GEOINFO
Brazilian Symposium on GeoInformatics

OPTIMIZATION OF TAXI CABS ASSIGNMENT USING A GEOGRAPHICAL LOCATION-BASED SYSTEM IN DISTINCT OFFER AND DEMAND SCENARIOS

Alocação Otimizada de Taxis Através de um Sistema Baseado em Localização Geográfica em Distintos Cenários de Demanda e Oferta

Matheus Patrocínio Souza[1], Abílio Augusto Marinho de Oliveira[1], Marconi de Arruda Pereira[1], Felipe Augusto Lima Reis[2], Paulo Eduardo Maciel Almeida[2], Eder Junio Silva[1] & Daniel Silva Crepalde[1]

[1]Universidade Federal de São João Del-Rei - UFSJ
Departamento de Tec. em Eng. Civil, Computação e Humanidades
Campus Alto Paraopeba, Rodovia MG-443, Km 7 - Fazenda do Cadete
Caixa Postal 131, Ouro Branco-MG, CEP 36420-000
{matheuspatrocinio, abilio_deoliveira}@hotmail.com, marconi@ufsj.edu.br, ederjrsilva@yahoo.com.br, danielcrepaldeufsj@gmail.com

[2]Centro Federal de Educação Tecnológica de MG - CEFET-MG
Departamento de Computação
Av. Amazonas, 7675 - Nova Gameleira - Belo Horizonte - MG – Brasil
falreis@gmail.com, pema@lsi.cefetmg.br

ABSTRACT

In this paper, different approaches are evaluated to assign taxi cabs to customers in geographical location-based systems. The main purpose of this work is to identify the solution in which all current customers are met in an acceptable time, however minimizing the distance traveled by existing free taxi cabs. Two aspects are considered: 1) the method to calculate the distance between vehicles and customers; and 2) a vehicle assignment strategy. The methods to calculate the distance between vehicles and customers are: a GPS-based routing (a shortest path algorithm) and the Euclidean distance. On the other hand, as vehicle assignment approaches, the considered strategies are: a greedy algorithm, which assigns each vehicle to the closest customer, and an optimization algorithm, which assigns vehicles considering the whole scenario, minimizing the global distance traveled by taxi cabs to meet the customers. This last strategy considers an optimization model in such a way that the calls are not readily answered. In this case, a short waiting window is implemented, where the calls are stored and then the optimization algorithm is executed, in order to minimize the required distance and to meet all current customers. The combination of the two methods of distance calculation and the two vehicle assignment strategies formed four possible approaches, which are evaluated in a realistic simulator. We propose some different traffic scenarios, varying the amount of calls and available taxis in order to better evaluate the algorithms´ performance. Results show that the approach which uses the shortest path algorithm and an optimization algorithm reduces the average service time by up to 27.59%, and the average distance traveled by up to 45.79%.

Keywords: Geographical Information Systems, Traffic Simulation, Vehicle Allocation.

RESUMO

Neste trabalho, são avaliadas diferentes abordagens de alocação de táxis para clientes, utilizando sistemas baseados em localização geográfica. O principal objetivo deste trabalho é identificar a solução em que os clientes são atendidos em um tempo aceitável, porém minimizando a distância percorrida pelos táxis até alcançar os clientes. Dois aspectos são considerados: 1) o método para calcular a distância entre os veículos e os clientes; e 2) a estratégia de alocação do veículo. Os métodos considerados para calcular a distância entre os veículos e os clientes estão: tamanho da rota obtida via sistema baseado em GPS (i.e. um algoritmo de caminho mais curto) e a distância Euclidiana entre taxi e cliente. As abordagens de alocação de um taxi para um determinado são baseadas em duas estratégias: um algoritmo guloso, que atribui o veículo "mais próximo" ao cliente (de acordo com uma das estratégias de cálculo de distância), e um algoritmo baseado num modelo de otimização, o qual atribui os veículos considerando contexto como um todo, minimizando a distância global a ser percorrida pelos táxis para atender todos os clientes. Esta última estratégia foi baseada num modelo o qual não realiza o atendimento às chamadas imediatamente: uma janela de tempo é usada, de tal forma a armazenar as chamadas que surgem no decorrer de uma fração de tempo e, em seguida, o algoritmo de otimização é executado, a fim de minimizar a distância requerida para atender a todos os clientes. A combinação dos dois métodos de cálculo da distância e as duas estratégias de atribuição de veículos formou quatro abordagens possíveis, que são avaliadas em um simulador realista. Foram propostos também alguns diferentes cenários na relação entre a oferta e procura pelo serviço, variando a quantidade de chamadas e táxis disponíveis, a fim de melhor avaliar o desempenho dos algoritmos. Os resultados mostram que a abordagem que utiliza o algoritmo de caminho mais curto e um algoritmo de otimização reduz o tempo médio de serviço até 27,59%, e a distância média percorrida pelo até 45,79%.

Palavras-chave: Sistemas de Informações Geográficas, Simulação de Trânsito, Alocação de Veículos.

1. INTRODUCTION

Taxi services are an alternative urban transportation mode which offers convenience and speed in relation to public transport. It is noticed that taxi service is not efficient because much of the time, about 50%, they are empty (REIS *et al.*, 2013). On the other hand, some passengers complain that some calls are missed. In Belo Horizonte, for example, about 10% to 15% of customers, who prefer the convenience of ordering a taxi by phone, give up because of the long wait for the service (CASTELO-BRANCO, 2012).

a) To make this service more efficient, several studies are being conducted and new attribution methods based on geographical location are being adopted (REIS *et al.*, 2013). These methods use GPS (Global Positioning System) to locate taxis and customers and propose a better way to assign each vehicle to attend the existing calls. Usually, two aspects are explored by these approaches:

b) The procedure used to estimate the distance traveled by a taxi to achieve a customer. Two very common possibilities are the Euclidean distance, and the shortest path between points. It is important to note that the evaluation of the distance between the taxi and the customer can be treated in a more elaborate way, taking also into account the route that the taxi will go, considering factors that complicate the taxi displacement, such as speed limits, traffic jams, among others;

The vehicle assignment strategy. A common vehicle assignment method is based on a greedy approach, where the taxi cab which is closest to a customer is designated to take the call. The intent is to minimize the passenger waiting time and to minimize the idle taxi time. Moreover, as all greedy heuristic, this solution can be a good one in the local space of solutions but can be a poor choice when considering global solution space. Thus, the problem can be modeled as an optimization problem, where it is wished to minimize the idle taxi time, treating the passenger waiting time as a restriction of this problem.

