Location-Based Service to reduce the waiting time for taxi services

Felipe A. L. Reis[1] **Marconi A. Pereira**[2] **Paulo E. M. Almeida**[1]

[1]CEFET-MG - Centro Federal de Educação Tecnológica de Minas Gerais, Brazil
[2]UFSJ - Universidade Federal de São João Del-Rei, Brazil

Abstract

One solution to increase the operational efficiency of the taxis systems is using of location-based services (LBS) in order to choose the best vehicle available to meet a request. In this work, it is evaluated three different algorithms to define the best taxi to attend the call, when the position of the vehicle and the passenger are known. The tests, performed in a stationary system, shows that the proposed LBS algorithms presents a reduction in the waiting time and a decrease of the distance to travel, when compared to broadcasting methods - currently used to choose the taxi for service.

Keywords: Location-based services (LBS), urban taxi transportation, real-time routing, assignment problems

1. Introduction

Taxi is an agil and comfort category of transportation very popular in urban areas. The taxi transportation systems is influenced by different factors, that change its availability. Among them, it is possible to mention:

- Events such as concerts, parties and conferences wich generate strong demand at a specific location, usually concentrated in a small period of time (closing time);
- Rainy days motivate people, who usually move by walk, subway or bus, to use taxi. This situation generates a scattered demand, that is, there is not a concentration of users in a specific location;
- Rush times (peak of traffic);
- Beginning or end of holidays, when users choose to travel to the airports, bus and train stations.

The operational efficiency of taxi service, generally, is unsatisfactory. The reasons are related to the lack of methodology to serve request and the organization of the systems service: phone call schedules, random search methods to take customers (taxi drivers and passenger look for each other) and use of taxi ranks [1].

In search of effective improve of taxi services, it is necessary a balance between taxi availability, number of requests by region and the knowledge of detailed information about preferences, traffic flow and route access. The join of these information can produce models that describes the behavior of taxi services. However, due their own characteristics - such as the number of taxis, autonomous stochatic demand in different locations and the absence of patterned behavior of drivers - creating those models of operation is not feasible, even for average size representations [1].

An alternative solution for this scenario is using trackers in the vehicles that send location information of drivers and passengers. The location-based services can increase the operational efficiency: the knowledge of geographical location about taxis and passenger allows to choose the best vehicle for each request, improving the global result of the system.

The use of global location technology with taxis request was already studied by [2] and [3]. In these papers, the request of taxi services is through telephone exchanges, in which the client informs his own location and the operator identifies the closest vehicle to serve the request. However, these papers do not describe the algorithms used to choose the best taxi.

This paper evaluates the gain, at the overall average time, of the GPS-based algorithms compared to the broadcasting method - mainly used today, to order a taxi. This work, also analyzes how the change in the offer of taxi affects the waiting time for this service. This paper measures the processing time of the proposed algorithms.

2. Background

The taxis systems is everywhere: big cities, medium-sized cities and even the small cities around the world has an infrastructure of these services. The taxi system meets a service demand of the native population and the tourist, that visit or works on these places.

At the most of cities around the world, it is possible to know that taxi services waste a large portion of time - about 50% - waiting for passengers [1]. Because of the high rate of waste time in taxis systems, there are a large number of studies that intend to improve the efficiency of the services, without increase costs [1].

The use of geographical location, obtained by trackers, has been recently incorporated into dispatch systems [3]. From the knowledge of the absolute position of a passenger, it is possible to estimate which services can interest this possible consumer and improve the offer and quality of products and services provided [4].

There are different techniques, used by taxi drivers, to find customers: from random search methods up to GPS-based methods, in which is known the location of taxi drivers and passengers. The knowledge of the passengers and the taxi positions allows a big advantage to anothers methods: it is possible to choose the best vehicle in the fleet to meet every request more efficiently. According to [3], the GPS-based methods will be used, in the future, more efficiently at vehicles dispatch systems to meet the demands of taxi.

