INITIAL POPULATION GENERATION CRITERIA FOR THE OPTIMIZATION OF BUS LINES WITH GENETIC ALGORITHMS

Criterios de generación de la población inicial para la optimización de líneas de autobuses con algoritmos genéticos

M. Lormeus Bloy Fujimoly[1], Ricardo Montoya Zamora[1]

[1]Universidad Autónoma de Querétaro (UAQ)

Autor de correspondencia
fujimoly_menard@yahoo.fr
Ricardo.montoya@uaq.mx

ABSTRACT

In recent years, genetic algorithms have become a tendency to find solutions to complex problems, such as the design and optimization of transport networks. During the optimization of the design of public transport bus networks by genetic algorithms, the way of generating the initial routes leaves us with many doubts, like the amount of routes and how to generate them before making the process of crossing and mutation to reach the network's optimal performance. In this work, a method of comparison between the routes generated by the algorithms of Baaj & Mahmassani, and Ceder & Wilson is used, with the purpose of looking for the amount of optimal routes within each algorithm and comparing them: between algorithms and then with the existing routes. Since the routes generated by both algorithms are different and without a pattern, each one can be used as a parent or mixed for the application of the genetic algorithm to later analyze the results obtained. The two initial route generating algorithms reduce the travel time by 8.80 % and 9.05 % respectively, and the travel distance by 17.63 % and 17.69 % respectively of the existing routes in the Querétaro metropolitan area.

Keywords: optimization, transport bus network, initial routes, algorithms.

RESUMEN

En los últimos años, los algoritmos genéticos han sido una tendencia para encontrar soluciones a problemas complejos, tales como el diseño y la optimización de redes de transporte. Durante la optimización del diseño de la red de autobuses de transporte público mediante algoritmos genéticos, la forma de generar las rutas iniciales nos deja con muchas dudas, por ejemplo, la cantidad de rutas y cómo generarlas antes de realizar el proceso de cruce y mutación para llegar a la red óptima. En este trabajo, se utiliza un método de comparación entre las rutas generadas por los algoritmos de Baaj y Mahmassani y Ceder y Wilson, con el propósito de encontrar y contrastar las rutas óptimas entre los algoritmos y, a su vez, compararlas con las rutas existentes. Como las rutas generadas por los diferentes algoritmos son diferentes y sin un patrón, se puede usar cada uno como padre o mezclarlos para la aplicación del algoritmo genético, para después analizar los resultados obtenidos. Los dos algoritmos generadores de rutas iniciales reducen el tiempo de viaje en 8.80 % y 9.05 % respectivamente, y la distancia de viaje en 17.63 % y 17.69 % respectivamente para las rutas existentes de la zona metropolitana de Querétaro.

Palabras clave: optimización, red de autobuses de transporte, rutas iniciales, algoritmos.

INTRODUCTION

Human beings have basic needs, one of them is mobility, which allows the individual to move from one point to another to carry out their daily tasks. Depending on these activities, purposes and hours, there are trips to meet these needs [1].

This research work is carried out considering the problems that exist when creating a public transport network in a metropolitan area, and when deciding the methods and tools that will be used to achieve the effective layout of this network. The task of providing optimal composite structures is becoming increasingly difficult. Lately, one of the trends is the use of algorithms to design and optimize public transport networks.

At the time of deciding the tools, authors [2], [3], [4] and [5] suggest the use of heuristic and metaheuristic algorithms as a genetic algorithm in order to optimize and design an efficient network. The application of genetic algorithms to solve compound optimization problems is increasing, especially in the last five years. The use of this algorithm raises doubts for the initial population generation, which is one of the basic components of this algorithm [6]. In this part, it only generates initial routes, but not a single criterion is established to determine the number of routes to be generated [7].

The purpose of this research is to determine the criteria for the generation of the initial population of routes to optimize public transport routes in the metropolitan area of Querétaro with the application of genetic algorithms.

