential pattern mining or path traversal pattern mining extract frequent patterns of all objects, meaningful movement characteristics of individual objects may be ignored.

Second, a sequential pattern or traversal pattern carries no time information between consecutive items, so they cannot provide accurate information for location prediction when time is concerned. Third, sequential patterns are not full representative to individual trajectories because a sequential pattern does not contain the information about the number of times it occurs in each individual trajectory. To discover significant patterns for location prediction and to mine frequent trajectories several methods has been proposed, where consecutive items of a frequent trajectory are also adjacent in the original trajectory data.

B. Clustering

Data clustering is a method in which we make cluster of objects that are somehow similar in characteristics. The criterion for checking the similarity is implementation dependent. Clustering is often confused with classification, but there is some difference between the two. In classification the objects are assigned to pre defined classes, whereas in clustering the classes are also to be defined. Precisely, Data Clustering is a technique in which, the information that is logically similar is physically stored together. In order to increase the efficiency in the database systems the numbers of disk accesses are to be minimized. In clustering the objects of similar properties are placed in one class of objects and a single access to the disk makes the entire class available.

C. Hierarchical Agglomerative methods

The hierarchical agglomerative clustering methods are most commonly used. The construction of a hierarchical agglomerative classification can be achieved by the following general algorithm.

1. Find the 2 closest objects and merge them into a cluster
2. Find and merge the next two closest points, where a point is either an individual object or a cluster of objects.
3. If more than one cluster remains , return to step 2

Individual methods are characterized by the definition used for identification of the closest pair of points, and by the means used to describe the new cluster when two clusters are merged. There are some general approaches to implementation of this algorithm, these being stored matrix and stored data, are discussed below

- In the second matrix approach , an N*N matrix containing all pairwise distance values is first created, and updated as new clusters are formed. This approach has at least an O(n*n) time requirement, rising to $O(n^3)$ if a simple serial scan of dissimilarity matrix is used to identify the points which need to be fused in each agglomeration, a serious limitation for large N.
- The stored data approach required the recalculation of pairwise dissimilarity values for each of the N-1 agglomerations, and the O(N) space requirement is therefore achieved at the expense of an $O(N^3)$ time requirement.

IV. METHODOLOGY

A. *GMP Mine Algorithm*

The GMPMine algorithm is comprised of four steps. First, extract the movement patterns of each object from the location sequence. Second, construct a similarity graph in which similar objects are connected by an edge. Third, extract highly connected components to derive the group information. Fourth, we construct a group PST for each group in order to conserve the memory space. In the next section, describe the steps in detail.

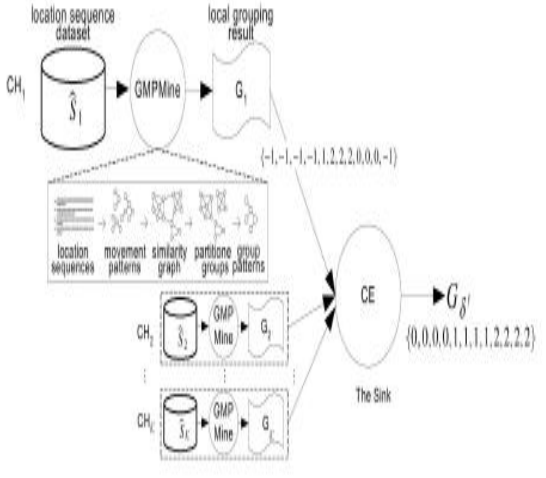


Fig.1. GMP Mine Technique

Location Sequence:

For the location sequence data set with N location sequences, we compute the movement patterns and generate N PSTs. For example, Fig. 6a shows the trajectories of three groups, each of which contains four objects. The mesh network is a two-layer structure with four clusters, each containing four sensors labeled as ∑= {a, b, c, d} The location sequence data set of CHs has defined. To facilitate collaborative data collection processing in object tracking sensor networks, cluster architectures are usually used to organize sensor nodes into clusters (with each cluster consisting of a cluster head and sensors).

Movement patterns

The goal is to propose an efficient data mining mechanism for deriving object moving patterns and utilize the object moving patterns for energy saving prediction-based object tracking sensor networks. To facilitate collaborative data collection processing in object tracking sensor networks, cluster architectures are usually used to organize sensor nodes into clusters.