This paper aims to explore these two aspects, in such a way to measure the effectiveness of the most common ways currently adopted to estimate distance and to assign taxi cabs to customers. Also, this work intends to propose an optimization approach to solve this problem, to objectively evaluate its performance by means

of micro-simulation and to draw some practical conclusion about the experiments undertaken. These experiments were carried out considering three different scenarios: a constant demand and constant supply of taxis during the simulation period (subsection 4.1), a high and variable demand (the double of the previous value), in order to simulate a peak time calls (subsection 4.2), and finally, a low offer with a large and variable demand of vehicles (subsection 4.3).

This article is based on Oliveira *et al.* (2015), previously presented at GEOINFO conference. The contribution of this new paper is the presentation of simulations in the three scenarios cited previously.

The remaining of this article is organized as follows: in Section 2 a historical background is revised, and some study cases are briefly presented. In Section 3, the tools and methods used to simulate and evaluate the discussed approaches are detailed. A practical experiment and some simulation approaches are presented in Section 4. Results and discussions about the experiments are shown in Section 5. Finally, in Section 6, some conclusions are presented along with some future work suggestions.

2. RELATED WORK

The use of mobile applications is increasing the productivity of resources, especially in the context of smart cities (STEENBRUGGEN *et al.*, 2015). (RAPER *et al.* 2007) present a several set of location-based applications, which can be broadly adopted in the context of this paper. A constantly growing demand, combined with an increasing traffic complexity, made the task to identify the best taxi to attend a call became very hard. This problem has been studied for some time around the world, showing that the problem is not exclusive to big cities, emphasizing the need for a solution to improve this picture.

In this sense, (XU *et al.*, 2005) describe a scenario in Shanghai, where taxis which joined Dazhong Company were modified and received GPS trackers, to indicate their location in real time. Customer requests arrive at a call center, which uses the trackers to know the nearest free taxi to answer each call, since the customer's location is also known. Then, the taxi drivers, within a certain range, respond to the call accepting it or not, via a button located on the equipment installed in the vehicles. Each consumer is then informed about how long it will take for his/her demand be met. The control of occupied and free taxi cabs and their distribution throughout the city is made in the dispatch Center of the company itself (DZDC - Dazhong Dispatching Center), which provides a higher speed in attendance.

In the city of Singapore, the system that handles the taxi service is the Automatic Vehicle Location and Dispatch System (AVLDS). This system, like the one used in Shanghai, consists of taxi cabs sending their positions to a central station. Customers calls arrive at this central, which congregates all taxi companies in the city. The system assigns the closest vehicle to each customer and waits around 10 seconds for the taxi driver to accept or decline the call. If he/she declines, then the system performs a new search for taxis using a new request for that service (LIAO, 2009) (YANG & WONG, 1998).

In both cities, the adoption of the mentioned systems provided a reduction of 16% to 32% on traveled distances without a passenger, and a reduction of 15 to 30 minutes in the customer average waiting time. These numbers were obtained considering the scenario with the traditional method, where the call center sends a broadcasting message, and a taxi driver who believes he/she is closer to the passenger offers to take the call.

(LIU *et al.*, 2015) directed their work to highlight the importance of communication between various taxi drivers, to predict where and when a customer will need a taxi. The past route information can be used in a machine learning process, combined to an optimization process in order to reduce the cabs idle time.

Conversely, (SANTOS & XAVIER, 2015) and (JUNG *et al.*, 2015) present a different practice that helps to reduce the idle traveled distance, using a cab sharing method (ride sharing). Each vehicle has a known location and all cabs are connected to a wireless network, so that the customers can find out if a vehicle will pass through the desired route. Then, they can share the same taxi cab with existing passengers. Therefore, this practice consists in helping both the driver and the customers, reducing both the global cost of travel and the overall traveled distances.

In none of the studies reviewed, it was noted the presentation in detail of an algorithm or method, that can reduce the distance to be traveled and the time required to collect the client.

3. TOOLS AND METHODS

In this section, the used tools and the methods will be presented, as well as the justification of use of them.

3.1 Tools

MATLAB (acronym for MATrix LABoratory) is a multi-purpose tool focused on numerical modeling, having matrices as its fundamental data structure. SUMO (Simulation of Urban MObility) is a simulation platform for discrete event traffic, microscopic, using continuous space and presenting inter- and multi-modal capabilities (KRAJZEWICZ *et al.*, 2012). SUMO was chosen because it is open source software, very popular in works focused on traffic simulations.

The relationship between SUMO and MATLAB is based on TraCI4Matlab (ACOSTA *et al.*, 2015), which is an API (Application Programming Interface) developed in MATLAB. TraCI4Matlab is an implementation of TRACI (Traffic Control Interface) protocol, which allows for interactions between MATLAB and SUMO in a client-server scenario. TraCI4Matlab can be used to implement vehicles control, traffic lights timing, intersections monitoring, among other applications of traffic for SUMO simulations.