Various algorithms are used to generate routes and transport networks, such as those of Pattnk aik *et al.*, where the ideal number of routes that will be in the solution is not easily known since this varies in each iteration; Krishna Rao *et al.*, that creates conflict by overlapping routes over short distances; Chien *et al.*, where the number of feasible routes increases dramatically with the number of links, therefore it is not treatable for a realistic urban network; and Fan and Machemechi, which assumes that the bus fleets are uniform and standard parameters are used. The methods of Baaj & Mahmassani, and Ceder & Wilson are chosen because they give better results when used as criteria for the generation of initial populations for the application of genetic algorithms, and because they reduce travel time and travel distance. Ceder & Wilson use an algorithm that follows a heuristic path and determines the optimal development in their network. Baaj & Mahmassani use an algorithm that has the advantage of adding new routes to meet new demands, in addition, it can reduce the search space. [8]

Gurevich presents algorithms as essential elements to process data. Many programs contain algorithms that specify the specific instructions that a team must carry out (in a specific order) to perform a specific task [9].

Optimization techniques are widely used in many different disciplines. They are best used to provide innovative solutions or to obtain information on complex problems. Evolutionary algorithms is one of the most popular categories of optimization techniques, especially in engineering design, as they are able to find solutions in large and complex search spaces. The genetic algorithms simulate the theory of evolution proposed by Darwin, they were developed by Holland and are robust stochastic processes that can be used to solve search and optimization through a fitness function or objective function. Genetic algorithms perform the search process in four steps: initialization, selection, crossing and mutation.

The genetic algorithm is considered one of the pioneering algorithms to be used for the problem; it is a general algorithm that is usually implemented in problem solving intractable Operational Research [6]. The highest proportion of algorithms used are genetic algorithms, which occupy 56 % of applications, showing the continued popularity of this type of optimization algorithm despite the increasing competition from other methods [10].

Genetic algorithms are based on the mechanisms of genetics and natural selection. As will be seen later, it is the only heuristic goal that works simultaneously with two sets of feasible solutions, which will consider them as individuals from a population that crosses, reproduces and can even mutate to survive [11].

In order to apply the genetic algorithm, the following five basic components are required [6]:

- A representation of potential solutions to the problem.

- A way to create an initial population of possible solutions (usually a random process).

- An evaluation function that plays the role of the environment, classifying the solutions in terms of their "aptitude".

- Genetic operators that alter the composition of the children that will be produced for the following generations.

- Values for the different parameters used by the genetic algorithm (population size, probability of crossing, probability of mutation, maximum number of generations, among others).

The choice of the initial population of individuals is important because it determines the speed of the algorithm by making the convergence towards the global optimum more or less rapid. This mechanism must be capable of producing a heterogeneous population of individuals.

Then we have to randomly generate individuals by making random draws and ensuring that the produced individuals respect the restrictions. If, on the other hand, a priori information on the problem is available, it appears to generate individuals in a particular subdomain to accelerate convergence [12].

For a problem NP-complete for optimizing operating costs of electricity, the genetic algorithms used give us different results that depend on the generation of the initial population and the type of genetic algorithm used [13].

In this study, the genetic algorithm that was used searches for the minimum number of solutions for each generation of time. In generating the initial population, the author defines a maximum number of solutions and a minimum number but does not specify the reason why he chose these numbers before determining the objective function, the ratio of mutations, the parameter p [14].

This article mentioned 321 articles published between 2008 and 2017, 17 magazines covering structures and composite materials are reviewed to understand the technical challenges associated with the use of the genetic algorithms and determine the implications for the future. Also included is a detailed review of the genetic algorithms in the evolutionary computing literature, with a description of their mechanisms and a prediction of their suitability for different compound optimization problems [10].

The analysis is available on an attached data sheet. A number of problems are presented in the literature and can be summarized as follows:

- Poor documentation of the genetic algorithms used.
- Little focus on categorizing optimization problems.
- A small range of genetic algorithms in computer science literature are compared.
- A focus on simplifying problems.

Importantly, there is no documentation linking genetic algorithm mechanisms for the type of problem, making it impossible to solve difficult and exciting problems of interest to the industry. Therefore, a series of recommendations are listed in this review, categorized according to problems found in the literature, to help to improve compound optimization. They are summarized as follows [10]:

- A need for accurate documentation of parameters and mechanisms in each document.
- Treating genetic algorithms as a specialization, using the latest modern algorithms with sensitive parameters.
- Conducting rigorous benchmarking of different algorithms to help characterize problem solving and generate a body of best practice.
- Solving problems with multiple objectives, reducing to a single objective and weighted objective instances.