Similarity Graph

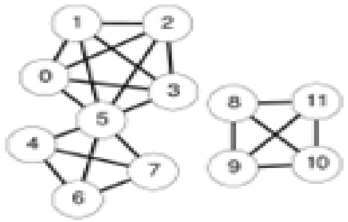


Fig.2. Similarity graph

Given: A data set X = {x1, x2, . . . ,xL}1. Create a graph G = (V,E),where V = {1, . . . , L} is the vertex set and E ⊆ V ×V represents the edge set. wijdenotes the edge-weight between sample I and sample j, where wij= exp(dist(xi, xj)/2σ2). Dist (・, ・) denotes a kind of distance. And then G is represented as an symmetric matrix A = (aij), which is defined as: aij= wijif (i, j) ∈ E 0 otherwise

Solve the quadratic program , which is solved by replicator equation: xi(t + 1) = xi(t) (Ax(t))I x(t)TAx(t) . A strict local solution is represented as x∗. Let Ω1 denote the vertex set {i : (x∗)i >0}and Ω2 denote the vertex set V \ Ω1.

Two sub-graphs G1 = (Ω1,E1) and G2 = (Ω2,E2) are obtained. G1 corresponds to a “similarity graph”.

The purpose of constructing similarity graph is to aggregate the temporal moving patterns into the memory in a compact form so that the mining of frequent patterns can be done efficiently. The main advantages of these techniques are 1) only one physical database scan is needed to mine all of the large patterns, and 2) the graph is compact so that the huge amount of data can be handled efficiently.

Partition Groups

For scalability, the cluster heads (referred to as CH) recursively form a hierarchical architecture for efficiently mining and queries. As such, a multi-resolution data collection mechanism is naturally supported by our HTM model, where the higher-level cluster heads will maintain coarse object moving patterns and the low-level cluster heads will have more precise object moving patterns. Based on the obtained object moving patterns, the cluster heads predict the object movements.

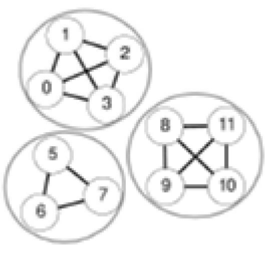


Fig.3. Result of partitioned group

In contrast to other clustering algorithms, such as K means, which partition objects into a predetermined number of groups, we consider the diversity of the number of groups and their sizes in the tracking applications. In addition, we trade off the grouping quality against the computation cost by adjusting the partition parameter of the CE algorithm.

Group patterns

In order to discover the temporal moving patterns, this study use time slot to uniformly segment the dimension of sequence for its real number characteristic.

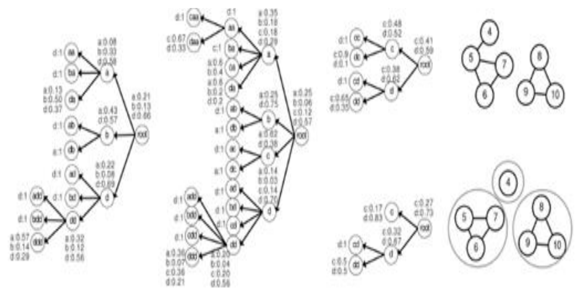


Fig.4. Identifying object from group pattern.

The above steps extract the group information and object movement patterns. In this step, we retain the most representative PST of a group of objects for storage efficiency. By integrating this kind of patterns into the prediction schemes, the error of predictions can be substantially reduced.

B. CE Algorithm

In the previous sections, each CH collects location data locally and generates group information with the proposed GMPMine algorithm. Since objects may not pass through all the clusters, and the group relationships of objects may vary in different areas, the local grouping results may be inconsistent. For example, if objects walk close together across a gap, it is reasonable to consider them a group. In contrast, objects scattered in grassland is hardly identified as a group. Furthermore, in the case where a group of objects move across the margin of a sensor cluster, the group relationship is difficult to determine. Therefore, the proposed CE algorithm to combine multiple local grouping results. The algorithm solves the inconsistency problem and improves the grouping quality.

