© 2018 IJSRSET | Volume 4 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

# Ship Route Mining and Path Optimization

**Shabana M A, Prof. Elizabeth Isaac, Dr. Surekha Mariam Varghese**
Mar Athanasius College of Engineering, Kothamangalam, Kerala, India

**ABSTRACT**

Mining trajectory data has been attracting significant interest in the last years. By analyzing trajectory data, we are able to discover the movement behavior and location aware knowledge, and then develop many interesting applications such as movement behavior discovery, location prediction, traffic analysis, and so on. This paper provides a study on different analysis done on maritime data and different methods for finding shortest path from a specific source to a specified destination in a trajectory dataset. The shortest path (SP) problem concerns with finding the shortest path from a specific origin to a specified destination in a given network while minimizing the total cost associated with the path.

**Keywords:** Ship Route Mining, Path Optimization, Behavior Discovery, Shortest Path, Computational Complexity, Real-Time Communications, ANN, TREAD

## I. INTRODUCTION

Maritime transportation represents approximately 90% of global trade by volume, placing safety and security challenges as a high priority for nations across the globe. Maritime surveillance data are collected at different scales and are increasingly used to achieve higher levels of situational awareness.

Trajectory data mining is a challenge task because of the trajectory data is available with uncertainty. The uncertainty may be produced by the inaccuracy of positioning device and asynchronous sampling. With location uncertainty, a moving objects movement may not exactly repeat the same trajectory even the object has the similar movement behavior with others. Furthermore, discovering the valuable knowledge from maritime trajectory is made even more difficult due to the maritime area is a free moving space. Unlike the vehicles movements are constrained by road networks, there is no such a sea route for ships to follow in maritime area. Obviously, the maritime trajectory is moving free and the data is more complex than the trajectories moving along the road network. A ships movement may not exactly repeat the same trajectory even the ship has the similar movement behaviour with others. Thus, mining the maritime routes where ships frequently navigate from collected ship trajectories is a challenging problem.

The shortest path (SP) problem concerns with finding the shortest path from a specific origin to a specified destination in a given network while minimizing the total cost associated with the path. This problem has widespread applications. Some important applications of the SP problem include vehicle routing in transportation systems, traffic routing in communication networks and path planning in robotic systems.

The SP problem has been investigated extensively. The well-known algorithms for solving this problem include the Bellman dynamic programming algorithm for directed networks, the Dijkstra labelling algorithm and Bellman Ford successive approximation algorithm for networks with non negative cost coefficients. These traditional algorithms have major shortcomings; firstly, they are not suitable for networks with negative weights of the edges. Secondly, the algorithms search only for the shortest route, but they cannot determine any other similar or non-similar short routes. Thirdly, they exhibit high computational complexity for real-time communications.

The proposed system provides a framework of ship route mining thereby analyse the huge volume of historical data of ship trajectories further the movement behaviour and find the shortest path between source and destination.

## II. LITERATURE SURVEY

In 1998, Masaharu Munetomo, Yoshiaki Takai, and Yoshiharu Sato introduced 'A Migration Scheme for the Genetic Adaptive Routing Algorithm'. Path genetic operators (path mutation and path crossover) are designed for the network routing algorithms to generate alternative routes in routing tables. The operators are constrained to topology of the target network. The path mutation operator mutates a route to create an alternative route. The path crossover operator exchanges sub routes between a pair of routes. These operators create alternative routes in a routing table to Find an optimal route which minimizes communication latency by employing together selection operators based on fitness values which are calculated from communication latency. It can also balance the load of links by distributing packets among the alternative routes.

Artificial neural networks (ANN) have been examined to solve the SP problem relying on their parallel architecture to provide a fast solution. In 2001 Filipe Arajo, Bernardete Ribeiro, and Lus Rodrigues introduces 'A Neural Network for Shortest Path Computation'. It presents a new neural network to solve the shortest path problem for inter network routing. The proposed solution extends the traditional single-layer recurrent Hopfield architecture introducing a two-layer architecture that automatically guarantees an entire set of constraints held by any valid solution to the shortest path problem.