Together with an analysis procedure and an improvement algorithm, this algorithm constitutes one of the three main components of an AI-based hybrid solution approach to solving the transit network design problem [15]. Such a hybrid approach incorporates the knowledge and experience of transit network planners and implements efficient search techniques using AI tools, algorithmic procedures developed by others, and tool modules implemented in conventional languages. The route generation algorithm implemented by Lisp (RGA) is a design algorithm that:

- Is strongly guided by the demand matrix.
- Allows the designer's knowledge to be implemented to reduce search space.

- Generates different sets of routes corresponding to different compensations between conflicting objectives (user and operator costs).

Due to the large number of external and operational factors involved in the design of a bus network (for example, financial, socio-economic, and political, etc.), it is desirable to establish a planning process that incorporates alternative levels of complexity. The acceptability of this process due to the properties of the bus depends on its simplicity, flexibility and practicality [16].

A two-tier methodological approach is proposed in this section, based on five main objectives:

- Developing an algorithm for optimal network design so that routes provide service between all origin and destination pairs and meet operational and service constraints on travel, transfer and wait times.
- Developing performance measures from the passenger, operator and community perspectives.
- Combining other operations planning components (schedule construction and vehicle scheduling) with network design procedures.
- Developing sensitivity analyzes to determine solution tolerances due to possible demand changes and constraints.
- Developing an interactive human-machine system so that the planner can change constraints or routes during the design process.

MATERIALS AND METHODS

Generation of initial routes with Baaj and Mahmassani algorithms.

In this stage of population generation, the routes were built sequentially as mentioned by authors [7] and [17]. To start the generation of routes, an initial node was determined and from this node the succeeding node would be selected. The process continued until nodes should be selected or the individual was limited by the predefined maximum path length. The starting node and the selection of the following nodes were random or probability-weighted according to the demand that the selection could cover. The steps were performed using the hybrid programming algorithm in C, and processed in route generation TransCAD implemented by Baaj and Mahmassani [15] for the initial generation of routes. The proposed methodology operates based on the generation, evaluation and improvement of routes. Initially, a set of routes is generated considering the origin-destination matrix as the main guide, and the two shortest paths are found among a subset of M pairs of high-demand nodes, considered in decreasing order of value. An input parameter specifies the proportion of demand that can be left unsatisfied. Additional nodes are inserted into this initial route skeleton, according to pre-established rules. The generation procedure is repeated with varying parameters to obtain solutions to different commitments between objectives. The main rule for assigning demand is the transfer minimization criterion; for each pair (*i*, *j*) of nodes it is checked if it is possible to travel without transfer.

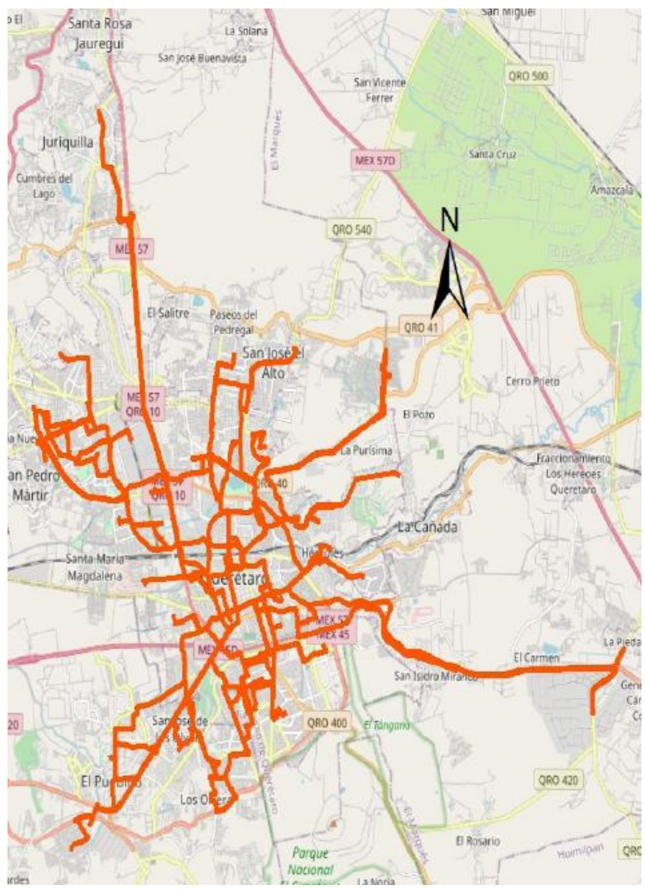

Figure 1. Some routes generated with the Baaj & Mahmassani algorithm.