3.2 Assignment Problem

Linear Programming Problems (LP) are those which present a linear objective function and linear constraints, regarding the existing design variables (TAHA, 2008). Certain special cases of LP, such as those involving network flow and multi-commodity flow, are considered so important that they have generated several researches and specialized algorithms in order to solve them.

There is a special case of LP which is called the "Transportation Problem" (TAHA, 2008). A common mathematical modeling approach to this case can be seen in equation (1), where m indicates the origins, n destinations, cij the cost per transport unit and xij the amount

sent from i to j.

Minimize:

$$z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} * x_{ij}$$

(1)

The solution of a Transportation Problem consists of determining xij values which will minimize the total cost of transportation and satisfy all constraints of demand and supply.

There is a special case of transportation problems, called "Designation Problem" (DP) (TAHA, 2008), which can be used in this context. The objective here is to systematically assign an available taxi cab to attend a customer, meeting some criteria z. A mathematical model to solve this problem can be seen in equation (2) (GOLDBARG & LUNA, 2005).

$$z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} * x_{ij}$$

(2)

Subject to:

$$\sum_{j=1}^{n} x_{ij} = 1, i = 1, \dots, n;$$

(3)

$$\sum_{i=1}^{n} x_{ij} = 1, j = 1, \dots, n;$$

(4)

For:

$$x_{ij} \in \{0, 1\}; \forall i = 1, \dots, n; j = 1, \dots, n$$

(5)

DP can be understood as the problem of allocating *n* producing cells to *n* tasks at a cost c_{ij}. If *i* receive the call to given *j*, x_{ij} turns 1, indicating the allocation of the call, otherwise, if *i* is not intended for the specific *j*, x_{ij} gets 0 (GOLDBARG & LUNA, 2005). Therefore, each variable x_{ij} is binary.

The solution of a DP can be obtained using the Hungarian Algorithm (HA) (JONKER & VOLGENANT, 1986). It works by analyzing a matrix consisting of j lines, indicating the number of calls to be answered, and i columns, representing each taxi cab available for service at any given cycle time.

In this paper, the Assignment Problem is used to identify which vehicle will met each customer call. To do this, it is necessary to calculate the distance (or cost) of each taxi to meet each customer. It can be done by a classic shortest path algorithm, like the Dijkstra algorithm (DIJKSTRA, 1959). For all free taxi cabs, the shortest path to meet the passenger is calculated; then, HA considers the costs of these paths to identify which is the best solution, considering the whole scenario and trying to minimize the criteria z.

In order to provide realistic calculations, Traci4Matlab library, coupled with SUMO, is used. It is important to note that these calculations go beyond to simply calculating the distance between sources and destinations. SUMO allows for a realistic simulation of the scene, where traffic signals, different speed limits, as well as the existence of other vehicles traveling on the streets, are simultaneously considered. This effect is similar to those GPS applications for smartphones, such as Waze and Tomtom, which use traffic information to calculate the best route to reach a given destination.

3.3 Methods

In this paper, the taxi service was handled as assignment problem, which can be solved with some naïve or with classical algorithms. We classify the taxi service in four ways, according to the presence or absence of information about geographical locations of taxis and passengers (REIS *et al.*, 2013). The first one, a "Random Searching" mode, occurs when a customer waits in a random location for taxis moving around on the way. This situation is characterized by the ignorance of positions of both, taxis and customers.

A second mode, "Fixed Stop", occurs when customers go after a taxi stop point.

It is characterized by the knowledge of taxi's positions, however the customer's initial position is unknown. The third mode is the "Broadcasting", which uses the geographical location of customers, but the positions of taxi cabs is unknown by the central control system. In an example of this mode, a customer calls the central service and the customer's location information is passed on by radio to the taxi drivers. The first driver to accept the request is confirmed by the central to pick up the customer. Since the position information of the available taxi cabs is unknown, this method does not guarantee that the closest driver or the most suitable vehicle is assigned to the customer.

The last mode is represented by "GPS-based models", which can identify the geographical position of both, customers and taxis. These GPS methods are distinguished from each other through the algorithms used to define the taxi driver which will be responsible for picking up each passenger. The geographical location of the actors involved underlies the criteria and costs used by the algorithms to choose the attendant to each request. Cost is a factor that determines the difficulty of a taxi driver to answer a call. Basically, two methods can be used: the Euclidean distance and the routing distance between each customer and the available vehicles (the distance that each taxi driver must travel through streets and avenues to reach the customer). The following sections explain some features of GPS-based algorithms, exploring both the distance calculation methods between actors, as the vehicle assignment strategy.

3.3.1 Greedy Algorithm based on Euclidean Distance

This allocation method is based on a Greed Approach (CORMEN *et al.*, 2001), where the closest taxi, according to the Euclidean Distance, is designed to attend the call. This method has the advantage of low computational cost required, compared to the next methods, due to the simplicity of the calculations involved. The operating principle is to calculate the distance between two points, taxi and client, then finding the taxi cab which has the lowest linear distance from the customer, calculated through Equation (6):

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (6)$$

where x_1 and y_1 are the coordinates of the customer and x_2 and y_2 are the coordinates of an available taxi cab.

For every customer call, the system checks which taxi is the best candidate to attend it, using therefore a greedy approach. As can be seen in the Figure 1, not always the result of

this method is the actual best. Due to the traffic structure of the cities, the Euclidean distance is not equal to the distance to be traveled to reach the destination. However, this is a solution to be analyzed, since it has a very large computational gain when compared to other methods based on shortest path algorithms.
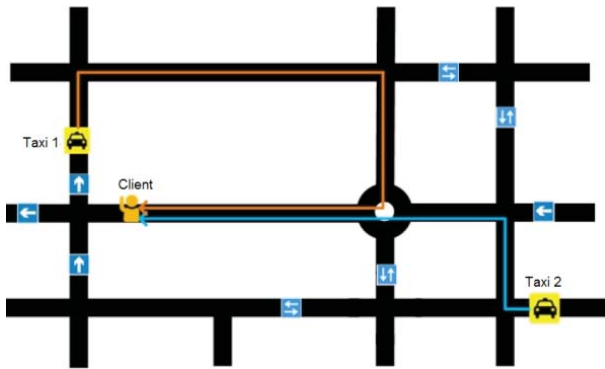

Fig. 1 - Problem of the Euclidean distance in the assignment of taxi cabs (REIS *et al.*, 2013).