A new Hopfield NN that aims to improve the reliability of the solutions, where reliability stands for successful and valid convergence. To achieve this, a new architecture, named dependent variables (DVs), which consists of a two-layer Hopfield NN is presented. This architecture automatically guarantees an entire class of restrictions, considerably increasing the reliability of the method. At the same time, the number of neurons is equal to the number of arcs in the graph rather than being equal to the squared number of nodes. Advantage: Improved solutions with a bigger number of nodes. Disadvantage: Poorer behavior in certain classes of graphs. The ANN approach has several limitations. These include the complexity of the hardware which increases considerably with increasing number of network nodes; at the same time, the reliability of the solution decreases. Secondly, they are less adaptable to topological changes in the network graph including the cost of the arcs. Thirdly, the ANNs do not consider sub-optimal paths. Among other approaches for this problem, the powerful evolutionary programming techniques have considerable potential to be investigated in the pursuit for more efficient algorithms because the SP problem is basically an optimal search problem.

In 2013 Pallotta G., Vespe M., and Bryan K introduced 'Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction'. It presents an unsupervised and incremental learning approach to the extraction of maritime movement patterns. The proposed methodology is called TREAD, which stands for Traffic Route Extraction and Anomaly Detection. TREAD converts raw data, i.e. ship position reports from different tracking platforms, into information that can be used to support decisions concerning the safety and security of shipping. The paper shows that understanding past maritime traffic patterns is a fundamental step towards Maritime Situation

Awareness applications, in particular, to classify and predict activities. TREAD is a basis for automatically detecting anomalies, using past ship tracks and traffic patterns as an input to a Decision Support System. TREAD builds a statistical model in which the traffic knowledge is extracted from the data by means of ship objects, created and constantly updated based on the AIS position data stream. The changes in the state vectors, i.e. the course and speed, of many ship objects generate a series of spatial events that are clustered around waypoints used to reconstruct the traffic routes. Tracks that substantially deviate from other vessel paths on the same route are considered outliers and eliminated from the analysis. The result of the data analysis is fed into the last module of TREAD which provides the anomaly detection and route prediction functions.

The problem of reconstructing shipping lanes in a particular area is presented by Fernandez Arguedas in 2014. The proposed algorithm automatically produces a network of maritime shipping lanes extracted from historical vessel positioning data, by detecting the entry and exit 7 points in the ocean region and the so called breakpoints which divide a ship track into shorter segments. The proposed applications are track reconstruction in cases of tracking gaps, destination prediction, and detection of anomalous behavior.

In 2014 'RouteMiner: Mining Ship Routes from a Massive Maritime Trajectories' was introduced by Yu-Ting We, Chien-Hsiang La, Po-Ruey Lei and Wen-Chih Peng. The objective of the system is to discover the ships movement patterns hidden in their historical trajectories, and then detect the ship route. More specifically, it not only discover the movement pattern but also define and detect the movement area of ship route in a free moving space. it propose RouteMiner to provide a framework for ship route mining. Given a set of ship trajectory in a monitoring area, RouteMiner explore the movement pattern from massive ship trajectories in a free moving space. Then, ship routes are defined and detected based on those behavioral pattern. Finally, the system generates a set of ship routes. It consists of four modules: pre-processing, pattern-based trajectory clustering, ship route detection, and visualization. The raw maritime trajectory data is recorded by GPS system. Each trajectory is represented by a sequence of spatial-temporal points,

$Ti = (x_1; y_1; t_1), (x_2; y_2; t_2), \ldots\ldots, (x_n; y_n; t_n)$.

Where n is the total number of points. The pre-processing module transforms each ship trajectory into a region based trajectory. The module includes two steps, frequent region detection and trajectory transformation. In frequent region detection, the GPS trajectory data is mapped into grid system, and then a cell is detected as a frequent region rj if the number of trajectories passed the cell has satisfied the user-defined minimum support threshold MinTs. Based on the discovered frequent regions, each GPS trajectory is transformed into the region-based trajectory with corresponding frequent regions. The points that are not in frequent regions will be regarded as noise. The pattern-based trajectory clustering module applies Prefixspan algorithm to mine the frequent sequential patterns (FSP).The ship route detection module provides a solution to detect a ship route from each trajectory cluster. More specifically, the task of the module is to detect a movement channel of a ship route followed by a cluster of ship trajectories those having similar behavior. Given a pattern-based trajectory cluster, the cluster provides a sequence of frequent regions and a set of the trajectories with similar movement behavior in those regions. It detect a ship route by performing statistical analysis of the trajectory points in each region sequentially. Finally, the visualization module visualizes the discovered ship routes on the top of Google Maps. The visualization provides the operators a better understanding of ships movement from ship trajectory data.