Generation of initial routes with Ceder and Wilson algorithms.

To generate routes in this step, the algorithm developed by Ceder and Wilson, with the system existing streets and one origin-destination matrix as input data; the existing bus stops collected data in the Institute of Transport of Queretaro (IQT). This algorithm allows the analysis of external and operational factors and is developed in C and processed in TransCAD as a macro.

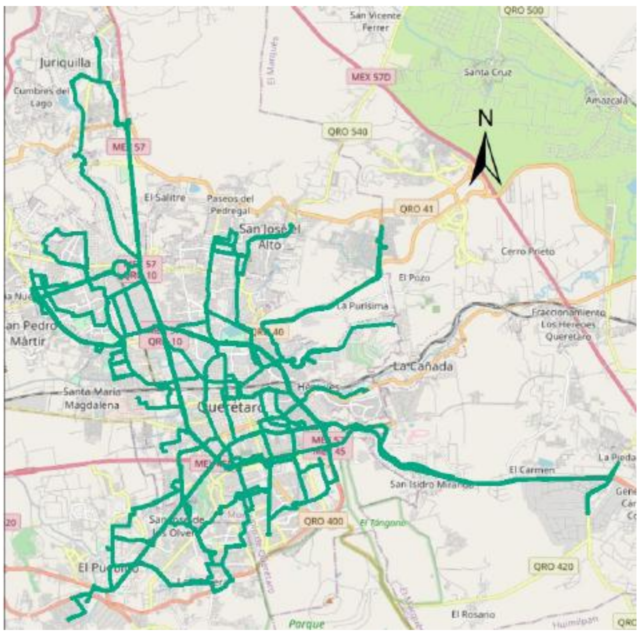

Figure 2. Some routes generated with the Ceder & Wilson algorithm.

Comparison of optimal results with existing network routes.

At this stage, data from the existing transport network in table 1 was analyzed, as well as the current distribution and the demand for trips; then it was compared with the results of the network obtained between the algorithms of Baaj & Mahmassani and Ceder & Wilson. This helped us to confirm the algorithm and the number of routes necessary to optimize the network with the genetic algorithm.

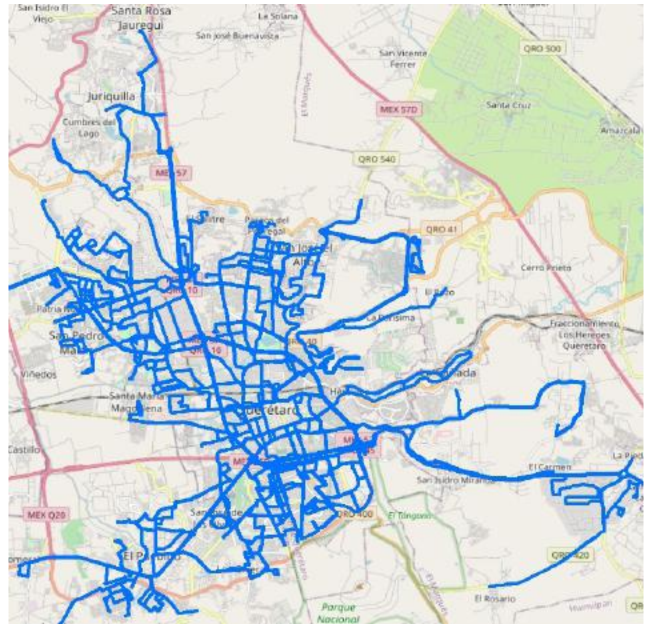

Figure 3. Map of Routes of the existing IQT network.

