ISSN:2348-2079

Volume-6 Issue-2

International Journal of Intellectual Advancements and Research in Engineering Computations

A Scalable Approach for Data-Driven Taxi Ride Sharing

[1]R.Kavitha, [2]B.Keerthana, [3]N. Gokulpriya, [4]R. Aravinthrajan

Department of Computer Science and Engineering,Velalar College of Engineering and Technology

[1] Assistant Professor, [2-4]UG Students

ABSTRACT:

Traffic congestion is a main issue all over the world. For that a proposed system called carpool is introduced. This system that will increase the number of occupation seats by decreasing the number of empty seats. In carpooling, drivers share their vehicles with one or more additional riders whose destinations are similar. It is good to traffic congestion, but also an environmentally sound transportation method. The genetic Algorithm is used for matching of which seekers goes with which driver. A Genetic algorithm is used for large number of users, so it gives proper match. This algorithm can because of exit from local optimal and with high percent produce the problem's optimal answer. The probability of results show that mooted algorithm has better performance regards as other clustering algorithms specially in two indices, the carefulness of clustering and the quality of clustering.

Keywords: Taxi ride sharing, Simulation, Shortest-path index, Scalability, Urban computing, Carpooling, Real-Time Tracking, Rating, Comment.

I.INTRODUCTION:

Several studies monitoring the environmental performance of different transportation modes in different countries and in accordance to the growing of transportation demand. The authors, as citizens of EU member States, reports statistics, analyses, strategies and policies, and targets which refers to the European transportation system and issues. In particular, the European Environmental Agency (EEA) and the Environmental Reporting Mechanism (EEA TERM) introduce a framework to integrate indicators useful to monitor both transport and environment issues and identifying rules and tools to improve the performance of a transport system. The analysis, control, and optimization of transportation modes and usage in accordance with climate change and environmental issues, are based on several focuses: focus on road transport, rail transport, air transport, waterborne transport, non-motorized transport (e.g., cycling and walking), land use and planning, etc. It is an interesting investigation on what car-use reduction measures are perceived by households to be feasible to reduce car driving. The study proposed in this paper deals with road transport, whose main two topics are:

1) Transport for the trading sectors and moving of materials, known as freight transport, i.e., logistics & distribution;

2) Transport for non-trading sectors and moving of passengers.

PROBLEM DEFINITION:

There is an acute problem of traffic on roads these days, and the increasing fuel prices add to the misery of daily users of personal vehicles. And use of vehicles causes pollution which has its adverse effects. Car sharing is a solution but issues like security and trust come into picture. Can this problem be solved? Solution to this problem is mobile based Carpool system. The Carpool system would enable its user a safe and secure way to share cars. This could include both short daily journeys such as going to workplace within the city and long inter-city trips.

II. LITERATURE REVIEW:

K-MEANS V/S K-MEDOIDS:

A COMPARATIVE STUDY:

Clustering is one of the most important research areas in the field of data mining. In simple words, clustering is a division of data into different groups. Data are grouped into clusters in such a way that data of the same group are similar and those in other groups are dissimilar. It aims to minimize intra-class similarity while to maximize interclass dissimilarity. Clustering is an unsupervised learning technique. Clustering is useful to obtain interesting patterns and structures from a large set of data. Clustering can be applied in many areas, such as marketing studies, DNA analysis, data planning, text mining, and web documents classification. Large datasets with many attributes make the task of clustering complex. Many methods have been developed to deal with these problems. In this work, two well-knownpartitioning-based methods k-means and kmedoids – are compared. The study given here explores the behavior of these two methods. [1]

FAST AND ACCURATE K-MEANS FOR LARGE DATASETS:

Clustering is a popular problem with many applications. We consider the k-means problem in the situation where the data is too large to be stored in main memory and must be accessed sequentially, such as from a disk, and where we must use as little memory as possible. Our algorithm is based on recent theoretical results, with significant improvements to make it practical. Our approach greatly simplifies a recently developed algorithm, both in design and in analysis, and eliminates large constant factors in the approximation guarantee, the memory requirements, and the running time. We then incorporate approximate nearest passenger search to compute k-means in o(nk) (where n is the number of data points; note that computing the cost, given a solution, takes 8(nk) time). We show that our algorithm compares favorably to existing algorithms - both theoretically and experimentally, thus providing state-of-the-art performance in both theory and practice. [2]

