INNOVATIVE ARCHITECTURE BASED ON BIG DATA AND GENETIC ALGORITHM FOR TRANSPORT LOGISTICS OPTIMIZATION.

Z. MOUAMMINE1, A. AMMOUMOU2, B. NSIRI3, S. BOUREKKADI4

1Laboratory Industrial Engineering, Information Processing and Logistics (GITIL), Faculty of Science Ain Chok. University Hassan II, Casablanca, Morocco
2Laboratory Industrial Engineering, Information Processing and Logistics (GITIL) High School of Technology University Hassan II, Casablanca, Morocco
3Research Center STIS, M2CS, National School of Arts and Crafts of Rabat (ENSAM), Mohammed V University in Rabat, Morocco
4Salmane BOUREKKADI, university of Poitiers, Poitiers-France, university of ibno tofail, Kenitra, Morocco.

Address for correspondence: E-mail: zakaria.moammine@gmail.com.

ABSTRACT

Many studies were carried out in the smart city field to improve the integration or to optimize the intelligent transport systems (ITS) in business. Yet, there were very few researches studying how ITS can still work outside of the context of a smart city. Taking this issue into consideration, this study was carried out to suggest a new system allowing shipping companies to keep working with their ITS even outside of the context of Smart city based on Genetic algorithm, Big Data, and multi-agent architecture. This is possible through implementing a transport information system, which will provide and exchange data about the optimum path to follow with vehicles (traveling vehicle problem) on the road. Theoretically speaking, this proposed system accelerates the processing time and enhances the quality of the obtained result with the Genetic algorithm to solve Salesman problem.

Keywords: Distribution Logistics, Intelligent Transport Systems, Traveling Salesman Problem, Genetic Algorithm, Big Data, Smart City.

1. INTRODUCTION

One of the most emerging matters all over the world is business competition. However, globalization nowadays affects the format of the markets and makes them available to overall the international dimensions. With regard to this fact, businesses that would like to globalize might encounter various organizational factors. These factors could promote or prevent the globalization process of the business. It is worth noting that they are very beneficial for the execution of a global strategy [8]. Among these factors is the ability to get along with this age of digitalization and new information techniques. Besides, the quantity of data created every year is much bigger than that created ever before, that is why our age is called the age of information (figure 1).

Figure 1: Exponential growth in digital data during actual decade [1].

Data analysis is considered as a vital necessity for transport and shipping companies in particular to improve or promote their operations. They can also be used to ensure that manufactured goods reach the customer quickly and reliably (Distribution Logistics). In fact, this change has led to the appearance of Intelligent Transport System (ITS).
For an effective integration of ITS, it is important to be within the context of a smart city. It is obvious that smart cities need special construction materials. In other words, expensive investments are needed to build a smart city. Among the foundations of a smart city we can include the construction of smart roads and highways, smart traffic lights, radio frequencies instruments, network of Internet of things (IoT), smart vehicles with sensors, and solar everywhere if possible so as to create the corresponding infrastructure. Some of the main requirements of ITS in Smart Cities are presented in Table 1 below.

Table 1: Requirements of Intelligent Transportation Systems in Smart Cities [10].

<table>
<thead>
<tr>
<th>S No</th>
<th>Requirements of Intelligent Transport System in Smart Cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Rapid intervention in an emergency situation.</td>
</tr>
<tr>
<td>02</td>
<td>Redesign and management of street as per the requirements of different transport modes.</td>
</tr>
<tr>
<td>03</td>
<td>Affordable by all section of the society</td>
</tr>
<tr>
<td>04</td>
<td>Congestion free route for faster service and safe mobility.</td>
</tr>
<tr>
<td>05</td>
<td>Accessible in most of the area of smart cities.</td>
</tr>
<tr>
<td>06</td>
<td>Sufficient spaces for parking of vehicles in different areas of smart cities</td>
</tr>
<tr>
<td>07</td>
<td>Faster service to reach different areas quickly</td>
</tr>
<tr>
<td>08</td>
<td>Safe mobility of public and vehicles in smart cities.</td>
</tr>
<tr>
<td>09</td>
<td>Real time information system for safe and efficient movement</td>
</tr>
<tr>
<td>10</td>
<td>Environmental friendly and energy efficient.</td>
</tr>
<tr>
<td>11</td>
<td>Better facilities to users during waiting at stops, travelling and transferring.</td>
</tr>
<tr>
<td>12</td>
<td>Efficient fare collection service</td>
</tr>
</tbody>
</table>