Later in 2017, 'Distributed Document Clustering Analysis Based on a Hybrid Method' was introduced by J.E. Judith, J. Jayakumari. It provides an overview of how Particle Swarm Optimization can be implemented in MapReduce Framework. PSO is used to take advantage of its global search ability to provide optimal centroids which aids in generating more compact clusters with improved accuracy. In order to support data intensive distributed applications, an open source implementation based on Hadoop is used for processing of large datasets. MapReduce is a functional programming model for distributed processing over several machines. The important idea behind MapReduce framework is to map the datasets into a group of <key, value> pairs, and then reduce all pairs with the same key. A map function is performed by each machine which takes a part of the input and maps it to <key, value> pairs. This is then send to a machine which applies the reduce function. The reduce function combines the result for further processing. The outputs from the reduce function are then fed into the appropriate map function to begin the next round of processing.

In this work, a hybrid PSOKMeans (PKMeans) distributed document clustering method is formulated for better speedup and accuracy of document clustering. The input to the Map function includes the document dataset stored in HDFS. The document dataset is split and <key, value> pairs are generated for each individual text documents in the dataset. The key is the doc ID and the value is the terms in the document <docID, term>. The Map function determines the term frequency of each term in the document.

The list of terms frequency of each term in the document collection are given as input to reduce function which combines the term frequency to form a Document Term Matrix. Document term matrix is normalized by including the inverse document frequency along with the term frequency for each entry in the matrix. PSO each particles location in the multidimensional problem space represents a solution for the problem. A number of candidate clustering solutions for document collection are considered as a swarm. A different problem solution is generated when the location of the particle changes. The problem solution moves through the search space by following the best particles. It moves through the search space looking for a personal best position. The velocity of each particle determines the convergence to an optimal solution. The inertia weight, particle personal experience and global experience will influence the movement of each particle in the problem space. A fitness value is assigned by the objective function to each particle which has to be optimized based on the position. The fitness function is defined according to the property of the problem where it is applied. The quality of clustering solution depends on the fitness function defined.

It represents the average distance between document vector and cluster centroids. Map function splits the documents and <key, value> pairs are generated for each individual text document in the dataset. In the first module the MapReduce job is used for generating optimal centroids using PSO. The map function evaluates the fitness of each particle in the swarm. All the information about the particle such as particleID, Cluster vector(C), Velocity vector(V), Personal Best Value(PB), Global Best Value(GB) are determined. The particleID represents the key and the corresponding content as the value. The collection of <key, value> pairs contained in _les is referred to as blocks. The particle swarm is retrieved from the distributed storage. For each particleID the map function extracts centroids vectors and calculates the average distance between centroids vector and document vector. It returns a new fitness value based on global best values to the reduce function. The reduce function aggregates the values with the same key and updates the particle position and velocity. The reduce function emits the global best centroids as the optimal centroids to be stored in HDFS for the next module.

Particle swarm optimization (PSO) is a population based optimization technique inspired by the social behavior of bird flock. The algorithmic flow in PSO starts with a population of particles whose positions, that represent the potential solutions for the studied problem, and velocities are randomly initialized in the search space. The search for optimal position (solution) is performed by updating the particle velocities, hence positions, in each iteration/generation in a specific manner as follows.

In every iteration, the fitness of each particles position is determined by some defined fitness measure and the velocity of each particle is updated by keeping track of two best positions. The first one is the best position (solution) a particle has traversed so far. This value is called pBest. Another best value is the best position (solution) that any neighbor of a particle has traversed so far. This best value is a neighborhood best and is called nBest. When a particle takes the whole population as its neighborhood, the neighborhood best becomes the global best and is accordingly called gBest. A particles velocity and position are updated as follows.

$$V_{id} = v_{id} + c_1 r_1 (b_{id} * x_{id}) + c_2 r_2 (b^n_{id} * x_{id}) \quad (1)$$