In general, location systems monitors the position of taxis by vehicle tracking. [2] and [3] describes real systems that store the absolute position of taxis and, after a call to a dispatch center, the system defines the best vehicle for each request, according to the client position.

Because of the increasing use of smartphones and tablets over the world, and de possible maintenance of this trend for the next years, it is possible to build systems that use integrated GSP / A-GPS trackers to locate vehicles and passengers. By the requests coming from these devices, the knowledge about taxi drivers and clients become innate and possible to provide taxi services with better quality.

3. Algorithms

3.1. GPS-based algorithm with lower estimated time for service

One of the options to choose the best taxi to meet a request is the development of an algorithm that choose always the best vehicle whose time attendance estimation is the lowest as possible.

To choose the best taxi, we have to estimate each travel time of each taxi, based on the real route between the vehicle and the passenger. There is, however, a problem in this simplified solution: the cost to evaluate all taxis make the request processing so slow, due the complexity of routing algorithms.

In order to decrease the response time, a pre-processing of the linear euclidean distance between the taxi and the client can provide decrease in the processing time, without affect the quality of the algorithm. In this approach, it is thought that exists a linearity between the distance and time to attendance.

Using the euclidean distance, it is possible to classify the taxis by its distance to the client. The first step is select vehicles in a range of distance and order then by ascending distance. If there are no taxis in a range defined at beginning, the algorithm proceed as follows:

1. Increase the range of distance to find taxis;
   (a) If at least one taxi was found in the new range, the algorithm processes the vehicles found;
   (b) Otherwise, the range is increase again;
      i. If at least one taxi was found in the new range, the algorithm processes the vehicles found;
      ii. Otherwise, the algorithm concludes that there are no vehicles near and puts the customer in a queue;

It is possible to better understand the algorithm with the Figure 1. The taxis are searched firstly in the lighter region of concentric circles. If they are not found, they are searched in the next region and, finally, if necessary, in the darker region (at the gray scale). If none taxi was found, the algorithm puts the customer in queue. Using the

Fig. 1: Method to filter taxis, by their distance to the customer.

circles concept defined above, it is possible to limit the maximal distance that a taxi will be evaluated to serve a request. The maximal distance limit aim to prevent that taxis so far from the customers can be choosen by the algoritm to meet the request. Far vehicles will have to travel big distances and, it is possible that the profit do not pay the costs of drive to the passenger. Futhermore, the clients will wait for long time - that is not a good thing.

If the system, using data filter, find a great number of taxis that fits for the service, the system limit a number of near N drivers, further reducing processing cost. Among these drivers, the system will estimate each travel time to the client and define the best taxi for the request.

In this paper, the measurement of time request estimation process is made by the Google Maps API, which gives the lowest actual route.

3.2. GPS-based algorithm with Euclidean distance

An alternative algorithm to the measurement of time request estimation process is just use the Euclidean distance. It can be used as a criterion for choosing the best taxi.

Analyzing this method, however, in an actual situation, the Euclidean distance may prove inefficient, since the nearest taxi, according to this criterion, it may take longer to travel to a customer than a taxi slightly farther, but to present a different route. It is possible to see this situation in the Figure 2. In the last figure, it is possible

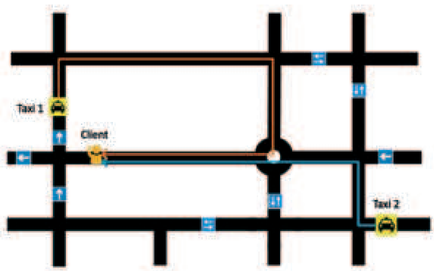

Fig. 2: Example of routes between taxis and a customer.

to see clearly that Taxi 1 is closer to Client 1 than Taxi 2, using euclidean distance algorithm. Nevertheless, Taxi 2 will travel a lower distance than Taxi 1 to serve the client.

Despite of the problems discussed above when using only the Euclidean distance to define the best taxi, this solution has a considerably lower computational cost, and can deliver good results in the average case.

Simiarly to the first method with the lowest estimated time of service, it is defined a maximal distance range of taxis that can serve the request.

