ISSN No. 0976-5697

Volume 8, No. 1, Jan-Feb 2017

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

Taxi Fleet Management System

Sujata Khedkar, Kunal Tolani, Amey Barapatre, Vineet Karkhanis and Nirbhay Pherwani

VESIT, University of Mumbai, Mumbai, India

Abstract: A taxi fleet management system is presented .The system consists of Clustering , Neuro fuzzy systems and Particle Swarm Optimization methodologies. The proposed system aims at maximizing revenue of cabs as individual entities and the cab aggregator simultaneously. Clustering of pick up requests is carried out using a variant of DBSCAN which uses Delaunay triangulation to recognise fare hotspots. Neuro Fuzzy system is used to evaluate the eligibility of taxis to contest for these hotspots .The Neuro Fuzzy System is trained using Particle Swarm Optimization method. Intelligent swarming of taxis according to their eligibilities for the hotspots is performed to maximize revenue of both cab aggregators and cabs.

Keywords: PSO, TSK Model, Taxi Fleet Management, Neuro Fuzzy Systems, Clustering, Fleet Management , Particle Swarm Optimization , Swarm Intelligence.

INTRODUCTION

Right from searching for a cab to booking one online the taxi world has undergone a drastic transformation. The focus of this change is the customer and follows the basic business rule 'customer is the king of the market'. Considering customers in the spotlight many cab aggregator firms have blossomed. These firms focus on spreading their footprint by increasing the number of cabs. in doing so their expenditure increases drastically.

These firms fail to route the taxis, under their command, to customer hotspots and fail to ensure cab drivers have completed the pre-determined number of rides or have accumulated the pre-decided amount of money through rides. so to conclude, in order to ensure the customers are benefited the firms and taxi drivers are incurring losses.

To solve this problem, a game-changing strategy was thought of which lead to the inception of this idea. Powered by the exhaustive dataset of NYC Taxis, it's beneficial to use data mining and swarming techniques to earn more profit while using available resources. Collection and implementation of real data for constant improvisation of fleet management are recommended. Factors like availability, reliability and punctuality of cabs and user demands are considered while working on the proposed solution.

Shortcomings of Existing Systems

In the current system the cab driver has to rely on his past experience to go to locations where the probability of acquiring a fare is high. Even when information about requests and their fares is available these aren't used in helping the driver making an informed decision. The customer request is shared with all drivers in the vicinity of the request. The one who accepts the request swiftly gets the fare. The firm fails to take into account the expense at which the driver completes a request as the customer needs are the final ends for the aggregator.

Even though sufficient information regarding the pickup and drop-off locations is available, it isn't put to use. The firm fails to consider which driver receives the request as long the request is being satisfied. So if a driver has completed more than requisite number of requests, his probability of acquiring a new fare is equal to that of a driver who hasn't completed his share of requests. There is no equitable distribution of fares. The firms fail to see effective revenue generation via taxi assignment to requests.

1. LITERATURE SURVEY

The proposed systems uses mainly two methodologies to serve its purposes viz. Clustering and Neuro Fuzzy System optimised by Particle Swarm Optimization.Literature reviews are thus categorised accordingly.

1.1 Clustering

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters). Clustering algorithms can be divided into partitioning, hierarchical, locality-based and grid-based. Partitioning based clustering creates partitions and divides the dataset into k clusters. This technique follows a simple and easy approach to arrange a given information set through a specific number of groups (assume 'k' bunches). The fundamental thought is to define 'k' centers, one for every cluster. These focuses ought to be set smartly due to various area causes distinctive result. The use of exclusive assignment - If there are two highly overlapping data then k-means will not be able to resolve that there are two clusters and it doesn't work with non-globular clusters. Grid based clustering are affected by noisy data, and doesn't not take into account the arbitrary sized clusters which will be found in our dataset. DBSCAN clustering does not require one to specify the number of clusters that may be formed, and only depends on the number of points in the locality (minimum Minpts points must be formed to form a cluster.) The points inside a cluster maybe distributed randomly, and the clusters can be odd shaped or arbitrarily shaped, as per requirement. It is resistant to noise, and can be further clustered in to distinguish sparse data.