However, all these requirements cost a huge amount of money to be implemented. This cost might be affordable in certain cities belonging to developed countries unlike in other developing countries where financial problems are encountered and essential information and communication exchange tools and devices are missed in roads. Thus, the question arises, what can serve as an alternative platform for intelligent transport system in these cities? For that reason, this study seeks to provide a solution that can replace the lack of the needed infrastructure for the ITS to work properly without a smart city equipment. This was carried out according to the Multi-agent architecture. Specifically speaking, data and information are exchanged through a permanent interaction of two principal agents for the sake of solving the traveling salesman problem. This demands finding out the best path that covers multiple cities to visit. Moreover, in the technical part of this solution, a novel approach is proposed based on Big Data tool to improve the result and accelerate the time of processing of genetic algorithm applied for solving the TSP. Overall, this paper tackles and discusses the mentioned issues in the next sections.

In general, this paper is divided into 9 main parts. An introduction as a first part followed by a literature review where the related works are stated. The third part is about Big Data in Intelligent Transport Systems followed by conceptualizing travelling salesman problem in the fourth part. After that, a description of the Genetic Algorithm principle is presented in the fifth part. The sixth part involves a description of the problem to solve whereas the seventh part includes a description of the proposed approach. This approach is introduced in the eighth part. Finally, a conclusion and implications of the study are presented in the ninth part.

2. RELATED WORKS

The author of the article [1] proposed a new crossing operator for the problem of moving vehicles in order to reduce the total distance. Moreover, the authors noted improvements while verifying the calculation results of the use of an ordered intersection and partial mapping of the path representation method, though this path method is old, and the new reference TSPLIB case for the circular intersection operator. Recently, other authors [2] suggested a new way to parallelize and adapt the GA. This was through the design of three kernels that are concurrent. Every kernel in this design runs GA operators in a dependent and effective way. Moreover, each kernel can be straightforwardly adapted to run on many core and also multi-core processors. In addition, threads that run any of the triple kernels are synchronized by a low-cost switching mechanism so as to best employ the valuable resources of such processors in parallel execution of the GA. In general, results demonstrate the efficiency of the proposed method.

On the other hand, a recent study suggested a hybrid algorithm. This algorithm was named “ant colony-partheno genetic” (AC-PGA). In fact, the combination of of partheno genetic algorithms (PGA) and ant colony algorithms (ACO) provided the AC-PGA. Besides, this study was carried out to divide the variables into two parts. In other words, it uses PGA to comprehensively look for the best value of the first part variables and after that uses ACO to determine the value of the second part variables. Their results were also effective.
3. BIG DATA IN INTELLIGENT TRANSPORT SYSTEMS

Big Data has been integrated widely in logistics industry fields, supply chain management, e-commerce, reverse Logistics etc. It has significantly improved these logistics operations significantly, the mean point of big data is to identify and extract information from the mass of data [4]. Meanwhile, predictive tools can be developed to anticipate the future.

As well as the previous logistics operations, the logistics transport is on continuous process of optimization and minimization of it cost, it took advantage of the era of technologies and IOT to emerge toward Intelligent Transport System (ITS). For the Intelligent Transport Systems, technology is applied via the integration of electronic satellites, computers, and sensors. This is carried out to increase the safety and efficiency of every transport mode and make it more energy saving. Besides, the reaction to dangerous conditions, congestion, and to the movement of goods and people’s optimization is a key element for the effectiveness of ITS, which depends on the gained transport-related information.

In order to address these challenges, it is important to consider that Big Data introduces aspects to consider according to its properties described by the 5V’s model [5]: Volume, Velocity, Variety, Veracity, and Value.

4. DEFINITION OF TRAVELING SALESMAN PROBLEM

An effective way of defining the traveling salesman/salesperson problem (TSP) is through giving examples. The problem of a salesperson travel is the act of facing difficulties to provide a good answer to a special issue. This issue is about selecting the shortest route, between many, that connects several cities. In other words, the question of the TSP is to find a way that connects two cities or more and goes back to the starting point. Assuming that n is the number of city to visit, and the number of possibilities is (n - 1)! the number of possibilities. It is an NP-hard problem in combinatorial optimization. This is very useful in operations’ research and theoretical computer science. In general, the objective behind solving this problem is to reduce the visited distance, but it can also work according to many other criteria.