COMPUTATION OF INITIAL MODES FOR K-MODES CLUSTERING ALGORITHM USING EVIDENCE ACCUMULATION:

Clustering accuracy of partitional clustering algorithm for categorical data primarily depends upon the choice of initial data points (modes) to instigate the clustering process. Traditionally initial modes are chosen randomly. As a consequence of that, the clustering results cannot be generated and repeated consistently. This work we present an approach to compute initial modes for K-mode clustering algorithm to cluster categorical data sets. Here, we utilize the idea of Evidence Accumulation for combining the results of multiple clusterings. Initially n F – dimensional data is decomposed into large number of compact clusters; the K-modes algorithm performs this decomposition, with several clusterings obtained by N random initializations of the K medoides algorithm. The modes that obtained from every run of random initializations are stored in a Mode-Pool, PN. The objective is to investigate the contribution of those data objects/patterns that are less vulnerable to the choice of random selection of modes and to choose the most diverse set of modes from the available Mode-Pool that can be utilized as initial modes for the Kmode clustering algorithm. Experimentally, we found that by this method we get initial modes that are similar to the desired modes and gives consistent and better clustering results with less variance of clustering error than the traditional method of choosing random modes.[3]

WEIGHTED QUANTUM PARTICLE ANT SYSTEM OPTIMIZATION TO ASSOCIATION RULE MINING AND PSO TO CLUSTERING:

In the area of association rule mining (ARM), the most major algorithms is Apriori algorithm. In the existing Apriori algorithm minimum support and confidence are determined subjectively or through trial and error method so, the algorithm lacks the objectiveness and efficiency. To improve the efficiency of association rules, Particle Ant system Optimization (PSO) algorithm is projected, which gives feasible threshold values for minimum support and confidence. This PSO algorithm, initially it looks for the optimum fitness value of each particle and finds their corresponding support and confidence as minimum threshold values. The difficulty of PSO algorithm is that, it guesses that the items have the same implication without taking into account of their weight/attributes within a transaction or within the whole item space. To overcome this drawback, this work proposes a weighted quantum particle ant system optimization algorithm (WQPSO) with weighted mean best position according to fitness values of the particles. WQPSO algorithm provides faster local convergence, fallout in better balance between the global and local searching of the algorithm, so it generates good performance. The proposed WQPSO algorithm is experienced with several benchmark functions and compared with standard PSO. The experimental result shows the supremacy of WQPSO and it is verified by applying the FoodMart2000 database of Microsoft SQL Server 2000. Likewise, in clustering, there are many unsupervised clustering algorithms have been developed one such algorithm is K-means which is

simple and straightforward. The main drawback of the K-means algorithm is that, the result is sensitive to the selection of the initial cluster centroids and may converge to the local optima. This is solved by PSO as it performs globalized search and produces clusters with high intra class similarity.[4]

III.SYSTEM ANALYSIS:

EXISTING SYSTEM:

Many carpool service systems have been proposed which can be divided into two broad categories based on their features. The first of these comprises systems which are web-based and which transmit carpool information to an online community platform. One such system is Carpool Global which supplies an interfacing service for willing drivers and passengers. These Systems do not include Location Display of user (GIS) & not real-time. The second category of carpool service systems provides digital GIS supports that match requests via location information an example system of this category is the Share YourRide platform by which users can readily submit carpool requests and offers via its map-based interface. Addition, Share YourRide supplies a GIS-based routing service.

DRAWBACKS:

- This system has limited functions in situations requiring instant service because it cannot support the use of Global Positioning System (GPS) handheld devices which provide pertinent information regarding user location.
- Many carpool systems have not been developed to decrease lessen traffic congestion.
- Of these, many systems supply simple carpooling functions including the option to send requests for a specified date and time, and search for applicable users.
- In addition, several systems feature a digital GIS mapping ability by which to provide a visual tool with accurate location information to users.
- Unfortunately, these systems are neither efficient nor convenient for users who need real-time carpool matches.
- Scalability limitations in terms of network bandwidth and server loading.
- Cannot able maintain best shortest path to the destination.
- Fewer data sharing accuracy.
- High data transfer time.
- Congestion may occur.
- Data transmission scheduling not made by data traffic.