[1] used an improved Density Peaks Clustering (DPC) approach to form clusters of demand hotspots. DPC algorithm states that demand hotspot clusters are usually points with higher densities compared to other points in the neighborhood.However, to properly calculate the density of points, they need to calculate distances between all points, which increases complexity anddefeats the purpose of using

DPC. So they have proposed an improved DPC, where instead of processing all data points, a density image of data points is generated, from which redundant points are removed, as well points very close to each other. Finally, image filtering is done to remove noise.
There can be a couple methods for deciding spatial patterns in mobility data- [2] is GPS based following that perceives the way taken, or source goal combines that may lead us to the grouping of data based on the origin and destination points. [3] utilizes a mix of shared nearest neighbors and Delaunay Triangulation to shape groups of closest neighbors. A Novel separation is computed between focuses utilizing shared nearest neighbors (SNN) and weights are doled out to figure out which group they have a place with. To guarantee spatial contiguity and accomplish productivity, a Delaunay triangulation (DT) is developed for all areas and is utilized to frame a grouping calculation.

[4] overcomes the drawbacks of vanilla K means clustering it proposes a K-means clustering algorithm based on the modified shuffled frog leaping algorithm and simulated annealing algorithm.The system updates the velocity of a particle in the state space using the original , local maxima , and global maxima so far in the state space.Simulated annealing is used to increase the convergence rate of the algorithm using the fitness variance as its measure.The algorithm thus is not only more accurate but also more efficient because of reduced iterations and computations.

2.2 Neuro Fuzzy Systems and Particle Swarm Optimization

Neuro Fuzzy System(NFS) , employs the ability of both Neural Networks and Fuzzy Sets . A typical neuro fuzzy system has the following layers, input layer, fuzzification layer , inference layer , and output layer . The fuzzification layer calculates the fuzzified input i.e membership of an input .The inference layer creates the antecedents of the rules from the fuzzification layer and give the output as the firing strength of these rules. The output layer takes creates the weighted output from the firing strengths of the rules to give the final output.They have a variety of applications in control systems.

PSO is a class of evolutionary algorithm. This algorithm searches the global maximum by group intelligence and communication between such groups among the state space to find the optimum solution. PSO can be applied to NFS to learn the antecedents of the rules.

[5] has discussed the performance of PSO in NFS training for any image processing function approximation. The system proposed in the paper uses Takagi-Sugeno-Kang model (TSK model) which uses a singleton output membership function instead of a fuzzy set, as used in Mamdani model inference system. PSO method was applied for learning process. For the PSO method, parameter setting was set empirically on the basis of the dimensioning and perceived difficulty of the problem. The NFS-PSO model had lesser number of rules and greater efficiency than other similar models.

[6] concentrates on how particles change their stages at an arbitrary rate, how these variables influence arrangement hubs and proposes an approach. It talks about how to handle permutation set, the issues that may rise and how to conquer these issues. The approach proposed is connected to N-queens issue and the consequence of its effectiveness is compared with other techniques.

2. PROPOSED SYSTEM

3.1. Neuro Fuzzy System , With PSO :

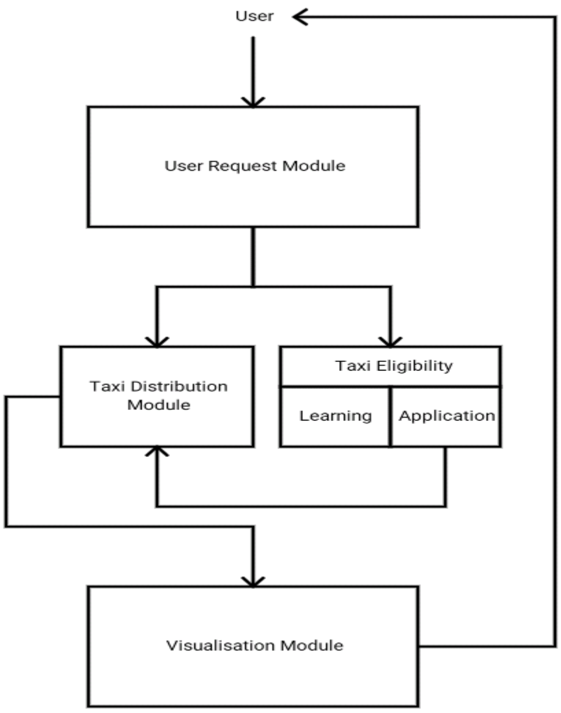
Neuro Fuzzy System is proposed to find the eligibility of the taxis for a particular ride request.A linear consequent Takagi-Sugeno-Kang model is proposed. A fuzzy rule R(r) form using the linear consequent TSK model turns to be as below: R(r) : If x1 is A1j AND .... AND xn is Anj then y is ax1 +bx2 + c. In the TSK model rule consequent either can be a linear function of the inputs or constant .In the proposed system the output of membership of a rule is assumed to be linear as the parameters of the linear function will be optimised during learning which eventually can reduce the linear function to a zero-order equation.The membership functions are gaussian functions around the cluster centers for the set.Thus these cluster centers are also a part of the parameters of the system.
PSO is used to optimize the parameters of the system . Initially random finite set of points and initial velocities for the points will be chosen in the state space defined by the constraints on the parameters.According to the fitness of these points , global maximum point and local maximum points the velocity of the points will be updated.Convergence is reached when reasonable variance is reached on the fitness function.