3.3. Broadcasting method

The broadcasting method is currently used for choice of the taxi driver responsible for meeting a client after a call to the call center. In that, the operator communicates, by radio, to the taxi drivers, the location of the passenger. One of the taxi drivers near the customer confirms that is available and can serve the request. Then it is applied for to meet that.

The broadcasting method does not guarantee that the near taxi will be chosen nor the system is optimized to meet the demands: the driver who first answers the demand is the one that will meet the request. It is possible to conclude that the choice of the driver is carried out in a way of random, since the driver that first response to the operator is chosen.

To simulate this method, an algorithm was developed. In it, after a request, is selected a set of taxi drivers near to the client (using the limit of distance like the GPS-based algorithms) and is raffled a taxi.

4. Tests and Results

The type of simulation engine designed to test the algorithms described in the last sections is classified as a Discrete Event Simulation.

Tests were performed for all algorithms under same conditions. Each method defined its best taxi, according to your own criteria. The time displacement of each taxi was computed in each of the algorithms.

To calculate the average time service was used Google Maps API in order to estimate the travel time between taxis and the customers. This estimate was defined as time to start the service. The simulator does not evaluate, at the tests, traffic informations.

4.1. Taxi distribution

To define the distribution of service demands in the area used as a representation of a city, was built a placement generator for taxis and passengers.

Here, it is estimated that the probability of a event in a peripheral area of the city is less than the probability of a event in the central region. This distribution represents a situation where a small place or a single place receives greater amounts of request. In large cities, at the most of times, several places concentrate the major number of requests, unlike a single area. However, these usual behavior is complex to be modeled, especially when related to stochastic demand for taxi services [1].

The taxis, likewise the demands, were not distributed uniform. Using the same algorithm, they occupy, at most, the central region. The peripheral area has less taxis availables and less requests. The taxi distribution, compared to the midpoint can be seen in the figure 3.

4.2. Service flow

To represent the dinamic of real systems, it was built a simulator that changes the position of taxis through the city. As the simulation was built on discrete events, these movements occur only after processing a request.

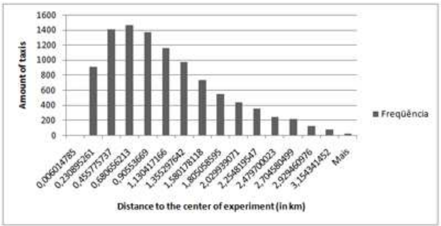

Fig. 3: Histogram with the distribution of taxi drivers in relation to the central point.

The simulator maintains the approximate reason between the number of occupied taxis and the amount of vehicles. The tests keeps on a stationary state, without tendency to increase or decrease supply in relation to the number of passengers.

The maintenance of taxis in a stationary state guarantee that the waiting time keeps equal. This analogy derives from the relationship proposed by [5], where the waiting time is directly related to amount of spare time of taxi drivers.

Another characteristic of the simulator is related to amendment of states of taxis. In a real scenario, taxis starts and release jobs all over the city. To represent this in a discrete way, the simulator, after a request processing, modifies the availability of a few taxi drivers.

It is necessary to inform that elasticity of demand was ignored by the simulator. In practice, there are variations of amounts of requests due to days of weeks, rush hours, climate conditions, among others. However, during the simulation tests, conducted under homogeneous conditions, there are not changes of the supply of services.

As a summary of the procedure used to perform tests, it is possible to describe in the following sequence:

1. The system defines a number N of taxis and puts them randomly in the city;
2. Half of the amount of taxis is set as occupied;
3. Taxi drivers moves randomly throughout the city in discrete events without changing significantly their current position;
4. Some of the taxis changes yours status of "Occupied" to "Free" and vice versa, simulating the beginning/end of a service;
5. The algorithm simulates a passenger request;
6. Each vehicle dispatching system choose the best taxi and the time until the service starts is measure for each algorithm;
7. If the number of request planned for the test is not reached yet, the system goes back to the step 3. Else, the system sums all the measured times for each algorithm;

4.3. Tests conditions

The city of Belo Horizonte (Brazil) was chosen as the place where the tests was performed, as a simulation of real conditions. Due to the existence of largest urbanized area in the city center, with a low number of large green areas and a extensive road network, the taxis and requests was distributed over the central area and vicinity (defined as the peripheral area).