The ensembling problem involves finding the partition of O that contains the most information about the local grouping results. Let C denote the ensemble of the local grouping results, represented as C = {$G_{0,}$ $G_{1,}$ G_{k-1}}where K denotes the ensemble size, i.e., the total number of CHs. The local grouping result Gi is obtained from CHi is a partition of O with mi disjoint groups, represented as Gi = {g^{i}_{0}, g^{i}_{1}; . . . ;

g^i_{m-1}.}

$$MI(G_i, G_j) = \sum_{a=0}^{m_i-1} \sum_{b=0}^{m_j-1} \widehat{P}(a,b) \log \frac{\widehat{P}(a,b)}{\widehat{P}(a) \times \widehat{P}(b)},$$

Where $\widehat{P}(a)$ denotes the probability function of Gi, defined as

$$\widehat{P}(a) = \frac{|g^i_a|}{|O|};$$

$$\widehat{P}(a,b) = \frac{|g^i_a \cap g^j_b|}{|O|}.$$

$$NMI(G_i, G_j) = \frac{\sum_{a=0}^{m_i-1} \sum_{b=0}^{m_j-1} \widehat{P}(a,b) \log \frac{\widehat{P}(a,b)}{\widehat{P}(a) \times \widehat{P}(b)}}{\sqrt{H(G_i) \times H(G_j)}},$$

Mutually informative. For a probable ensembling result G, the summation of NMIs of G and every Gi 2 C represents the amount of information that G contains with respect to C. Therefore, the ensembling result G0 that contains the most information about C is given by

$$G' = arg \max_{G \in \widetilde{G}} \sum_{i=0}^{K-1} NMI(G_i, G),$$

where˜G denotes all possible ensembling results. However, enumerating every G 2 ˜G in order to find the optimal ensembling result G0 is impractical, especially in resource constrained environments. To overcome this difficulty, it proposes the CE algorithm detailed. The algorithm leverages the information in C to generate the ensembling result δ € 0 , and trades off the grouping quality against the computation cost by adjusting the partition parameter D, i.e., a set of thresholds with values in the range [0,1]. A finer grained configuration of D achieves a better grouping quality, but the penalty is a higher computation cost. The proposed CE algorithm is comprised of three steps. First, it utilizes the Jaccard similarity coefficient to measure the similarity of each pair of objects. Second, it partitions the objects for every δ € D; and third, it employs NMI to optimize the ensembling result. In the following, it describes the three steps in detail.

δ	G_δ	$\sum NMI(G_i, G_\delta)$
0.1	{{0,1,2,3},{5,6,7},{8,9,10,11}}	2.322
0.2	{{0,1,2,3},{5,6,7},{8,9,10,11}}	2.322
0.3	{{0,1,2,3},{4,5,6,7},{8,9,10,11}}	2.636
0.4	{{0,1,2,3},{4,5,6,7},{8,9,10}}	2.401
0.5	{{0,1,2,3},{4,5,6,7},{8,9,10}}	2.401

Fig.5.Output of Ensembling

In the first step, the algorithm measures the similarity of each pair of objects to construct a similarity matrix based on the local grouping results, as shown in Lines 5-7. In tracking applications, the trajectories of moving objects span multiple sensor clusters, i.e., moving objects are present in partial sensor clusters and absent from the others. Hence, counting the absence of both objects in a pair makes no meaningful contribution to the similarity measurement. To address this issue, it utilizes the similarity coefficient to incorporate the opinions of the CHs whose opinions are known, when measuring the similarity between a pair of objects. The similarity coefficient, which compares the similarity between two binary vectors, is widely used in information retrieval applications where the importance of positive and negative opinions is asymmetric.

The metric is more objective and reasonable for our study than other measures, such as the simple matching coefficient and the overlap coefficient. Let $\pi_k(o_i)$ denote the mapping of the local group ID and o_i obtained from CHk; and let IkðoiÞ π 1 indicate that _kðoiÞ 6π _1. The group relationship of o_i and o_i in Gk, denoted by ckðo_i; ojÞ, represents that objects o_i and o_i belong to the same group in Gk;

To design an energy-efficient OTSN, we leverage thegroup information and the object movement patterns derived in Section 3. In a conventional update-based OTSN, sensors are assigned a tracking task, i.e., to update the sink with the location data of moving objects at every tracking interval. When a sensor detects an object of interest, it sends an update packet upward to the sink. Since the sensing area of sensors may overlap, a data aggregation step is required to avoid making unnecessary location updates. For example, at best, eight update packets passing through K layers to the sink are required for an object that moves from sensor s1 to s8.

C.Energy Efficient Object Tracking Sensor Network (Otsn)

WSNs designed for tracking the locations of moving objects are called OTSNs. Conserving energy in OTSNs is more difficult than in WSNs that monitor immobile phenomena, such as humidity, vibrations, or sound, because the target objects are moving. Hence, OTSNs require special consideration and designs to track moving objects efficiently.

```
Algorithm: Cluster Ensembling
Input: O = {o_0,o_1,...o_N}, C = {G_i |0 ≤ i<k }, D = {δ_i |0 ≤ i<d }
Output : G_δ'
init sum[]
init SM[][]
idx = 0
max = 0
/*building similarity matrix by Jaccard*/
for 0 ≤ i<N-1
    for i+1 ≤ j<N
        SM[i,j] = getS_ij(C)
/*select the partition with max ∑NMI(G_δ, G_i)*/
for 0 ≤ i<d
    Graph(V, E) = Convert2Graph(SM,δ_i)
    G_δ_i = HCS(Graph(V, E))
    sum[i] = ∑_{0≤j<k} NMI(G_j, G_δ_i)
    if sum[i] >max then
        max = sum[i]
        idx = i
G_δ' = G_δ_idx
return G_δ'
```

Fig.6.Cluster Ensembling Algorithm

V .RESULT AND DISCUSSION

A distributed mining algorithm identifies a group of objects with similar movement's patterns. It's used comprises a local mining phase and cluster ensembling phase Data collect from locally and generates the group of information with GMP Mine algorithm. CE algorithm comprised of three steps. First collect the similarly coefficient pair of objects.That means presently moving object is present in partial clusters and absent from others.

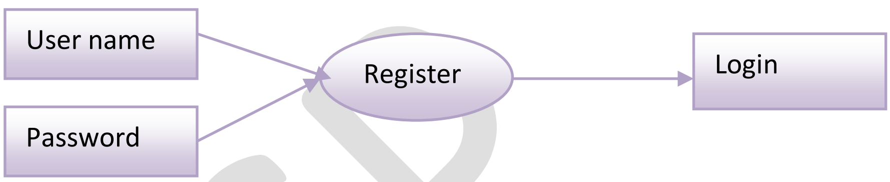
Second the coefficient object are same group or different group, is that simple match that coefficient underestimates the object's correlations. Final step find the normalized mutual information to select the ensembling result from the group of objects. In network data aggregation in improves the scalability and reduces the long-distance of communication demands and thus saves energy. To determine conduct three experiments to evaluate the performance of GMP-Mine and CE under different system conditions by varying the parameters in terms of number of objects, support threshold, and segmentation unit, respectively. Meanwhile, the effects of varying these system parameters were also studied.

δ	G_δ	$\sum NMI(G_\delta, G_i)$
0.1	{{0,1,2,3},{5,6,7},{8,9,10,11}}	2.322
0.2	{{0,1,2,3},{5,6,7},{8,9,10,11}}	2.322
0.3	{{0,1,2,3},{4,5,6,7},{8,9,10,11}}	2.636
0.4	{{0,1,2,3},{4,5,6,7},{8,9,10}}	2.401
0.5	{{0,1,2,3},{4,5,6,7},{8,9,10}}	2.401

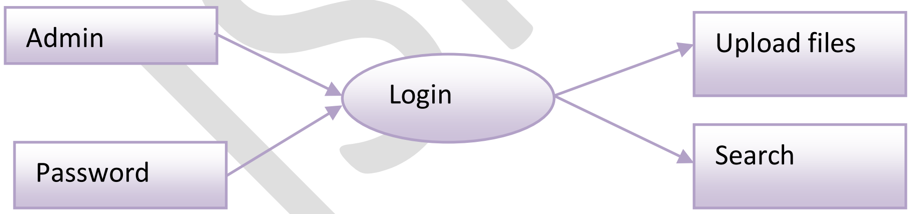
Fig.7. Result of CE algorithm

The following charts show the amounts of discovered patterns under different object numbers. As indicated, the relative variation is small but there exists an increasing trend. By having a check on the datasets, we find that some patterns with support slightly smaller than the specified threshold turn to be frequent when the number of objects increases. This study has been implemented and evaluated in .Net framework by creating a client server application on the network. The flow of the methodology implementation has been described in the following diagrams.