PROPOSED SYSTEM:

Our proposed system incorporates the Shortest path computation with pool share and communication technology with GIS to create a carpool service which is operable in real time. Subsequently, users can instantly submit carpool requests to the intelligent carpool system which reflect their current locations via the use of smart, handheld, communication devices which feature GPS capabilities. The system will use the carpool matching algorithm to generate and return match results within a short amount of time. GENETIC Based Carpooling, which is an application of car-sharing (also called lift-sharing or ride-sharing) in which drivers (alone-riders) who are travelling to work alone can ask for fellow passengers through our application.

In those who use public-transport system to go to work daily can use this application to find drivers who are travelling to the same destination and willing to share ride. This will not only get rid of the extra journey time of passengers but will also help environment by reducing pollution and traffic on roads. This social networking application is also called fare-sharing and time sharing as you are sharing both with fellow travelers.

MERITS:

- The server periodically updates the travel/ Passenger details and times on these paths based on the latest traffic, and reports the current best path to the corresponding user.
- Efficiently maintains the index for live traffic circumstances.
- Low communication time between a sender and receiver.
- Data transfer rate is high with minimal time.
- High security and safety journey for a passenger with minimum fare amount.

IV. MODULE DESCRIPTION:

Initial Population Management Module:

In the initialization procedure, each passenger is randomly chosen and assigned to a driver in the assignment layer of a chromosome. The chosen passenger is marked to prevent the passenger from being assigned twice. Fig 1.1.

Profile Fitness Evaluation Module:

To find the quality of the profile in the population the fitness function is used to determine the travel cost for each driver. To calculate the fitness value, we need to find the most efficient route for picking-up and dropping off passengers for each corresponding driver. Fig 1.2.

Selection Passenger Module:
The first phase involves sorting the profile into a descending order according to their fitness values, and selecting those with the highest values in the population. This gives with the highest fitness values from one generation to the next level. Fig 1.2.

Fixing the Car for Pooling:

After the optimal profile have been selected, the chromosome crossover procedure is utilized to recombine the profile of selected parents to simulate the natural process of evolution. E. Mutation It is used to change the allocation of the passengers mutually. Fig 1.3.

Online Route API Module:

This module is to compute the shortest route between two points on a road network, based on live traffic. It has the latest road network G with live travel time information. Mobile User. Using a mobile device (smartphone), the user can acquire his current geo-location q and issue queries to a location-based driver. In this module, we consider range and KNN queries based on live traffic. Fig 1.3.

SYSTEM REQUIREMENTS:

HARDWARE REQUIREMENTS:

System : PentiumIV 2.4 GHz.
Hard Disk : 40 GB.
Monitor : 15 VGA Color.
Mouse : Logitech.
Ram : 1 GB DDR2 RAM.

S O F T W A R E R E Q U I R E M E N T S :

Operating system : Windows 7.
Coding Language: Java
IDE : Netbeans 8.1

Database : MYSQL

DATA FLOW DIAGRAM:

LEVEL 0:

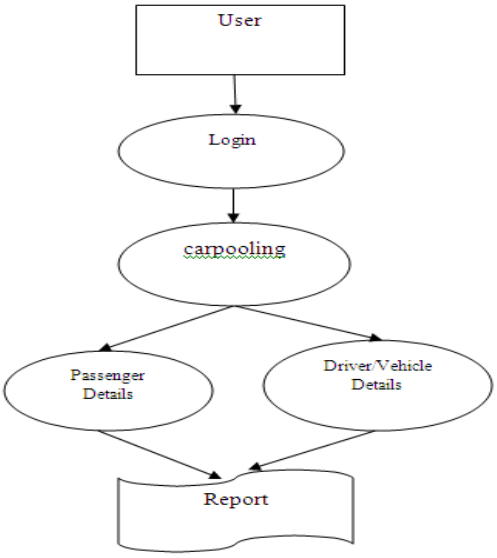

Fig 1.1

LEVEL 1:

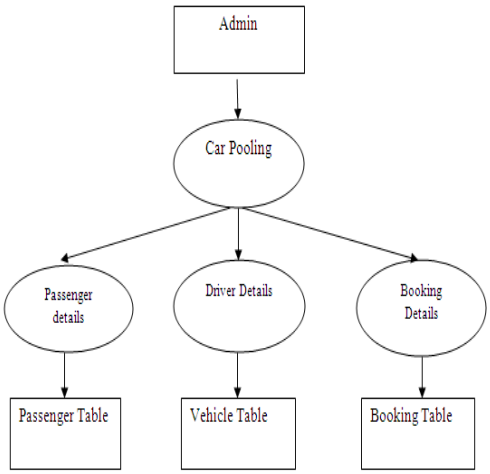

Fig 1.2

Level 2:

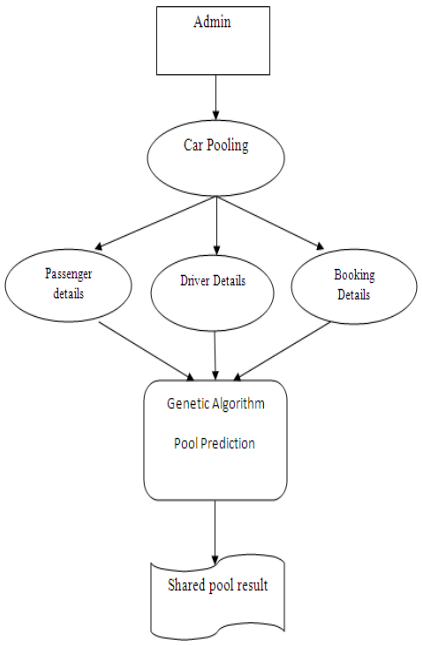

Fig 1.3

SYSTEM FLOW DIAGRAM:

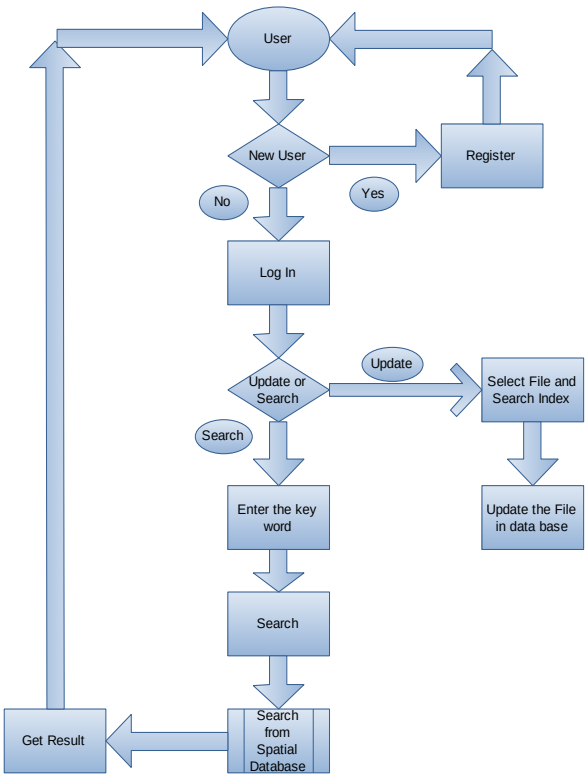

Fig 1.4

EXPERIMENTAL SETUP:

During the configuration step, a list of scenarios is generated: each scenario represents the configuration of a population. During the experiments, we have changed some parameters of the population to evaluate the effects on the KPIs; these parameters are listed in Table fig 1.5,1.6. The performed experiments have been conducted by selecting an area of interest in the city of Cagliari (centered at: 39.23, 9.14) and with an area (A) of about 64 km2, which is where the users can operate.

This area is of interest for this study since the majority of users mainly operate inside this boundary. Furthermore, this area is representative of medium-small cities with numerous residential areas, commercial sites, factories and historic neighborhoods within its metropolitan boundaries. In the performed emulations, we also refer to the timeout T, which ranges from 1 to 30 min. For simplicity, we assume the same timeout T for each rider and each driver.

The setup step consists of the generation of each member of the population for a single run. During this step, the total population N is divided into Ld drivers and Lp passengers. The emulator assigns each user departure and destination locations chosen randomly and uniformly in the selected area, with the following two constraints:

(1) Both locations fall on a source to destination

(2) It is actually, possible to travel from the departure to the destination, i.e., a path exists between these points.