Besides the actual geographic location where tests were run, adjustments were made to the model in order to simulate conditions close to the average in the city of Belo Horizonte. The next values defines the scenario of the tests:

- Average load factor: 50%;
- Maximal numbers of taxis evaluated in circles next to the customer: 7;
- "Circle" search radius: 0,7km, 1,2km e 1,5km;
- Probability of moving of a taxi: 90%;
- Maximum probability of change the taxi status ("Free" to "Occupied" ou vice versa): 10%.

In the first test, using the real average conditions in Belo Horizonte, seted up a representation in which 300 taxis were spread over an area of approximately 34.84 km^2. The average rate of taxis per km^2 was 8.61, or a taxi every 0.12 km^2.

In a second test, it was decided to restrict the number of taxis, in order to analyze the influence of the availability of services. For that, was defined a amount of 200 taxis in the same area (average rate of 5.74 taxis per km^2).

4.4. Results

The results of the tests described in this section were subjected to analysis of variance (One-way ANOVA) followed by Tukey test for multiple comparisons averages. It was adopted a confidence interval of 95%, with differences being considered significant when $P<0.05$. Statistical analysis was done by Graph Pad Prism software (version 5.01).

To reduce the name of the algorithms at the graphics on this section, according to the type of test, was used the naming standard described at table 1.

4.5. Test I

The first result is the customers average waiting time to a taxi service. The graph on the figure 4 shows averages times for each algorithm. The

Algorithm	Time	Distance	Processing
GPS-based algorithm with lower estimated time	GPS-T	GPS-D	GPS-P
GPS-based algorithm with Euclidean distance	EUC-T	EUC-D	EUC-P
Broadcasting Method	BRC-T	BRC-D	BRC-P

Table 1: Reduced naming pattern of algorithms.

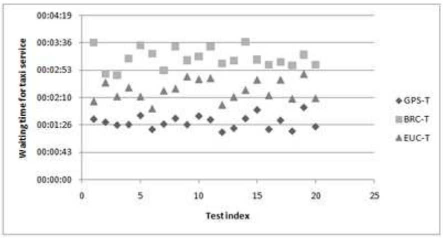

Fig. 4: Customer's average time to service.

GPS-based algorithm with lower estimated time for service had lower values, when compared to the GPS-based algorithm with Euclidean distance and the Broadcasting method. Therefore it can be inferred, according to tests carried out, that this algorithm has lower average waiting time to customers. These results were obtained with 95% confidence and shows that there is a statistically significant, favoring the GPS algorithm with Lower Estimate Time for Service.

The results corresponding to the average distance to travel by taxis can be seen in figure 5. It is possible to

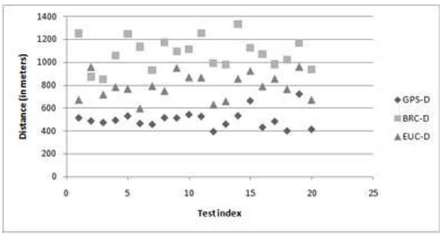

Fig. 5: Taxi average distance to serve a request.

see that the average distance traveled by taxi drivers in the GPS algorithm is smaller than the distance algorithms at EUC and BRC algorithms.

As the proposed tests do not consider traffic information, and according to the results, it is possible to correlate the average waiting time and the distance to travel by the taxis. As already discussed, the average time to serivce is related to distance that the driver must go to the customer: as far from the driver, the longer the customer will wait until be serviced. The correlation can see in the figure 6.