5. GENETIC ALGORITHMS PRINCIPLE

Genetic Algorithms are a search heuristic inspired by Charles Darwin’s theory of natural evolution, they belong to the larger class of evolutionary algorithms (EA), They reflect the process of natural selection where the fittest individuals are selected according to selection criteria for reproduction in order to produce ofspiring of the next generation, there are 3 categories of Genetic algorithms:

a. simple genetic algorithm (SGA)
b. Parallel and Distributed Genetic Algorithms (PGA and DGA)
c. Hybrid Genetic Algorithm.

by trying to emulate the process of natural selection, GAs can find an optimum solution to many difficult problems, The working mechanism of an algorithm genetic is presented in the figure 2.

When it comes to apply the GA on a given problem, the researcher has to choose, relatively to the specificity and complexity of the problem to solve, Among the existing presentations; matrix, binary, path, ordinal or adjacency.

Figure 2: Genetic Algorithms Mechanism
6. PROBLEM DEFINITION

The current section provides an example to define the problem in this study. The situation is when a vehicle \( V_1 \), existing in Casablanca intends to deliver commodity in many different places/cities without visiting them more than once, for instance, the table 1 shows a set of targeted cities to visit and the distance between them.

The issue here is how to find out the optimal path (shortest and/or cheapest) which will allow the vehicle \( V_1 \) to serve the whole cities at a lower cost.

7. THE PROPOSED APPROACH.

The proposed approach in this study is a multi-agent system (Fig. 5). This is to say that this approach is based on genetic algorithms reliability in term of finding an optimal or near-optimal solution to a given problem and on Big Data advantages in term of distributing computing and executing complex processing on data in a very short time. This proposed system will be used as a framework to find out the optimal or near-optimal trajectory for traveling salesman problem, which will reduce the transport cost for both logistics and shipping companies alike. Adopting the Multi-Agent Architecture: the solution of this study was designed so as to be composed of four main agents working independently as a team and interacting with each other to reach the same goal (refer to figure 3).

A. The Agent behavior in this Distributed Architecture for solving TSP.

a. The Master Agent, its mean function is to gather data and information which are sent by the GUI agent, after a storing operation, it delivers them to the Genetic Optimization Algorithm agent using Agent Communication Languages (ACL), this agent ensures and controls the communication between different agents composing the system, when the
Figure 4: Interaction diagram for distributed Model for solving Traveling salesman problem.

Solution of the best path is found it sends it back to the GUI Agent.

b. The Genetic Algorithm Agent: its mean objective is finding the optimal path between a set of targeted cities which fit the preferences of users (drivers), after generating, at first sight, an undirected weighted graph this agent ensures the processing of the Genetic algorithm optimization by assigning this complex task to the MapReduce function, it also calls, at second sight, a specific function which updates the values of the graph edges with accurate data gathered from Driver Agents at current tours.

c. The GUI agent: in this level, The Driver (user) chooses, on the graphical User Interface, a number of cities in which he aims to deliver commodity, this agent, basing on its knowledge base, extracts information about different distances between these cities, theoretical time needed to move between two points (cities or neighborhoods), and sends them to the Master Agent, once the best Path is found, it gets delivered to this agent.

d. The last one, is the Driver Agent and its essential role is to provide in real or near-real time data and information about the road conditions, traffic jam, and effective time needed for moving between cities which is based on the current tour experience, it sends these data in a frequency of 1 hour to the Master Agent.

The GUI Agent, which is installed in the traveling vehicle, creates in a given time $T_1$ a request that contains a set of targeted cities or neighborhoods. This also contains the different distances between the targeted cities or neighborhoods and the theoretical time needed to move between every two points. This time is sent through an internet network to the Master Agent. Later on, it asks the Master Agent for the most favorable way to travel to the whole cities. This master agent saves data and transmits them to the GA agent. It is worth noting that the exchange and communication between the agents is a responsibility of the master agent.

Genetic Algorithm Agent, which has the domain expertise for finding the optimal path serving the targeted cities considering the required parameters, it receives these parameters (cities or neighborhoods, distances, time of moving between each pair of cities, current position), then it constructs an undirected weighted graph $G= (V, E)$ where:

- $V= \{1,2,..,n\}$ N a set where items are called nodes, the number of nodes is equal to the number of targeted cities.
• $E = \{i, j\} \in V$ and $i \neq j$. A set of edges (possible paths between cities)

We assume that the undirected weighted graph should note that: $(j, i) = (j, i)$.