3.2 Clustering

The presence of a large number of noise points, no guarantee regarding the convex shape of the clusters and no prior information about the number of clusters ‘k‘ makes partitional clustering algorithm such as k-means unsuitable for clustering this type of dataset.The proposal made is to use a combination of Delaunay Triangulation and DBSCAN method to reach the optimum clusters of request data. Delaunay Triangulation is being used for two purposes here. First being to ensure spatial contiguity and second to get an understanding of specific regions from where DBSCAN can pick up and identify the coarse data and then in the second iteration take into consideration the sparse data as well.The task as summarized above is to use Delaunay Triangulation as a method to obtain specific regions which help in returning coarse clusters taking into consideration the major hotspots and minimize the data points which are not in the coarse clusters as some clusters may cover a very broad region due to the dense and wide distribution of data points in some areas. After this step the system uses a density based clustering scheme DBSCAN to find cluster structure in the sampled data points and later extend it to the rest of the dataset using nearest prototype rule as sparse data should also be considered no matter how unimportant it is.Also as a matter of fact DBSCAN defines clusters as areas of higher density than the remainder of the data set.It does not require the number of clusters, k to seek as an input and can find clusters of any arbitrary shape.

3.3. Swarming

The hotspots recognised from the clustering system and the eligibility memberships from the NFS are used to create a system of particles where each particle represents a taxi . The fitness of the system is determined by the expected revenue at current position of the particles . The particles are assigned new directions according to the local maxima , global maxima and current position.The effect of the updation parameters can be varied according to company policies



.

Fig 1: Conceptual System Diagram

3. CONCLUSION

The system proposed overcomes the computation drawbacks of DBSCAN algorithm by applying Delaunay Triangulation .Clustering is thus achieved efficiently without losing the sparse data.The nonlinear nature of the membership function of eligibility of the taxis is traced effectively by the use linear consequent TSK model. PSO optimizes the TSK model offline without getting trapped at local maxima on contrary to traditional optimization algorithms like backpropagation. Swarming of cabs according to set parameters ensures that aggregator has control over profitability of the end users of the system.

4. REFERENCES

[1] Dongchang Liu ,Shih-Fen Cheng & Yi-ping Yang,2015,Density Peaks Clustering Approach for Discovering Demand Hot Spots in City-scale Taxi Fleet Dataset .

[2] Diansheng Guo,Xi Zhu,Hai Jin,Peng Gao,Clio Andris , 2012,Discovering Spatial Patterns in Origin-Destination Mobility Data.

[3] D. T. Lee and B. J. Schachter , Received July 1978; revised February 1980 ,Two Algorithms for Constructing a Delaunay Triangulation 1.

[4] Rongzhi Zhang, Chenchen Liu, Shining Liang, Xueni Zhang, Wenyu Dong, Wanli Zuo , 2016,An Improved Clustering Algorithm Based on Multi-swarm Intelligence.

[5] Manel Elloumi ,Mohamed Krid , Dorra Sellami Masmoudi, 2015,Neuro-Fuzzy System based on Particle Swarm Optimization Algorithm for image denoising application.

[6] Xiaohui Hu, Russell C. Eberhart and Yuhui Shi , 2003 ,Swarm Intelligence for Permutation Optimization: A Case Study of n-Queens Problem.