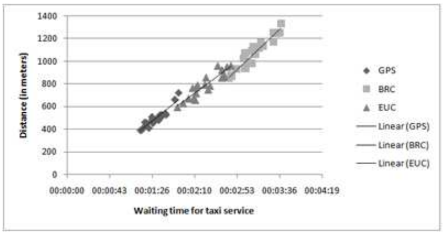

Fig. 6: Correlation between service time and distance.

Finally, should evaluate the runtime of the algorithms, to be able to measure its efficiency in the search for best taxi available to server the request. The figure 7 shows the average time of computational processing.

The analysis of figure 7 shows that the GPS-based algorithm with Euclidean distance and the Broadcasting algorithm have significantly lower processing time than GPS-based algorithm with lower estimated time for service.

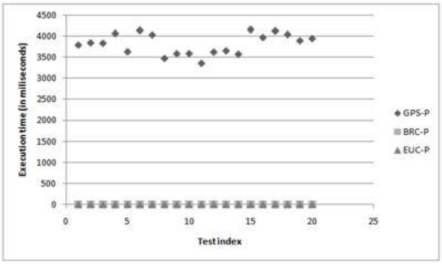

Fig. 7: Average processing time of algorithms.

Despite of the values of processing time of GPS algorithm, when compared to the other algorithms, its absolute value is low, and it can be used in a real application. Due to the its own characteristics, where the number of processed taxis is limited by restrictions of the algorithm and, the major time of response is caused by web service calls, it can said that the processing time will be stable even with a increase of taxis and customers in the a real system.

4.6. Test II: Decrease of taxi avalability

In the second test, was attempted to decrease the supply of taxis. To do this, was reduced the number of taxis available in the same area of the city. In this second experiment, the availability of taxi vehicle was 1 each 0.17 km^2.

To evaluate the impact caused by the changes in the test, the results of each algorithms, in new conditions, can be seen in figure 8. In the second test, at the same way as the first one, it is possible to say, with 95% that the GPS-based algorithm with lower estimated time for service has significantly smaller times than GPS-based algorithm with euclidean distance and the

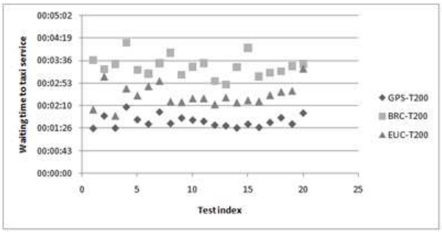

Fig. 8: Customer's average time to service, by each algorithm.

broadcasting algorithm.

The results of the average distance to travel by taxi drivers in the second test can be seen in figure 9 As in the

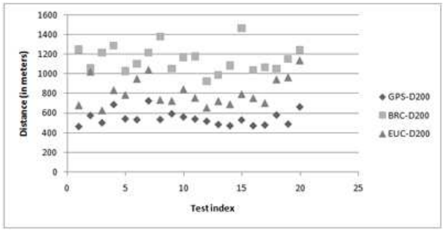

Fig. 9: Taxi average distance to serve a request, by each algorithm.

first test, the average distance traveled by taxi drivers in GPS-based algorithm with lower estimated time for service is smaller than the distances found in the other algorithms.

5. Discussion

The major criterion for the choose of the best algorithm among those described in this article is the contribution of each one to reduce the average waiting time for taxis services. A second criterion used in this article is the computacional cost of each algorithm.

In the average taxi waiting time criterion, it is possible to see that the GPS-based algorithms had a decrease of at least 25% when compared to the broadcasting algorithm. From a carefull analisys, the GPS-based algorithm with lower estimated time for service decrease 52.8% when compared with the broadcasting method. The decrease of time by the GPS-based algorithm with euclidean distance is 26.1%. At least, the gain of he GPS-based algorithm with lower estimated time for service over GPS-based algorithm with euclidean distance is 36.1%.

In the second test, the software results indicates that the reduce of availability of the services slightly decrease the gain of the GPS-based algorithms when compared to the broadcasting method. The GPS-based algorithm with lower estimated time for service was reduced in 51.3% and the GPS-based algorithm with euclidean reduced 27.1%, when compared to the broadcasting algorithm.