RESULTS AND DISCUSSION

Table 1 shows the lengths of the routes and the travel time of some routes generated with the algorithms of Baaj & Mhamassani and Ceder & Wilson that we can compare with the lengths and the travel time of the existing routes in the metropolitan area of Querétaro. The lengths of the routes generated with the two algorithms vary randomly for each route without following a variation pattern and, in most cases, they are shorter than the lengths of the existing routes. The travel time follows the random variation of the lengths of the routes, although they are not proportional because the difference in length can be very small and the travel time varies greatly. Baaj & Mahmassani's algorithm minimizes total passenger transfer times and the required fleet size, subject to frequency, load factor and fleet size restrictions. The main aspects of the problem are taken into account, as well as a variety of parameters and restrictions (load factor, for example). It is flexible, as it allows the incorporation of user knowledge; for example, restrictions on the minimum proportion of demand covered based on trips without transfers or with at least one transfer can be added when applying a resolution method; however, its main limitation is that it does not propose a systematic way of varying the parameters to generate different solutions. The Ceder and Wilson algorithm solves the problems of designing routes and schedules simultaneously, and of non-linear mathematical programming with mixed variables and multiple objectives (minimization of travel times and minimization of fleet size). The algorithm of these authors has the advantage of providing a certain degree of interactivity to define some restrictions and parameters; it is flexible due to its modularity, it allows both medium and long term planning. Considering the results in Table 2, the two algorithms offer considerable percentages of reduction in a matter of time and average travel distance. The total time and total travel distance after the application of the algorithms decrease compared to the existing network and have a value. For Baaj & Mahmassani, Ceder & Wilson and IQT, respectively: 5867.00, 5863.00 and 9945.00 min; and 1938.98, 1933.68 and 2968.38 km. The average time and the average distance of the algorithms and the existing network are closer but there is a reduction, expressed as a percentage for each algorithm: 8.80 % and 9.05 % respectively for the distance; and 17.63 % and 17.69 % for the time. Baaj & Mahmassani's algorithm takes less execution time than Ceder & Wilson's algorithm. In a matter of overlapping routes, coverage based on the kilometers traveled by the routes, the Baaj & Mahmassani's algorithm has a better result because its coefficient is higher.

Table 1. Length and travel time of some routes generated with the two algorithms and the existing routes.

	IQT		Ceder y Wilson		Baaj y Mahmassani	
Número de ruta	Longitud (Km)	Tiempo de recorrido (min)	Longitud (Km)	Tiempo de recorrido (min)	Longitud (Km)	Tiempo de recorrido (min)
5	33.592	167	33.0898	100	33.1262	110
7	35.967	146	35.986	118	34.893	105
9	37.754	118	36.513226	110	35.9764	100
10	55.989	174	45.555031	136	46.8765	145
12	46.289	184	32.944081	99	32.5567	96
13	21.312	75	24.9875	70	25.061712	73
14	32.475	131	34.0765	94	32.5076	97
17	46.375	159	45.7843	125	47.949582	136
19	39.817	147	49.5674	148	49.27899	145
20	40.142	145	40.5654	125	39.332738	118
21	46.631	186	22.0543	70	21.701156	65
24	44.468	148	43.9867	130	43.18119	126
27	52.097	256	48.5786	144	49.47993	148
28	45.879	149	43.1224	160	42.515297	155
29	32.534	124	32.6554	97	33.402882	100
31	42.913	140	42.1324	128	41.590152	124
33	17.636	79	16.7654	48	17.640682	52
36	38.889	145	39.6543	117	38.348066	114
37	40.046	156	38.987	116	40.095405	120
38	31.605	124	33.5134	102	31.696159	95

Table 2. Summary table of the two algorithms.

	Baaj and Mahmassani algorithm	Ceder and Wilson's algorithm	IQT
Total travel distance (Km)	1938.98	1933.68	2968.38
Average travel distance	36.58	36.48	40.11
% reduction	8.80 %	9.05 %	
Total travel time	5867	5863	9945
Average travel time	110.7	110.62	134.39
% reduction	17.63 %	17.69 %	
Km of routes	747.94	747.94	747.94
Route overlap	2.5924272	2.58534107	3.96874081
Computing time (min)	43	174	

CONCLUSIONS

Baaj & Mahmassani's algorithm generates shorter routes between pairs of high demand nodes, it is an exhaustive algorithm and generates many feasible routes. It also improves route merging and splitting but there is no parameter domain exploration. Ceder & Wilson's algorithm generates a set coverage and improves local search with cycle prevention. Each of the algorithms used has its advantages, the Ceder & Wilson's algorithm has a better application in terms of satisfying demand and users of the service; and the Baaj & Mahmassani's algorithm has a better application in terms of the time the user spends in a unit and in the optimization of the travel distance. In this study, the results obtained are better for Ceder and Wilson's algorithm because the reduction percentages are better than those of Baaj & Mahmassani's, but these percentages are quite close to each other, which leads to say that the generation of the initial population can be done with the two algorithms and then apply the genetic algorithm and compare the results to be obtained.