3.3.2 Greedy Algorithm based on Shortest Path

This variation of the previous greedy algorithm aims to partially repair the deficiency pointed by Figure 1. The operating principle is to calculate the actual distance, finding the taxi cab which has the lowest driving distance to the customer.

This method still uses a greedy approach, since, for each customer call, a taxi cab which has the shortest driving distance to the customer is selected to do the service. The following methods to be detailed, on the other hand, try to consider a global view of the scenario. In some way, they effectively try to minimize the total distance that vehicles have to travel through the streets, without a passenger, before reaching the customers.

3.3.3 Hungarian Algorithm based on Euclidean Distance

The Hungarian Algorithm – HA – (KUHN, 1955) is a popular procedure for solving assignment problems. In this work, it analyzes the matrix formed by customer calls collected in a time window of 50 seconds and the taxi cabs available to attend, so that the total cost of the chosen solution pairs is minimized. So, this is a very different approach from Greedy Algorithm. The cost factor used, as this model's name implies, is the Euclidean distance between customers and taxi cabs.

If the number of calls is less than the number of available vehicles, some taxis are not included in the search, but all calls are answered in such a way that the total cost is the lowest possible. The same goes for the opposite case, when the number of taxis is lower. Some calls are not answered on that iteration, always looking to find a solution which brings the lowest overall cost.

The evidence that HA determines the best assignment among taxi drivers and customers is given by the Great Allocation Theorem and the König Theorem (KONIG, 1931). The Great Allocation Theorem says that if a real number is added or subtracted from all entries of a row or column in a cost matrix, then an optimal assignment for the resulting cost matrix is also an optimal assignment for the original cost matrix (KUHN, 1955).

3.3.4 Hungarian Algorithm based on Shortest Path

In this approach, the cost calculation used in HA is based on the actual distance to be traveled by the taxi cab to meet a customer. The calculation consists on the analysis of a matrix formed by available taxis and customer calls, on a time window of 50 seconds. Thus, the algorithm minimizes the total cost of taxi assignments. This optimization approach provides an improvement in the previous algorithm, since now the actual distance to be traveled by each driver is considered.

The use of HA does not imply a significant rise in processing time when comparing this approach with greedy algorithms. Its major issue is the need for a set of calls to improve the results, which demands the time window. The value of 50 seconds was chosen with the intention to avoid leading to excessive idleness of the vehicles, and conversely, a big amount of calls to be answered, keeping a trade-off duration during the processing of customer calls.

4. EXPERIMENT AND SIMULATIONS

The four algorithms presented in subsection 3.3 were implemented in practice: Greedy Algorithm based on Euclidean Distance (GED), Greedy Algorithm based on Shortest Path (GSP),

Hungarian Algorithm based on Euclidean Distance (HED), Hungarian Algorithm based on Shortest Path (HSP). The data sets were generated by SUMO simulations. The simulations used in each algorithm were carried out under the same conditions, considering a same number and an equivalent geographical distribution of taxi cabs and other vehicles. Likewise, customer requests are located in the same places and amount for all simulated algorithm. The choice of each taxi to meet the requests followed the criteria for each algorithm. Spent time and distances traveled by taxi cabs to service the requests are then made available by SUMO to MATLAB, allowing fair comparison among the methods.

The simulation environment tried to be as realistic as possible, considering two way and one-way streets, traffic lights and other vehicles traveling freely. The streets and avenues cover an area of 54 km2 (see Figure 2). Border roads were designated as rapid traffic routes, having a speed limit of 80 km/h. Vertical roads were considered arterial roads with a limit of 60 km/h and horizontal roads were considered collector roads with speed limit of 40 km/h.

The amount and types of vehicles which formed the traffic flow were set in proportion to the environmental area of the city of Belo Horizonte (CASTRO, 2014) (IBGE, 2014).
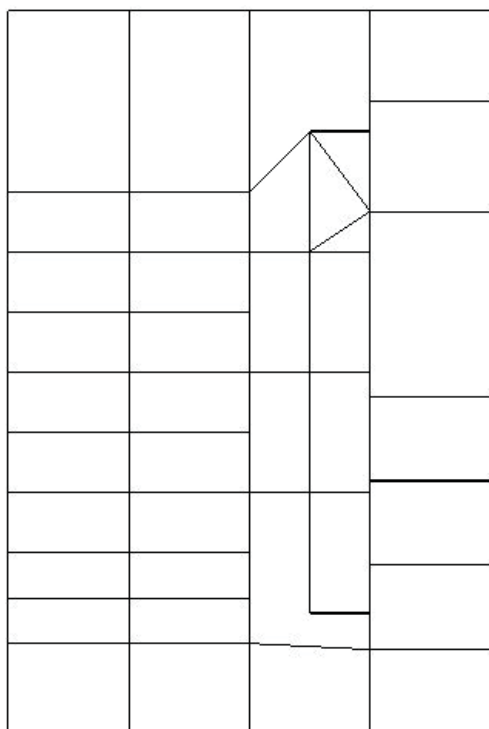

Fig. 2 - Simulation environment in SUMO.