As the results in figure 6, it is possible to indicate a corelation between the distance to travel by the taxi drivers and the wait time for a taxi service. However, the relation between the both values has kept uniform because the traffic conditions was not evaluated in the tests.

In a real scenario, traffic information may cause important changes on the values found. A different route, that is usually slowly, may be the fastest one in a rush hour or in bad traffic conditions. But the traffic conditions, may also not affect the algorithm. To know the impact of traffic conditions on the alogorithms, is necessary to do the next analysis:

1. In a bad traffic situation limited to a small area, the speed of a congested road will significantly decrease and the linearity of distance-time can be changed: a longer route may be faster than a route that uses congested roads.
2. In a rush hour, for example, there are a overall decrease of speed of all routes. In this scenario it is not possible to say that rerouting is always justified. In this rush times, it is possible that distance-time corelation keeps existing.

The response time of the GPS-based algorithm with euclidean distance and the broadcasting algorithm are similar, with negligible value. The response of the GPS-based algorithm with lower estimated time for service is significantly slower than other algorithms, with average response in 3.8 seconds.

A software that takes a long time to respond a request may impact the final result of the algorithm, once the taxi can move through the city. The best vehicle, after a while, can start to meet another request or can not be the best choice, because it has moved to another direction. A example of how processing time can affect the result is a vehicle at 60 km/h over a 1 minute algorithm time response: the vehicle may goes 1 km away from the point evaluate from the algorithm.

The results of GPS-based algorithms when compared to the broadcasting method using computational cost evaluation must be taken with care, because the broadcasting algorithm do not really exists - it is a interaction between taxis and the call center operators.

After this alert, it is possible to calculate the impact of the slow processig of the GPS-based algorithm with lower estimated time for service. In a 60 km/h speed, it is impact on the original position of the taxis is, at most, 63.3 meters.

6. Conclusion

From the study, it is possible to indicate that GPS-based algorithms has better results than broadcasting algorithm in a taxi wait time criterion. The results indicate numerically that GPS-based methods can improve the quality of services, as was expected by [3].

Other goal of this work was to indicate the correlation between the distance to travel by the taxi to serve a request and the taxi waiting time. This correlation influences the method to choose the best taxi to the client - nearby taxis must decrease the average time of the system.

The analysis of processing time of the algorithms indicates that the described algorithms has low response time. Due its own characteristics, as well as the times of processing found in the simulations, it can be stated that the solutions presented have practical feasibility.

7. Future Work

In a future works it is possible to improve the developed models, making them closer to the reality. It is possible to consider different environmental conditions in order to produce more accurate results. The evaluation of external factors to the taxi search, as rains or concerts, in a similar way that was shown by [6] to predicting bus arrival times can be used by the algorithm for more accurate results.

References

[1] S. Cheng and X. Qu, X. A Service Choice Model for Optimizing Taxi Service Delivery, *Research Collection School of Information Systems*, v. 209, 2009.

[2] Z. Liao. Real-Time Taxi Dispatching Using Global Positioning Systems. *Communications of ACM*, v. 46, n. 5, may 2009.

[3] Z. Xu, et al. Investigating the Value of Location Information in Taxi Dispatching Services: A case study of DaZhong Taxi. *PACIS 2005 Proceedings*, v. 111, 2005.

[4] B. Rao and L. Minakakis Evolution of Mobile Location-based Services. *Communications of the ACM - Mobile computing opportunities and challenges*, New York, NY, USA, v. 46 , n. 12, p. 61 - 65, December 2003. ISSN 0001-0782.

[5] H. Yang and S. C. Wong. A Network Model of Urban Taxi Services. *Transport Research Board-B*, v. 32, n. 4, p. 235-246, 1998.

[6] W.-H. Lin and J. Zeng. An experimental study on real-time bus arrival - Time prediction with GPS data. *Transportation Research Record*, n. 1666, p. 101-109, 1999. ISSN ISSN: 0361-1981, ISBN 0309070619.