ACKNOWLEDGMENT

The authors thank the Consejo nacional de ciencia y tecnología and the Autonomous University of Querétaro, especially for coordinating the master's degree in landways and mobility engineering and their support throughout this research.

REFERENCES

[1] S. A. Obregón and E. Betanzo, "Urban mobility analysis of a mexican middle city, case study: Santiago de Querétaro," *Economía, Sociedad y Territorio,* vol. 15, no. 47, pp. 61-98, 2015.

[2] A. T. Buba and L. S. Lee, "A differential evolution for simultaneous transit network design and frequency setting problem," *Expert Systems With Applications,* vol. 106, pp. 277-289, 2018.

[3] O. J. Ibarra, F. Delgado, R. Giesen and J. C. Muñoz, "Planning, operation, and control of bus transport systems: A literature review," *Transportation Research Part B: Methodological,* vol. 77, pp. 38-75, 2015.

[4] L. Ahmed, C. Mumford and A. Kheiri, "Solving urban transit route design problem using selection hyper-heuristics," *European Journal of Operational Research,* vol. 274, no. 2, pp. 545-559, 2019.

[5] A. Fieldbaum, S. Jara and A. Gschwender, "Optimal public transport networks for a general urban structure," *Transportation Research Part B: Methodological,* vol. 94, pp. 298-313, 2016.

[6] J. Allen, J. C. Muñoz and J. d. D. ortúzar, "Modelling service-specific and global transit satisfaction under travel and user heterogeneity," *Transportation Research Part A: Policy and Practice,* vol. 113, pp. 509-528, 2018.

[7] S. J. Berrebi, K. E. Watkins and J. A. Laval, "A real-time bus dispatching policy to minimize passenger wait on a high frequency route," *Transportation Research Part B: Methodological,* vol. 81, no. 2, pp. 377-389, 2015.

[8] R. Ortega, "Comparativa de Algoritmos Empleados en la Optimización de Sistemas de Transporte Público Urbano," Master thesis, Universidad Autónoma de Querétaro, Queretaro, Mexico, 2019.

[9] Y. Gurevich, "The sequential ASM thesis," *Bulletin of European Association for Theorical Computer Science,* vol. 67, pp. 93-124, 1999.

[10] Z. Wang and A. Sobey, "A comparative review between Genetic Algorithm use in composite optimisation and the state-of-the-art in evolutionary computation," *Composite Structures,* vol. 233, pp. 1-51, 2019.

[11] M. Zouita, S. Bouamama and K. Barkaoui, "Improving genetic algorithm using arc consistency technic," *Procedia Computer Science,* vol. 159, pp. 1387-1396, 2019.

[12] I. Vlasic, M. Durasevic and D. Jakobovic, "Improving genetic algorithm performance by population initialisation with dispatching rules," *Computers & Industrial Engineering,* vol. 137, pp. 2-38, 2019.

[13] P. Cortés, J. Muñuzuri, M. Berrocal-de-O and I. Domínguez, "Genetic algorithms to optimize the operating costs of electricity and heating networks in buildings considering distributed energy generation and storage," *Computers and Operations Research*, vol. 96, pp. 157-172, 2018.

[14] G. Ahn and S. Hur, "Efficient genetic algorithm for feature selection for early time series classification," *Computers & Industrial Engineering*, vol. 142, no. 106345, pp. 1-16, 2019.

[15] M. H. Baaj and H. S. Mahmassani, "Hybrid route generation heuristic algorithm for the design of transit networks," *Transportation Research Part C: Emerging Technologies*, vol. 3, no. 1, pp. 31-50, 1995.

[16] A. Ceder and N. H. M. Wilson, "Bus network design," *Transportation Research Part B: Methodological*, vol. 20, no. 4, p. 331–344, 1986.

[17] M. Owais and M. K. Osman, "Complete hierarchical multi-objective genetic algorithm for transit network design problem," *Expert Systems with Applications*, vol. 114, pp. 143-154, 2018.