Passenger success rate	Path duration Shared Pooling
5	10
10	20
15	30
20	40
25	50
30	60

Fig1.5

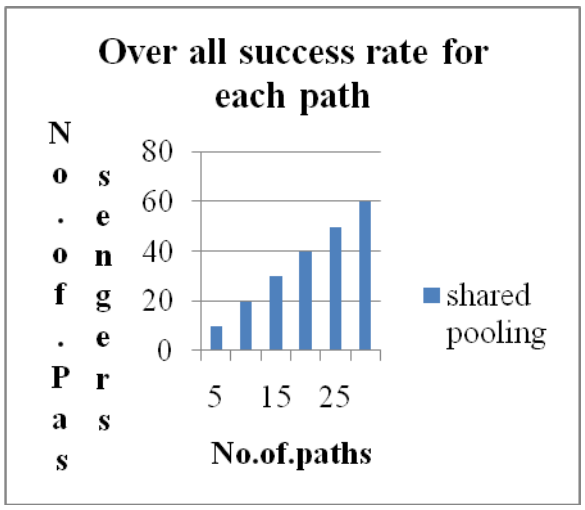

Fig 1.6

SYSTEM IMPLEMENTATION:

After having the user acceptance of the new system developed, the implementation phase begins. Implementation is the stage of a project during which theory is turned into practice. In this phase, all the programs of the system are loaded onto the user's computer. After loading the system, training of the users starts. Main topics of such types of training are:

- How to execute the package
- How to enter the data
- How to process the data (processing details)
- How to take out the reports

After the users are trained about the computerized system, manual working has to shift from manual to computerized working. The following two strategies are followed for running the system:

Parallel run: In this strategy, for a certain defined period, both the systems i.e., computerized and manual are executed in parallel. This strategy is helpful because of the following:

- Manual results can be compared with the results of the computerized system.
- Failure of the computerized system at the early stage does not affect the working of the organization, because the manual system continues to work.

Carpooling is also seen as a more environmentally friendly and sustainable way to travel as sharing journeys reduces carbon emissions, traffic congestion on the roads, and the need for parking spaces. Authorities often encourage carpooling, especially during high pollution periods and high fuel prices. We intent on making a Java based application that will enable to let people know if vehicles are available for carpooling in their desired path they can sign in for it. This will enable people using this application to share expense, not worry about hiring a cab and making new connections. People having this application on their cell phone can easily carpool with unacquainted people without worrying about security.

V.CONCLUSION & FUTUREWORK:

Carpooling system is an effort to reduce consumption of fuel, our most important non-renewable resource and traffic congestion on roads by encouraging people to use car sharing. Therefore, it is an environment-friendly social application also helps people to reduce their journey time. Future scope would include

Chat Option:-In future we are also going to implement chatting functionality for the convenience of our users and which will also be more user-friendly.

Inter-personal Issue and smoking issues Users (drivers) can mention if they are comfortable with passengers who smoke. It is difficult to adjust with smokers if other passengers are non-smokers. Punctuality maintains for Passengers should directly come to their meeting point after office as other fellow passengers will be waiting for them.

REFERENCES:

1. Chih-Ping Wei, Yen-Hsien Lee and Che-Ming Hsu, **Empirical Comparison of Fast Clustering**. Algorithms for Large Data Sets,2000, Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000.

2. DW van der Merwe, AP Engelbrecht, **Data Clustering using Particle Ant System Optimization,** Conference Publications, 2003.

3. L.E. Agustın-Blas, S. Salcedo-Sanz, S. Jimenez-Fernandez, L. Carro-Calvo, J. Del Ser, J.A. Portilla-Figueras, **A new grouping genetic algorithm for clustering problems,** Elsevier ad hoc 2012.

4. Li-Yeh Chuang, Chih-Jen Hsiao b, Cheng-Hong Yang, **Chaotic particle ant system optimization for data clustering,** Elsevier ad-hoc, 2011.

5. P.S. Shelokar, V.K. Jayaraman, B.D. Kulkarni,' **An ant colony approach for clustering,** Elsevier, 2003.

6. Sandeep Rana, Sanjay Jasola, Rajesh Kumar, **A hybrid sequential approach for data clustering using K-Means and particle ant system optimization algorithm"** ,2010, International Journal of Engineering, Science and Technology, 2010.

7.Shalini S Singh, N C Chauhan, **K-means v/s K-medoids**: A Comparative Study, National Conference on Recent Trends in Engineering & Technology,2011.

8.Suresh Chandra Satapathy, Gunanidhi Pradhan, SabyasachiPattnaik, JVR Murthy, PVGD Prasad Reddy, **Performance Comparisons of PSO based Clustering**,2009, InterJRI Computer Science and Networking

9.Taher Niknam, Babak Amiri, **An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis**, Elsevier Engineering Applications of Artificial Intelligence, 2010.

10. Zhexue Huang, **Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values,**1998.