• A path between two nodes $i$ and $j$ of the graph is a series of arcs linking $i$ to $j$. It is noted $Path(i, j)$. The set of the existing paths is noted $SPath(i, j) = \{Path(i, j)\}$.

Formally: $Path(i, j) = (w_0, w_1, w_2, ..., w_n)$ with:

- $w_n \in N$, $\forall k \in [0, n]$
- $w_0 = i, w_n = j$
- $\forall w_n, w_{n+1} \in SPath(i, j)$ $(w_n, w_{n+1}) \in E$

Actually, the GA assigns this task to a function which takes as parameters a set of cities, distances, and other parameters related to the roads conditions, then it returns an initial undirected weighted graph as result.

After the generation of the graph representing the eventual geographical map of the targeted city, in this stage, an evaluation phase of the real-time road conditions is necessary, for this purpose the Genetic Algorithm Agent exploits different data sent each 60 seconds by the Driver Agents of the other vehicles which are in current tours, this data contain information about the effective time needed to move between the different targeted cities, real-time traffic jam, weather, road incidents, all this data is analyzed and used to modify the previous saved values of edges, then the costs of edges for the whole graph get updated with the new values.

In fact, this research uses a combination of an equal rate of the road distance and the time needed for moving between two cities for evaluating the edges cost. The outcome of this step is an accurate updated graph.

The second step is the application of the principle of Genetic Algorithm on the given graph so as to find the optimal path/trajectory to be followed.

Application of the genetic algorithm:
In the current stage, it is required that a step of adaptation for the G.A should happen to a certain problem domain. Meanwhile, it is important to take into consideration and determine some related parameters:

• The representation of the problem: as the GA on TSP is applied, it is important to choose one of the existing presentations of a round trip (matrix, binary, path, ordinal, and adjacency). In this paper, the natural way is adopted, which is probably by using path representation. For example, a round trip can be represented simply as a series of numbers, each number represents a city, let’s have: $(1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7) = \{RABAT, ELJADIDA, BERRECHID, BENIMELLAL, KENITRA, TETOUAN, KHIRIBGA\}$.

• The number of cities or places to visit during the trip (this paper worked on 7 cities travelling case)

• The number of generation is a parameter which is strongly related to the size of the initial population, it can be large if the
The number of the initial population is small and vice versa, in this study, it is designed to be configurable.

- The determination of the abovementioned parameters is more or less simples. On the other hand, the starting initial population size is among the most demanding and challenging matters of genetic algorithms. It is obvious a former population that is well set secures diversity. Diversity is among the most important items of achieving a better quality of the solution. A factorial function represents the number of the potential permutations. Indeed, this makes it difficult to run over all the possible permutations and pick up the optimum. In this case study, there are only 7 cities, which means that the size of the search space stands as 7! (5040 possibilities). However, a number of 15 cities for instance, gives 1 307 674 368 000 possible permutations of routes, and this is impossible to go through using the classical computing system.

- In fact, the quality of the solution is affected by the size of the initial population. This is to say that when the size increases, the quality increases too. This process needs additional time of computing. While using the old version of the algorithm application, the computing time was an obstacle to getting a larger population size. An additional reason is the classical computing methods that were used, which were weak. However, the current study provides a well-designed solution that can be of service to multiple vehicles at the same time. The Master agent has to resolve the objective functions for many Driver agents in a very short time according to the following formula.

Objective function for TSP modeling :
\[ \sum_{i,j \in V; i \neq j} c_{i,j} x_{i,j} \]

- \( x_{i,j} \) indicates respectively the inclusion or exclusion of every edge.
- \( x_{i,j} \in \{0,1\} \forall (i,j) \in E \)
- \( c_{i,j} \) The cost (Distance, time or other criteria).