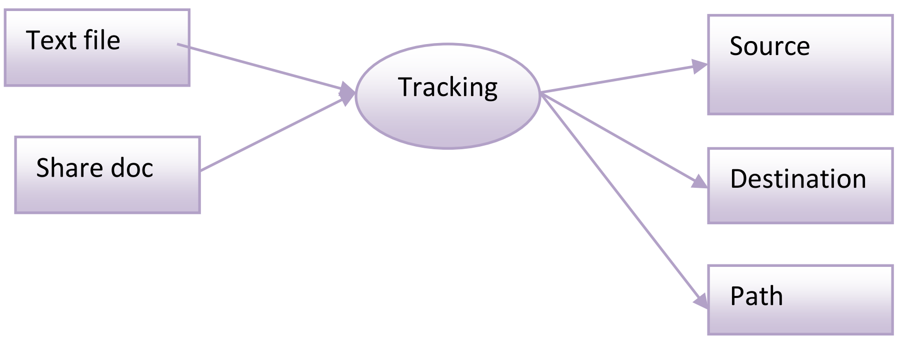
Level 0:



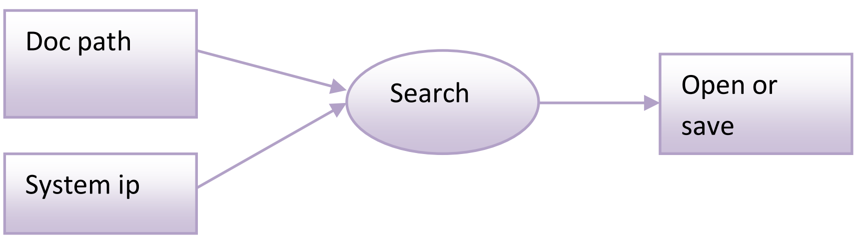
Level 1:



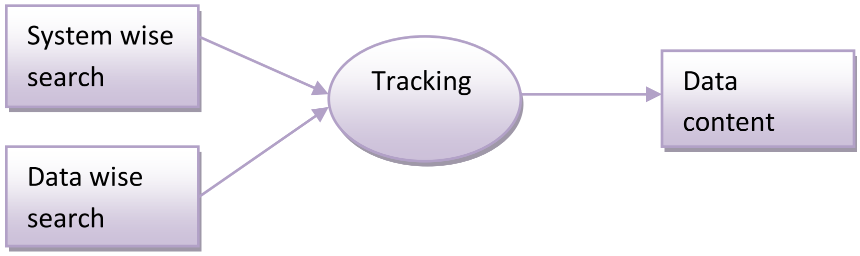
Level 2:



Level 3:



Level 4:



Level 5:

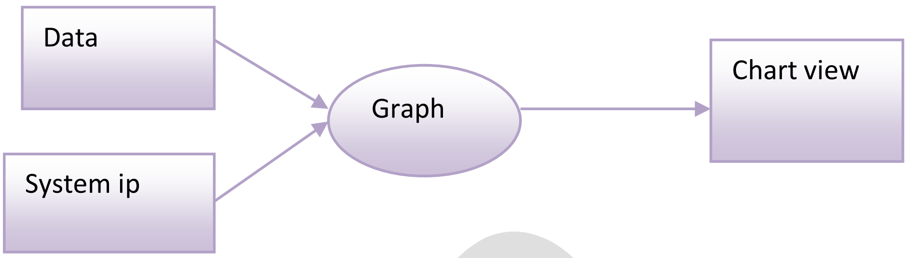


Fig.8. Methodology Implementation

TABLE 1.

An integrated log of temporal moving sequences.

Object Id	Moving Sequence
1	(a,1)(e,3)(c,5)(b,10)>
1	(a,3)(b,5)(c,7)(d,12)>
3	(a,1)(e,2)(c,5)(b,10)>
4	(f,0)(e,5)(b,13)>
5	(a,4)(b,6)(c,7)(d,12)>
6	(f,0)(a,4)(c,6)(d,10)>