The routes for all private vehicles (non taxi) were randomly determined. The simulation included: 9,333 private cars, 3,443 motorcycles, 155 buses and 450 taxi cabs, among which 225 were walking on the map and 225 are stopped. The number of taxi cabs was defined according to the average value of cabs into each 0.12 km2 in the city of Belo Horizonte (REIS *et al.*, 2013).

In order to better exploit the capacity of each allocation method, three different scenarios were proposed: (1) constant number of calls and high supply of vehicles; (2) wide demand for vehicles and high supply; and (3) varied demand of taxis and reduced supply of vehicles. Each scenario will be explained in the following subsections.

4.1 Constant Demand and High Offer

In this scenario, the demand is constant and the offer of taxis is high, both distributed evenly and randomly throughout the map. The call occurrence was also random. The same configuration is shown for each of the four proposed algorithms in order to obtain a fair comparison of them. To better represent the dynamics of the system, taxis have a specific initial operation area, which is amended in accordance with the duties to answer calls.

Calls are generated with constant frequency throughout the simulation. This steady state ensures that waiting times for services keep up uniform, directly related to the amount of time the taxis run free, without customers (YANG & WONG, 1998). However, in actual life, for each day of the week, time, region, events, weather, etc., there are variations in the number of requests. So, for this set of simulations, it is assumed to be part of a day, with a constant number of requests, giving a better scenario for comparisons.

Another feature, not included on the simulations, was the concurrency requests.

To statistically validate the results, the number of simulations conducted for each scenario was equal to 30, with a duration time of 4 hours, totaling 120 hours (or 5 days), to comply with Central Limit Theorem. On average, a taxi cab serves 15 calls per day, working for 13 hours (two or more drivers can use the same vehicle, in complementary time (LOPES, 2014)). Therefore, for each simulation of 4 hours, a number of

approximately 4 calls per taxi was randomly generated and distributed for both, instant of time and occurrence position, totaling 54,000 requests (approximately 8 calls / minute) simulated for each algorithm.

The requests were generated until a predetermined final time was achieved, and the same conditions were repeated for each simulation and for each algorithm, considering place, calling time and traffic, in such a way to ensure that the same factors were present under the same conditions, for each algorithm. Therefore, the algorithms defined the best taxi cab for each request, and the values of the waiting time and the distance were extracted and evaluated.

In our experiments, each simulation can be objectively described as follows:

- Vehicles are randomly distributed on the map: private cars, motorcycles and taxis. In each second, vehicles can enter to and exit from the network, simulating the traffic flow. However, the number of taxi cabs remains constant throughout the simulation;
- Requests are made and the simulation is independently carried out in 4 scenarios, one for each algorithm. Each algorithm determines the best taxis, according to its own criteria, to meet each customer;
- A taxi cab state changes from free to busy and it is classified as ongoing service, avoiding to be assigned to other calls. A minimum cost algorithm determines the best route taken by the driver to the customer;
- Once a taxi meets a customer, the distance and the traveled time spent to arrive to the passenger's location are stored.
- High Offer and Variable Demand
- The second scenario is a high supply of vehicles, but with a variable demand. The total calls value was twice the amount used in the previous scenario. This type of configuration is quite common on rainy days and near major holidays, when, not rarely, the demand for taxis can become twice the regular amount. The total simulation time remains the same, 4 hours, but in the first 1/3 of the time (80 minutes) are held 25% of the calls (approximately 12 calls / minute). The second 1/3, concentrate 50% of the calls (approximately 23 calls / minute). In

the remaining time, occur 25% of the calls.
- The intention of dividing this scenario in periods is to ensure realism in the simulation, since the taxis occupancy rate varies greatly during the day when, at peak times (e.g. 18:00 to 19:00) it can reach the values between 90% and 100% (YANG & WONG, 1998).

4.2 Reduced supply and Variable Demand

The third scenario is a change in the previous scenario, where in addition to a variable demand over time, the supply of vehicles was reduced by half. Here, the inspiration is the occurrence of big events, where the majority of taxis are moved to specific locations, and the rest of the city ends up with a very tight supply, while maintaining its usual demand in traditional regions.

5. RESULTS AND DISCUSSIONS

The four algorithms were tested considering the same conditions for the simulation in SUMO, in the three proposed scenarios. The results can then be analyzed to ensure that the comparisons do not place a positive bias for some of the algorithms to the detriment of others. The average time and traveled distance in each algorithm as well as the sum of these two measurements were calculated and are presented below. The standard deviation of each of these variables was also calculated.

As can be seen in Table 1, average and standard deviations of both attendance time and the total distance traveled were calculated for each algorithm, considering all 54,000 simulated calls divided into a total of 30 simulations performed for each algorithm.

The first column shows the mean time ± standard deviation of each method. The average waiting time for each call was measured from the time that the request was made by the client until the instant that a taxi arrived. For algorithms based on the Euclidean distance, it was observed that instantaneous vehicle's allocation algorithm (GED) was approximately 7.98% more efficient in attendance time, but the HED was 1.09% more efficient at average traveled distance to meet the customer. Variation of the average values for both, time and traveled distance, were very close. It was observed also that, when using

the Euclidean distance as a cost criterion in the algorithms, the implementation of the Hungarian Algorithm does not become more efficient, not generating significant results.

The highlight is due to the algorithms based on the actual distance to be traveled between customers and taxi cabs. The GSP showed up a gain of 46.48% in the average attendance time, when compared to GED, and a gain of 38.66% in average distance traveled, when compared to HED.

The proposed HSP algorithm presented an average attendance time of 27.59% lower than GSP. As the total distance traveled in attendance (Figure 4), HSP presented an even better result, approximately 45.79% below to the GSP. Another point to note is the stability of responses observed through standard deviation estimations. HSP showed up a variance around 25.10% lower than GSP in service time and a variance 30.66% lower than GSP, when comparing the average distances traveled.

Table 1: Attendance Time And Distance (Average ± Standard Deviation) – Scenario 1

	Time (minute)	Distance (km)
GED	4min36s ± 5min48s	3.37 ± 3.05
HED	5min ± 5min51s	3.34 ± 3.05
GSP	2min28s ± 2min11s	2.05 ± 1.99
HSP	1min47s ± 1min38s	1.11 ± 1.38

Considering the algorithms based on shortest path, these results mean an average gain of 0.9369 Km for HSP in relation to GSP in a single request of attendance. As 54,000 calls were simulated, the total gain was around 50,620 km relative to GSP. The result becomes even more significant considering, for example, the city of Belo Horizonte, where around 96,000 calls happen in a day (REIS *et al.*, 2013). In this configuration, HSP would imply a gain in relation to GSP by approximately 89,991 km daily.

Figure 3 shows another view of the obtained results, detailing the sum of the attendance times for calls. The algorithms based on the Euclidean distance presented a big difference when compared to algorithms that use actual shortest path. GED, however, showed up a total gain of 21,400 minutes compared to HED. In the graphic of distance traveled, shown

in Figure 4, the discrepancy between the lines of the algorithms based on the Euclidean distance is smaller, with a lead of 2,000 km for Hungarian algorithm (HED).
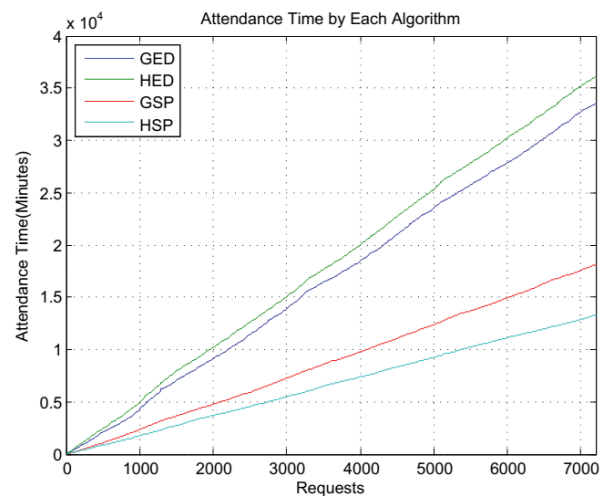

Fig. 3 - Sum of the attendance time by each algorithm - Scenario 1.

The algorithms based on the shortest path stood out. The total attendance time of GSP is similar to the attendance time of 29,000 requests of GED. GSP exceeded HSP in 36,680 minutes. The total time of HSP service (to 54,000 calls) is similar to the attendance time of 39,000 requests of GSP.

Even more significant are the results found for the traveled distance (Figure 4). The total traveled distance by taxi cabs to all calls of GSP amounts to the same distance of HED in approximately 32,860 calls. HSP again achieved the best result. It was smaller than GSP in 50,620 km. Its total distance is similar to the distance of GSP in 29,300 calls. So, HSP would save 50,620 km of travel, which is equivalent to 24,700 requests.

Table 2 presents the results obtained in scenario 2, where the supply is high and demand is twice the demand presented in the previous scenario and varying in time. The occurrence of a moment where there was a greater concentration of calls, as well as increased global demand for vehicles, did not generate significant change in the average attendance time, or the average traveled distance. On the other hand, the total time (Figure 5) and the total traveled distance (Figure 6) increased significantly. For both variables, the HSP remain the algorithm that performed better.
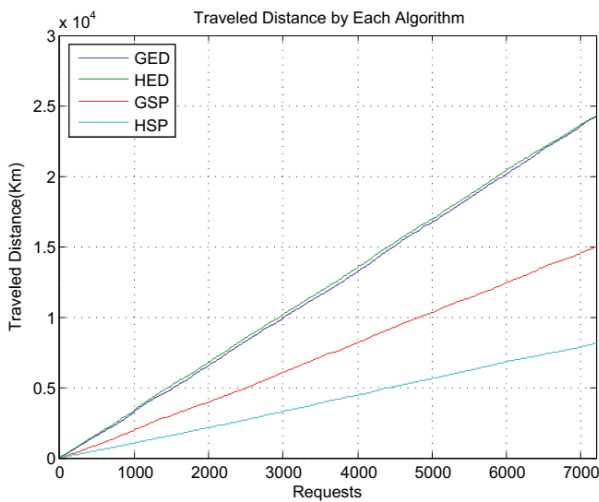
Fig. 4 - Sum of the traveled distance by each algorithm - scenario 1.

Table 2: Attendance Time And Distance (Average ± Standard Deviation) – Scenario 2

	Time (minute)	Distance (km)
GED	4min16s ± 3min20s	3.43 ± 3.08
HED	4min28s ± 3min17s	3.44 ± 3.01
GSP	2min18s ± 2min4s	1.93± 2.10
HSP	1min58s ± 1min59s	1.32 ± 1.78

Finally, in scenario 3, where the supply of taxis has been reduced by half, but demand remained the same as presented in scenario 2, where it changes over time, the difference between the performances of the algorithms that use the same distance calculation criterion is even closer (Table 3). Considering the average Time parameter, the algorithms based on Euclidean distance still losing to that based on the shortest path. Moreover, using a greedy strategy or using the proposed optimization model, provided no large difference in the mean time, or in the total time (Figure 7). Even when we consider the distance, it cannot be said that the there is a significant difference between a method that uses optimization or a greedy strategy, when the distance calculation strategy is based on Euclidean formula. However, the results make it clear that in the optimized taxi allocation criterion, using the shortest route, is more efficient. Here, the optimized method can stand out from the greedy method, achieving both a lower average distance as a total distance (Figure 8).
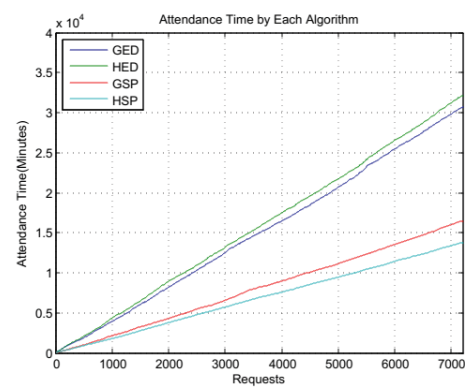

Fig. 5 - Sum of the attendance time by each algorithm - Scenario 2.
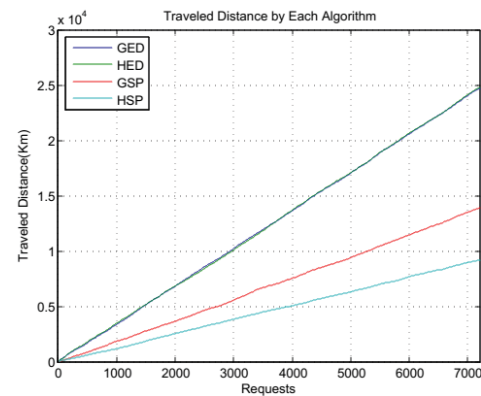

Fig. 6 - Sum of the traveled distance by each algorithm - scenario 2.

Table 3: Attendance Time And Distance (Average ± Standard Deviation) – Scenario 3

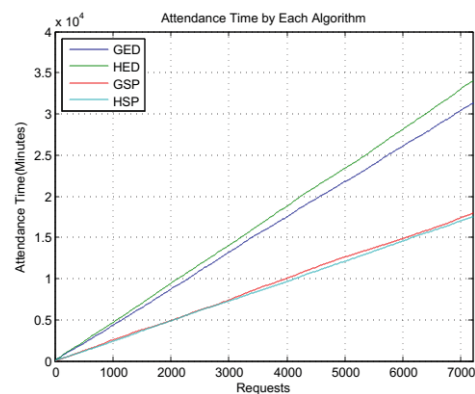|	Time (minute)	Distance (km)
GED	4min21s ± 3min26s	3.55 ± 3.04
HED	4min44s ± 3min24s	3.53 ± 3.01
GSP	2min31s ± 2min11s	2.14 ± 2.11
HSP	2min26s ± 1min57s	1.74 ± 1.89

Fig. 7 - Sum of the attendance time by each algorithm - Scenario 3.

6. CONCLUSIONS

This paper analyzed some of the most common approaches used to choose the taxis to attend the passengers, exploring different methods to calculate the distance between customers and vehicles, and strategies to taxi's designation. The paper presented an algorithm for choosing taxis based on a greedy strategy that uses the Euclidean distance (GED), a greedy algorithm based on the shortest path (GSP), an optimized method that uses the Hungarian algorithm (HA) based on the Euclidean distance (HED), and finally a method also based on Hungarian algorithm, but that uses the classic shortest path algorithm (HSP). GED and GSP can be considered the most popular algorithms, because they resemble the most used way to order a taxi: either via radio or via smartphone, vehicle assignment systems attempt to identify an available taxi, which are closest to the customer, so that he can answer the call. The use of an optimization algorithm, such as HA, to the taxi allocation process achieved positive results, especially when using the shortest route compared to the instant selection algorithm based on greedy strategy.
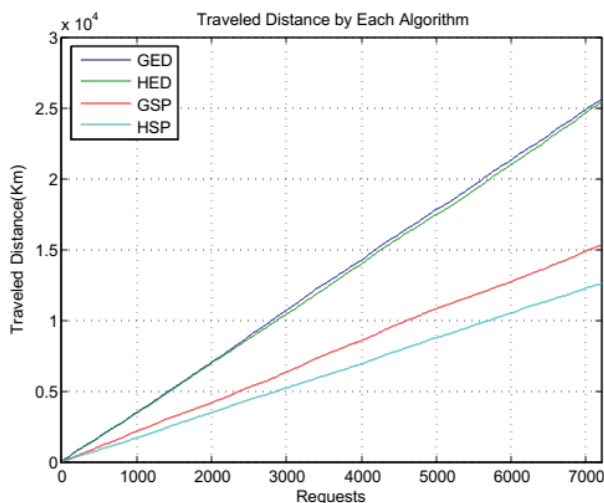

Fig. 8 - Sum of the traveled distance by each algorithm - scenario 3.

The simulations were based on a map with a complexity compatible to a city of small or medium size (different speed limits, traffic lights, two-way and one-way streets, and so on). This complexity can be increased, considering a differentiated distribution of taxi cabs along the map.

The simulations used three different scenarios, with demand and differentiated vehicle offerings, which sought to give more realism to the simulations. In fact, the demand and supply of vehicles can suffer a lot of change over time, according to day, time and holidays.

In summary, the use of an optimization model was very significant, even when the simulation setup was more adverse (scenario 3), providing a lower total distance to be traveled by the vehicle. New scenarios can be proposed in order to continue this process of validation of HSP algorithm. A relevant question to be considered is the use of a map of a city with more complexity. The map used in these experiments has rectangular quadroons and in a small number. As future work, we want to use a more complex map, such as the Belo Horizonte city, to verify the performance of the algorithms studied, especially if the HSP will continue to present significant gains.

Furthermore, the time window used for the two algorithms using the optimization model was arbitrarily chosen to be 50 seconds, and in special days, such as rainy or with special events (concerts, sports), this interval may have a different values. We intend to study the time window, so that the result can be improved even more, according to the scenario configuration.

This work has as its main focus at allocation of taxis, but these algorithms can be perfectly applied in different situations, as private fleet companies and even in the allocation of autonomous cars.