$i = 1; 2; :::;N_s$

$d = 1; 2; :::;D$

$$X_{id} = x_{id} + v_{id} \quad (2)$$

Where $c_1$ and $c_2$ are positive constants, called acceleration coefficients, Ns is the total number of particles in the swarm, D is the dimension of problem search space, i.e., number of parameters of the function being optimized, $r_1$ and $r_2$ are two independently generated random numbers in the range [0,1] and n represents the index of the best particle in the neighborhood of a particle. The other vectors are defined as: $x_i = [x_{i1}, x_{i2}, . . ., x_{iD}]$ is the position of $i^{th}$ particle; $v_i = [v_{i1}; v_{i2}; . . . ; v_{iD}]$ is the velocity of $i^{th}$ particle; $b_i = [b_{i1}, b_{i2}, . . ., b_{iD}]$ is the best position of the $i^{th}$ particle ( pBesti), and $b^n_i = [ b^n_{i1}, b^n_{i2}, . . . ; b^n_{iD} ]$ is the best position found by the neighborhood of the particle i (nBesti).

Equation (1) calculates a new velocity for each particle based on its previous velocity, the particles position at which the best possible fitness has been achieved so far, and the neighbors best position achieved. Equation (2) updates each particles position in the solution hyperspace.c1 and c2 are two learning factors, which control the influence of pBest and nBest on the search process.

The pseudo-codes for general algorithmic flow of PSO are listed bellow.

```
Initialize the position and velocity of each particle
in the population randomly.
Calculate fitness value of each particle.
Calculate pBest and nBest for each particle.
Do
Update velocity of each particle.
Update position of each particle.
Calculate fitness value of each particle.
Update pBest for each particle if its current fitness
value is better than pBest.
Update nBest for each particle, i.e, choose the
particle with the best fitness value among all the
neighbors as the nBest for a specific neighborhood
topology.
While termination criterion is not attained.
```

## III. CONCLUSION

Particle Swarm Optimization will take more time to process networks having large number of links. So in the future work the PSO is going to implement in Hadoop, in order to increase the throughput and reduce the time needed to execute.

## IV. REFERENCES


[1]. Ammar W. Mohemmed, Nirod Chandra Sahoo, Solving shortest path problem using particle swarm optimization, Applied Soft Computing, 2008.

[2]. C.W. Ahn, R.S. Ramakrishna, A genetic algorithm for shortest path routing problem and the sizing of populations, IEEE Trans. Evol. Comput. 2002.

[3]. F. Araujo, B. Ribeiro, L. Rodrigues, A neural network for shortest path computation, IEEE Trans. Neural Netw, 2001.

[4]. Giuliana Pallotta, Michele Vespe and Karna Bryan, Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction, Entropy 2013.

[5]. J.E. Judith, J. Jayakumari, Distributed Document Clustering Analysis Based on a Hybrid Method, China Communications , February 2017.

[6]. M.K. Ali, F. Kamoun, Neural networks for shortest path computation and routing in computer networks, IEEE Trans. Neural Netw. 1993.

[7]. M. Munemoto, Y. Takai, Y. Sato, A migration scheme for the genetic adaptive routing algorithm, in: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 1998.

[8]. Yu-Ting Wen, Chien-Hsiang Lai, Po-Ruey Lei and Wen-Chih Peng, RouteMiner: Mining Ship Routes from a Massive Maritime Trajectories, IEEE 15th International Conference on Mobile Data Management, 2014.

[9]. Ying-Tung Hsiao, Cheng-Long Chnang, and Cheng-Chih Chien, Ant Colony Optimization for Best Path Planning, lnternational Symposium on Communications and Information Technologies, 2004.

[10]. 0La Spezia, Contextual Anomalous Destination Detection For Maritime Surveillance, Maritime Knowledge Discovery and Anomaly Detection Workshop, 2016.

[11]. Mark P. Wachowiak, Member, IEEE, Mitchell C. Timson, and David J. DuVal, Adaptive Particle Swarm Optimization with Heterogeneous Multicore Parallelism and GPU Acceleration, IEEE Transactions, 2016.

[12]. Qisheng Cai, Teng Long, Zhu Wang, Yonglu Wen, Jiaxun Kou, Multiple paths planning for UAVs using particle swarm optimization with sequential niche technique, IEEE Transactions, 2016.

[13]. Giuliana Pallotta, Michele Vespe and Karna Bryan, Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction, entropy ISSN 1099-4300, 2013.

[14]. Weather Routing, Principles Of Weather Routing, National Imagery And Mapping Agency.