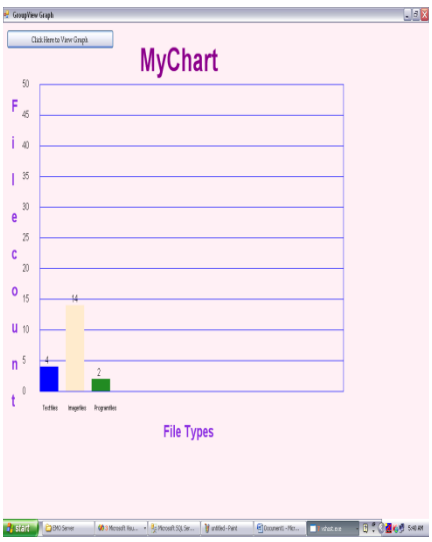


Fig.9. Chart for File Types

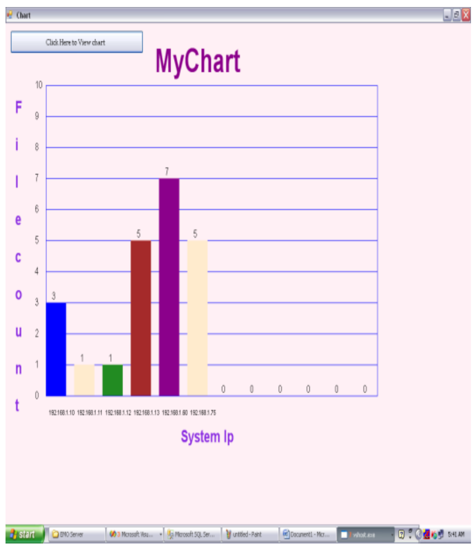


Fig.10. Overall File Count in the Network

VI. CONCLUSION AND FUTURE WORK

Exploit the characteristics of group movements to discover the information about groups of moving objects in an OTSN. In contrast to the centralized mining technique, mine the group information in a distributed manner. A novel mining algorithm, which consists of a local GMP Mine algorithm and a CE algorithm, to discover group information. Our algorithm mines object movement patterns as well as group information and the estimated group dispersion radius. Other than clustering trajectories, can apply the distributed clustering approach to heterogeneous and distributed sequential data sets, such as web logs or gene sequence. Using the mined object movement patterns and the group information, design an energy-efficient OTSN. The contribution of our approach is three fold:

1) It reduces energy consumption by allowing CHs to avoid sending the prediction-hit locations, because the locations can be recovered by the sink via the same prediction model;

2) It leverages group information in data aggregation to eliminate redundant update traffic; and

3) It sets the size of an SG adaptively to limit the amount of flooding traffic.

Our experimental results show that the proposed mining technique achieves good grouping quality.The results of experiments show that the proposed mining algorithm achieves good grouping quality, and the mining technique helps reduce the energy spending by reducing the amount of data to be transmitted.Furthermore, the proposed OTSN with PST prediction, group data aggregation, and in-network data aggregation significantly reduces energy consumption in terms of the transmission cost, especially in the case where moving objects have distinct group relationships. The proposed method has concentrated on the group of objects moving in a minor cluster way. In future this can be implemented in a grid environment to identify the last movement and location of a particular file. This will reduce the resource to search the file from the base source. Implementing the proposed concept with some enhancement in a grid environment will give better output.

REFERENCES

[1]. Hsiaso-Ping Tsai,DeNian Yang, and Ming-Syan Chen,(2011).Mining Group Movement Patterns for Tracking Moving Objects Efficiently:IEEE transactions on knowledge and Engineering,vol.23.

[2]. L. Chen, M. Tamer O ¨ zsu, and V. Oria, (2005). Robust and Fast Similarity Search for Moving Object Trajectories: Proc. ACM SIGMOD, pp. 491-502.

[3]. M. Morzy,(2007) . Mining Frequent Trajectories of Moving Objects for Location Prediction: Proc.FifthInt Conf. Machine Learning and Data Mining in Pattern Recognition, pp. 667-680.

[4]. M. Nanni and D. Pedreschi,(2006).Time-Focused Clustering of Trajectories of Moving Objects:J. Intelligent Information Systems, vol. 27, no. 3, pp. 267-289.

[5]. W.-C. Peng, Y.-Z.Ko, and W.-C. Lee, (2006).On Mining Moving Patterns for Object Tracking Sensor Networks: Proc. Seventh Int'l Conf. Mobile Data Management, p. 41.

[6]. Y. Wang, E.-P.Lim, and S.-Y. Hwang (2006). Efficient Mining of Group Patterns from User Movement Data: Data Knowledge Eng., vol. 57, no. 3, pp. 240-282.