ACKNOWLEDGMENTS

REFERENCES

ACOSTA, A., ESPINOSA, J., & ESPINOSA, J. "Traci4matlab: Enabling the integration of the sumo road traffic simulator and matlab R through a software re-engineering process". In Behrisch, M. & Weber, M., editors, **Modeling Mobility with Open Data, Lecture Notes in Mobility**, p. 155–170. Springer International Publishing, 2015.

CASTELO-BRANCO, A. "Demora no atendimento de táxi em BH leva 15%

dos passageiros a cancelar pedido", 2012. Website. Disponível em: http://www.em.com. br/app/noticia/gerais/2012/06/12/interna_gerais,299515/demora-no-atendimento-de-taxi-em-bh-leva-15-dos-passageiros-a-cancelar-pedido.shtml; Acessado em 16 de novembro de 2016.

CASTRO, C. M. D. "Viaduto da floresta tem 40mil carros por dia, diz pesquisa origem destino", 2014. Website. Disponível em: http://g1.globo.com/minasgerais/noticia/2012/10/viaduto-da-floresta-tem-40-mil-carros-por-dia-diz-pesquisaorigem-destino.html. Acessado em 16 de novembro de 2016.

CORMEN, T. H.; STEIN, C.; RIVEST, R. L. & LEISERSON, C. E. "Introduction to Algorithms" (2nd ed.). **McGraw-Hill Higher Education**, 2001, 1312 p.

DIJKSTRA, E. "A note on two problems in connexion with graphs". **Numerische Mathematik**, 1(1): p. 269–271, 1959.

GOLDBARG, M. C. & LUNA, H. P. L. "Otimização combinatória e programação linear: modelos e algoritmos". **Elsevier**, Rio de Janeiro, RJ, Brasil, 2005. 649 p.

IBGE. "Área territorial de Belo Horizonte", 2014. Website. Disponível em: http://www.ibge.gov.br/home/geociencias/areaterritorial/area.php?nome=Belo Acessado em 16 de novembro de 2016.

JONKER, R. & VOLGENANT, T. "Improving the hungarian assignment algorithm". **Operations Research Letters**, 5(4): p.171–175, 1986.

JUNG, J., JAYAKRISHNAN, R., & Park, J. "Dynamic shared-taxi dispatch algorithm with hybrid-simulated annealing". **Computer-Aided Civil and Infrastructure Engineering**, p. 275–291, 2015.

KONIG, D. "Gráfok és mátrixok". **Matematikai és fizikai lapok**, 38, p. 116–119, 1931.

KRAJZEWICZ, D., ERDMANN, J., BEHRISCH, M., & BIEKER, L. "Recent development and applications of SUMO - Simulation of Urban Mobility". **International Journal On Advances in Systems and Measurements**, 5(3&4): p.128–138, 2012.

KUHN, H. W. "The hungarian method for the assignment problem". **Naval Research Logistics Quarterly**, 2: p. 83–97, 1955.

LIAO, Z. "Real-time taxi dispatching using global positioning systems". **Communications of ACM**, 46(5): p. 81–83, 2009.

LIU, S., WANG, S., LIU, C., & KRISHNAN, R. "Understanding taxi drivers' routing choices from spatial and social traces". **Frontiers of Computer Science**, 9(2): p. 200–209, 2015.

LOPES, V. "Bhtrans fecha o cerco ao táxi ocioso", 2014. Website. Disponível em: http://www.em.com.br/app/noticia/gerais/2013/02/15/interna_gerais,350571/bhtransfecha-o-cerco-ao-taxi-ocioso.shtml. Acessado em 16 de novembro de 2016.

OLIVEIRA, A.; SOUZA, M; PEREIRA, M.A.; REIS, F. A. L.; ALMEIDA, P. E. M.; SILVA, E. J.; CREPALDE, D. S. "Optimization of Taxi Cabs Assignment in Geographical Location-based Systems". **XVI Brazilian Symposium on GeoInformatics**, GEOINFO2015. Campos do Jordão, SP, Brasil, p 92-104 2015. http://mtc-m16c.sid.inpe.br/col/sid.inpe.br/mtc-m16c/2015/12.10.17.17/doc/proceedings2015_p8.pdf

Raper, J., Gartner, G., Karimi, H., & Rizos, C.. 2007. "Applications of location-based services: a selected review". **Jounal of Location Based Services**. 1, 2 (June 2007), p. 89-111. DOI=http://dx.doi.org/10.1080/17489720701862184

REIS, F. A. L., PEREIRA, M.A., & ALMEIDA, P. E. M. "Location-based dispatch to reduce the waiting time for taxi services". **Fourth International Workshop on Knowledge Discovery**, Knowledge Management and Decision Support, 2013, Mazatlán. Proceedings of the Fourth International Workshop on Knowledge Discovery, Knowledge Management and Decision Support, 2013, 10 p.

SANTOS, D. & XAVIER, E. "Taxi and ride sharing: A dynamic dial-a-ride problem with money as an incentive". **Expert Systems with Applications**, 42(19): p. 6728–6737, 2015.

STEENBRUGGEN, J., TRANOS, E., & NIJKAMP, P. "Data from mobile phone operators: A tool for smarter cities? New empirical

approaches to telecommunications economics: Opportunities and challenges Mobile phone data and geographic modeling" **Telecommunications Policy**, 39(3-4): p.335–346., 2015.

TAHA, H. A.. "Pesquisa Operacional: uma visão geral". **Pearson Prentice Hall**, São Paulo, SP, Brasil, 2008, 360 p.

XU, Z., YUAN, Y., JIN, H., & LING, H. "Investigating the value of location information in taxi dispatching services: A case study of dazhong taxi". **Pacific Asia Conference on Information Systems**, 2005, 12 p.

YANG, H. & WONG, S. "A network model of urban taxi services". **Transportation Research Part B: Methodological**, 32(4): p. 235–246, 1998.