While the evaluation function (fitness function) assigns to each chromosome a value, in the TSP problem the fitness value is the inverse of the length of the related round trip. For each chromosome \( Cr=\{G1, G2... Gn\} \) the fitness function is given with:

\[ f(Cr) = \frac{1}{c_{Gn,G1} + \sum_{i=1}^{n-1} c_{Gi,Gi+1}} \]  

Taking in consideration the number \( k \) of eventual TS that can ask simultaneously for the shortest path, the AG calculates in each iteration \( I \) the fitness function for each chromosome \( C \). This of course takes a considerable time of processing, for that reason satisfying the demand of a multiple TS within a reasonable time is a challenging part of the Genetic Algorithms, yet with the appearance of distributed computing systems and the powerful Big Data architecture, it is henceforth possible, first to serve multiple TSP, and second to take a considerable population size.

For the sake of achieving high-quality results, the initial population size was extremely raised in the current study. This obviously expanded the processing time. Several studies and experiments were performed to control and find out the appropriate solutions for the speed restrictions. This was carried out by endorsing Big Data to enhance the process of velocity. The Hadoop MapReduce function was implemented in the current study to distribute computing on many machines. For this, technologies like Apache spark and Hadoop yarn were used as it is shown in figure 4. Theoretically speaking, this has greatly enhanced the time of computing.
computers’ network was used. As displayed in figure 6, the Genetic Algorithm Agent, the GUI agent, the Master Agent and the driver agent make up the four important containers of the algorithm.

- Parameters related to the execution environment.
  - Hp Omen = CPU 2.2 GHZ (12 CPU) and 16G of RAM
  - Eclipse IDE
  - 5 JAR files serving for the JADE environment

- Parameters related to the study case.
  - Number of iteration: 50
  - Length of population: 20
  - Number of cities: 7
  - Mutation Probability: 1.5%
  - The crossover probability: 0.5%
  - Number of Vehicles: 60

B. At the launch step: The user/client selects the cities or neighborhoods to visit in addition to the distance to activate the process. The master agent receives the information and handles the other steps.

C. For the final step, which is the execution, it is based on some rules: The Master Agent controls all of the existing agents on the environment via command messages. The different steps of the execution have been already described in the paragraph dedicated to the behavior of Agent in this Distributed Architecture for solving TSP, at the end, the GUI draws the best path to follow on the dashboard of the travelling vehicle.

The configuration of our solution is done on three essential steps:

a. Creation of the essential agents.

b. Initialization of the parameters for the GUI agent

c. Execution

A. At the creation step: we start by creating a Main container, then, the four additional containers are created for the four essential agents, they all include the address of the Main container. The GUI Agent, is the only agent created with a user graphical interface.

The enhancement in terms of the quality of the solution in addition to the processing time that was reduced can be observed. Moreover, vehicle V1 receives the best way taking into consideration 50% criteria for road condition and the same percentage for the time in an optimized time of 0.1317 minutes as displayed in figure 7.

The GUI agent running on the vehicle V1 receives the information about the optimal trajectory to follow. After that, it notifies the driver of the
trajectory through showing a map on the vehicle dashboard (fig.8) with the order of cities to visit.

![Figure 10: Road network for the VI Travelling Salesman.](image)

9. CONCLUSION

It was recognized that there is a lack in terms of the required technological infrastructure for the use of the ITS. Therefore, the current study was performed to tackle this issue and suggest solutions. Among the suggestions, we revealed an approach that lets the shipping companies continue the work and use of the ITS though there is no smart city context. A new transport information system was implemented based on Genetic algorithm, Big Data, and multi agent architecture. This IS exchanges information with vehicles (travelling vehicle problem) on the road so as to provide them with data about the optimum path to follow. Secondly, the proposed model provides a real-time updating for the graph edges representing the road condition in the TSP modeling, which guarantees results that are more efficient. The third important point of this study is the improvement of the efficiency of the genetic algorithm by expending the size of the initial population and decreasing the processing time of the algorithm.

For further study, we would like to consider other application of genetic algorithm optimization for instance, we intend to apply our framework to solve other problems related to the transport and shipping especially the milk run problem, which is used to transport mixed loads from different suppliers to one customer. Instead of each supplier dedicating a vehicle every week to meet the demands of one customer, one truck (or vehicle) visits the suppliers to pick up the loads for that customer.

For further study, we would like to consider other application of genetic algorithm optimization for instance, we intend to apply our framework to solve other problems related to the transport and shipping especially the milk run problem, which is used to transport mixed loads from different suppliers to one customer. Instead of each supplier dedicating a vehicle every week to meet the demands of one customer, one truck (or vehicle) visits the suppliers to pick up the loads for that customer.

REFERENCES:


