Analysis and development of a new method for defining Path Reliability in WebGIS based on Fuzzy Logic and Dispersion Indices

Walter Balzano, Antonio Lanuto, Erasmo Prosciutto, and Biagio Scotto di Covella

Abstract When planning a trip, we wish to receive a precise itinerary, taking into account various factors such as traffic, distance, types of roads. However, we often have to deal with unreliable and inaccurate information. Evaluating the reliability of a route is a crucial aspect to improve the quality of navigation services and guarantee the user an experience that is not only effective but also efficient in terms of cost and travel time. The problem was approached with an innovative solution, which uses fuzzy logic and dispersion indices to measure variations in the traffic situation at different times of the day and on different days of the week, with tests carried out on real routes collected by the Google Maps platform API.

Key words: Path Reliability in WebGIS; Fuzzy Logic; Dispersion Indices

1 Introduction

The advent of mobile devices and real-time navigation applications such as Google Maps has made navigation an essential part of daily life. However, one of the main limitations of these applications is the absence of a *reliability* attribute for a given route. Currently, navigation applications only provide an estimate of travel time, which can cause problems for users who need to plan their time, especially in situations where time is limited or when there are delays.

To ensure the safety and punctuality of users, it is critical to define the *reliability* of a route. A more reliable route is less prone to unexpected traffic

University of Naples, Federico II
Balzano W. e-mail: walter.balzano@gmail.com · Lanuto A. e-mail: antoniolanuto@outlook.it · Prosciutto E. e-mail: erasmo.prosciutto@gmail.com · Scotto di Covella B. e-mail: biagioscottodicovella@gmail.com

delays or traffic jams, which means users will arrive at their destination more accurately than a less reliable route. In addition, the definition of route *reliability* can be used to determine the safety of the route itself, such as avoiding roads with a high accident rate.

The travel time estimation provided by navigation applications does not take into account route variability. Two alternative routes might have the same travel time estimate, but one might have greater variability than the other. This means that a route with a higher *reliability* might be preferred over one with a lower value because it offers greater safety and punctuality.

The following sections are divided as follows: in section 2 the background and the related works are described; in section 3 the proposed solution is presented with an in depth look at the structure of the algorithm; Section 4 deals with the evaluation with the results of the simulation of the algorithm execution in the city of Naples; in section 5 some future developments are introduced; finally in section 6 the conclusions are explained.

2 Background

Current research is focusing on analyzing the *reliability* of routes in transportation networks, using measures such as the probability of arriving at the destination on time. To do this, scholars have developed algorithms based on stochastic networks [2], which represent a type of mathematical modeling that uses probability to describe dynamic systems. In essence, these networks use a graphical representation to describe the relationships between system variables and their transition probabilities over time.

However, these algorithms have some limitations. First, they require a large number of calculations to determine the transition probabilities, especially for systems with a large number of possible states and transitions. This can make them difficult to use for large or real-time problems. Second, they may have difficulty handling imprecise or non-quantitative uncertainties, and they may provide inaccurate results if the data change.

Despite this, these algorithms mainly focus on the *reliability* of route time estimation, using algorithms based on historical data and statistical analysis. However, route *reliability* studies are still under development and have not yet been widely adopted in GPS navigation because they cannot predict extraordinary events or traffic conditions that may affect the actual travel time.

2.1 Related Work

The problem of road traffic congestion is still one of the most important challenges that many researchers are working on (see [1] for a survey), and there are many studies on how to optimise the distribution of vehicles on city streets, an example of which is the [4], where the problem is addressed by posing an isomorphism with resource management in operating systems, or in [3] were an innovative hierarchical model, based on two abstraction levels, prevents the traffic congestion. Moreover, as the population increases over the years, as seen in [14], so does the number of vehicles.

The problem of road congestion is closely linked to the problem of route *reliability*, that has been addressed several times from different aspects [13]. Many of these systems use stochastic networks to create mathematical modelling that uses probability to describe dynamic systems [12, 5].

Another interesting aspect to note is that analyzing the *reliability* of routes in transportation networks also shares methodological parallels with advancements in other computational fields. For instance, similar to the use of stochastic networks in route reliability, Quantum Natural Language Processing (QNLP) also represents an innovative blend of complex theoretical models with practical applications. In QNLP, the principles of quantum mechanics are applied to language processing tasks, demonstrating the integration of advanced theoretical frameworks with real-world problems, akin to our approach in route analysis using stochastic networks [7].

Moreover, the challenge in route *reliability* of handling imprecise or non-quantitative uncertainties is reminiscent of the challenges addressed in the multilingual BERT (mBERT) language model study. The mBERT research investigates cross-lingual syntactic transfer, revealing how complex syntactic knowledge in one language can be applicable in other languages. This kind of cross-domain application of theoretical models is analogous to our approach in route *reliability*, where we deal with the variability and unpredictability of real-world traffic conditions [8].

In conclusion, the widespread use of fuzzy logic in these fields such as, routing algorithm [6], shortest path [10], has led us to the use of such technology to better address the *reliability* problem.

2.2 The key role of Variance for reliability

Variance is a widely used statistical measure in computer science. It allows us to represent the dispersion of data with respect to the mean, giving us an estimate of the variability of the data. For task management of a processor, scheduling algorithms use *variance* to evaluate adaptivity. The goal is to minimize the waiting time for task execution.

Choosing a scheduling algorithm with a lower *variance*, even if it results in a slight increase in average wait time, means opting for a more adaptive solution. In this way, tasks will run more smoothly and predictably, which is essential for optimal utilization of processor resources.

Variance plays a key role in the design of processor scheduling algorithms. This has been demonstrated by several studies, including [9] and [11] . *Variance* minimization can improve the quality of service (QoS) of systems involved in task scheduling, and it can significantly affect system performance and user satisfaction. Therefore, consideration of *variance* in the design of scheduling algorithms is a crucial aspect of ensuring stable and predictable performance.

3 Proposed Solution

In the previous section, we explored how *variance* plays a key role for process scheduling algorithms in terms of *reliability*. Similarly, the proposed system uses fuzzy logic to define the *reliability* of a path based on the *standard deviation* (i.e. the mean square deviation) of the travel time estimates. It represents precisely the dispersion of values around the mean and can be used as a measure of the *variance* of the data. The mean square deviation is equal to the square root of the *variance*, which allows the increasing and decreasing characteristics of the original function to be maintained, thus preserving information on the properties of the data distribution. Therefore, using the *standard deviation* to assess the *reliability* of transport routes is a consistent and reasoned choice, as it allows us to describe the dispersion of travel times with respect to the mean, and to assess the regularity of the route. This means that the system will be able to assess its variability, providing a more complete view of the route. In this way, it will be possible to more consciously choose the most suitable route to take, rather than opting for the one with the lowest estimated travel time.

We aim to highlight the significance of route *reliability* by employing Google APIs to build a comprehensive data-set stored in a database. The data-set will track traffic patterns on different routes during various times of the day. Although the starting and finishing points of each route will be the same, each route will differ in terms of its trajectory and traffic. We will then perform a process of converting the data into fuzzy logic for each route, analysing the *standard deviation* from the average travel time. These results will be represented using graphs to facilitate visualisation and better understanding of the information. This approach will help us to develop a more accurate understanding of route *reliability* and enable us to identify key factors that impact travel time.

3.1 Path Reliability Estimation Algorithm

In this section, we will take an in-depth look at how the *Path Reliability Estimation Algorithm* works. The algorithm consists of two main parts, each of which plays an important role in data processing:

1. **The route reliability storage part -** Here we will define the structure of the database that handles the *reliability* estimation data. It will also be important to analyse the manipulation of this data, together with the management of the *WebGIS* calls needed to derive the estimates.
2. **The data processing part -** It plays a crucial role in processing the data collected in the first part. Using fuzzy logic, we will process the data to obtain a more accurate estimate of path *reliability*. Finally, we will transform the processed data into graphs for a more intuitive visualisation.

3.1.1 Route Reliability Storage

In this section, we will delve into the process of data collection for determining route *reliability* estimates. To do so, we will examine two fundamental structures that govern this process:

- **The first structure** is a simple custom-built database that stores the routes to be analyzed along with some crucial values. Each route has a base time in the absence of traffic, and is associated with a data set that takes into account the estimated values of time with traffic on different days. These values will allow us to subsequently analyze the *reliability* of each route.
- **The second structure** is a Python file that utilizes the Google Maps *Distance Matrix* API to store the values of duration and expected duration in traffic for a given route. This part of the algorithm retrieves the data received from Google and processes it in the format required by our database. This operation is repeated every 5 minutes in the background.

3.1.2 Data Processing

The algorithm implements a fuzzy system to assess the *reliability* of a travel route, considering two main factors; *distance* and *standard deviation* percentage from the average estimate of travel time as a function of time. Using these values, we can construct three models as a function of *distance*, *standard deviation* and route *reliability*.

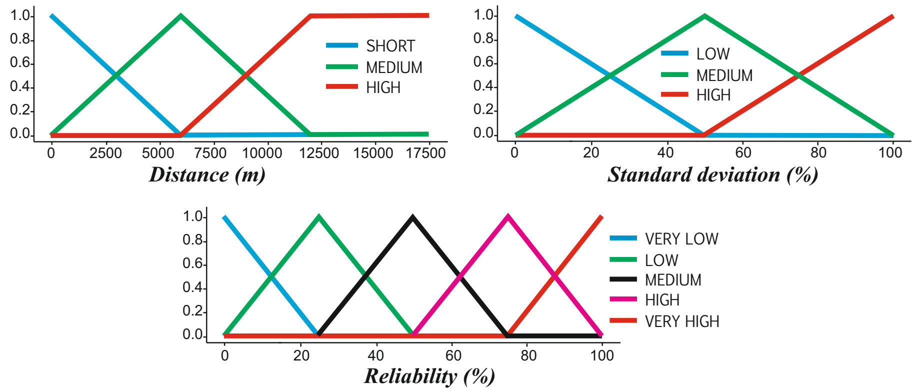

Fig. 1 Linguistic variables of the fuzzy system.

As can be seen in the figure 1 each of the first two characterise the main *reliability* model by going to base our inference engine on the variability of the data obtained. The results of the algorithm are then subsequently aggregated, using the centre of area (COA) method. Finally, graphs are generated to facilitate data analysis.

4 Evaluation

In this section we want to provide some simulation results of the algorithm execution, as well as to analyse the different results obtained on different routes in the city of Naples. In order to obtain a simulation faithful to a real situation of path *reliability*, we chose four paths with different trajectories all starting and arriving from the same point. We analysed these four routes for thirty days, for a total of over **300,000 records**, which allowed us to analyse certain points in detail.

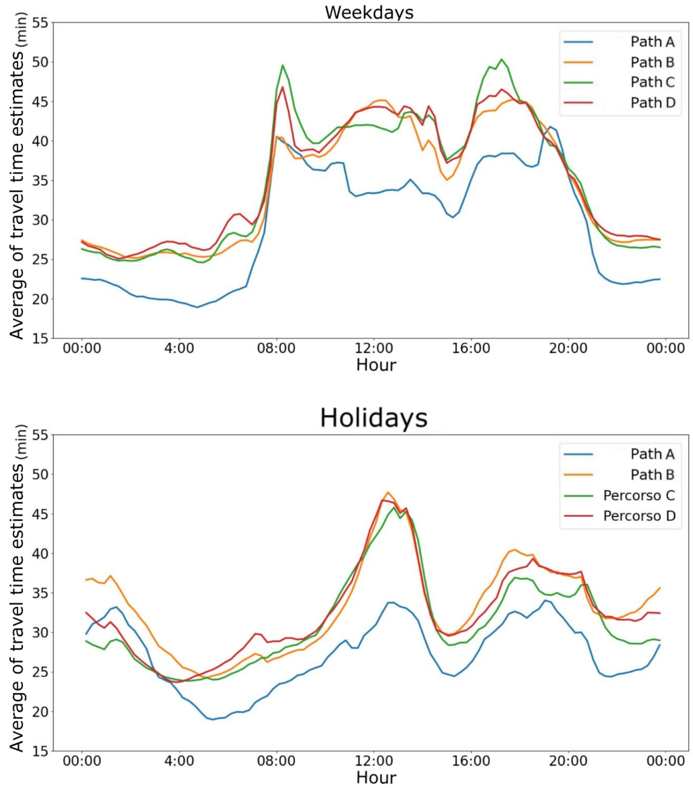

Fig. 2 Average travel times in weekdays and holidays.

As can be seen from the graphs of the *average travel time* estimates, Figure 2, **Route A**, despite being the route with the maximum traffic-free duration, is definitely *the most advantageous*, since it maintains an average that is almost always at least *five minutes less than the other routes in all time slots*, both on weekdays and holidays, with the gap reaching as much as fifteen minutes at 1:00 PM local time on holidays.

Given the unpredictability of variables such as traffic, bad weather, traffic accidents, and so on, it is not possible to be certain of arriving in the times indicated by the Google Maps estimate. In this sense, the *standard deviation* graphs in Figure 3 showed that although in some places one route may be better than another in terms of *average travel time* estimates, it recorded a higher percentage of *standard deviation* from the average. This means that the variability of this route is higher than that of the others.

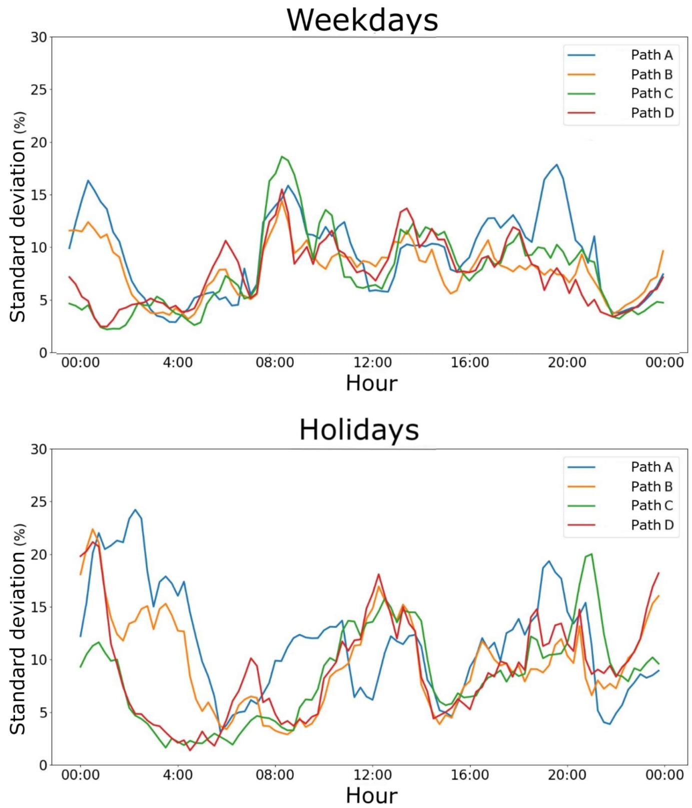

Fig. 3 Standard deviation in weekdays and holidays.

In the images, it can be seen that the *standard deviation* trend is much more erratic than that of the *average travel time* estimates, synonymous with the fact that for a more conscientious choice of which route to take, we cannot consider only its estimate. In fact, **Route A** records a *standard deviation* percentage of **25%**, far higher than the others, on holidays at about 02:30 AM. It is most likely that this is due to the large amount of traffic on "Via Caracciolo" at that time; for that matter, the *standard deviation* percentages of **Route C** and **Route D** turn out to be **5%** as they pass through "Via Manzoni."

It is also possible to observe that **Route A**, just before 8:00 PM, has a *standard deviation* percentage that is *almost twice that of the other routes.*

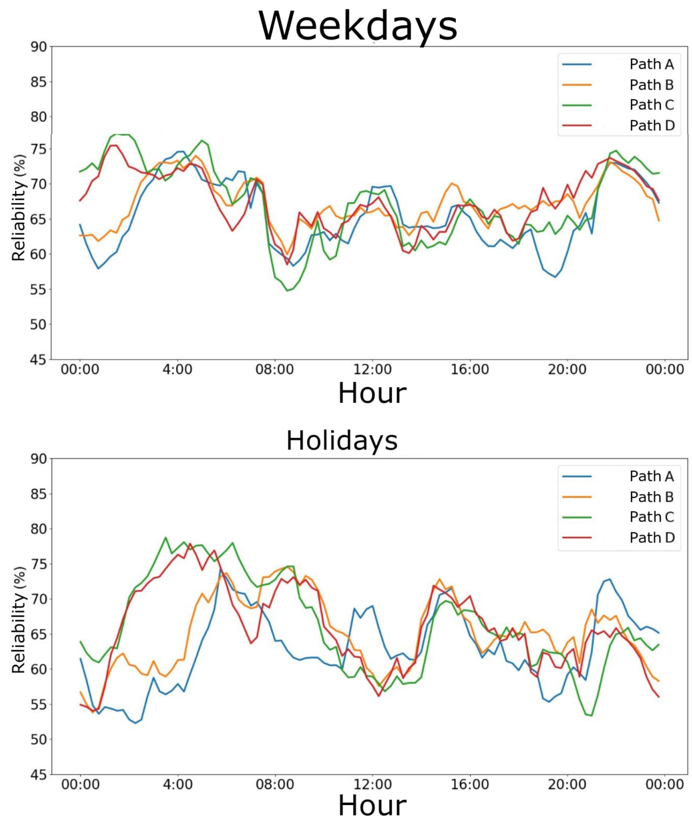

Fig. 4 Routes reliability in weekdays and holidays.

In the *reliability* graphs of the Figure 4, what has been said so far about the variability of the routes studied is evident: *greater variability resulted in lower reliability and vice versa*. From this point of view, if one were at 4:00 AM on a public holiday and had the four routes analyzed, it would be preferable to choose **Route C**, despite it having a higher *travel time estimate* than **Route A**, as the *reliability* values are about **80%** and **57%**, respectively.

It is evident from the results that there is no unambiguous answer to whether it makes sense to choose a route with lower travel time estimate and higher *standard deviation*, however, *more tools can be provided to the user so that he or she can choose the route to take more conscientiously.*

5 Future Works

Regarding the future evolution of the application, there are several possibilities for improving the efficiency and effectiveness of its functionality:

- **Implementation of a more universal and versatile fuzzy logic** that can be applied in contexts other than urban settings. This would broaden the utility of the application and make it better suited to the needs of a wide range of users.
- **Optimization of the source code** in order to improve the performance and the scalability of the application. This could involve the introduction of more efficient programming techniques or the use of external libraries to simplify the development process and improve software efficiency.
- **Implementation of the application on real *WebGIS*** such as Google Maps and Waze. In particular, it would be important for these services to add a new attribute that defines the adaptability of a route, similar to how they implemented the attribute that defines which route is less fuel-intensive. This would allow this functionality to be integrated into applications already used by millions of users, improving the user experience.

6 Conclusions

The objective of this paper is to demonstrate that *the choice of a route to follow should not be based solely on the estimated travel time proposed by Google Maps*. In fact, in addition to the duration of the route, it is also important to consider its variability, i.e., its ability to be predictable and not undergo unpredictable changes over time. In fact, due to factors such as traffic, bad weather and traffic accidents, the estimate provided by the *WebGIS* may be inaccurate and the proposed route may be less reliable than it initially seems. The goal is to provide the user with tools to make a more informed choice, taking into account the variability of travel time and its *reliability*. In this way, the user will be able to choose the route best suited to his or her needs, even taking into account of any delays or unforeseen events that could affect the duration of the route.

References

1. Tanzina Afrin and Nita Yodo. A survey of road traffic congestion measures towards a sustainable and resilient transportation system. *Sustainability*, 12(11):4660, 2020.

2. A. Arun Prakash. Algorithms for most reliable routes on stochastic and time-dependent networks. *Transportation Research Part B: Methodological*, 138:202–220, 2020.
3. Walter Balzano, Antonio Lanuto, Carmine Mascia, and Silvia Stranieri. Hierarchical vanet: A traffic congestion management approach based on critical points. In Leonard Barolli, editor, *Advances on P2P, Parallel, Grid, Cloud and Internet Computing*, pages 192–202, Cham, 2023. Springer International Publishing.
4. Walter Balzano, Erasmo Prosciutto, Biagio Scotto di Covella, and Silvia Stranieri. A resource allocation technique for vanets inspired to the banker's algorithm. In Leonard Barolli, editor, *Advances on P2P, Parallel, Grid, Cloud and Internet Computing*, pages 222–231, Cham, 2023. Springer International Publishing.
5. Rajesh S Prabhu Gaonkar, Akshay V Nigalye, and Sunay P Pai. Possibilistic approach for travel time reliability evaluation. *International Journal of Mathematical, Engineering and Management Sciences*, 6(1):223, 2021.
6. Golnoosh Ghalavand, Azadeh Ghalavand, Arash Dana, and Mahnaz Rezahosieni. Reliable routing algorithm based on fuzzy logic for mobile ad hoc network. In *2010 3rd international conference on advanced computer theory and engineering (ICACTE)*, volume 5, pages V5–606. IEEE, 2010.
7. Raffaele Guarasci, Giuseppe De Pietro, and Massimo Esposito. Quantum natural language processing: Challenges and opportunities. *Applied Sciences*, 12(11):5651, 2022.
8. Raffaele Guarasci, Stefano Silvestri, Giuseppe De Pietro, Hamido Fujita, and Massimo Esposito. Bert syntactic transfer: A computational experiment on italian, french and english languages. *Computer Speech & Language*, (71/101261), 2022.
9. Anas Jebreen. On minimum variance cpu-scheduling algorithm for interactive systems using goal programming. *International Journal of Computer Applications*, 135:51–59, 02 2016.
10. Esmaile Keshavarz and Esmaile Khorram. A fuzzy shortest path with the highest reliability. *Journal of Computational and Applied Mathematics*, 230(1):204–212, 2009.
11. Satyasundara Mahapatra, Rati Ranjan Dash, and Sateesh Kumar Pradhan. A heuristic job scheduling algorithm for minimizing the waiting time variance. In *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, pages 446–451, 2015.
12. Feng Pan and David P Morton. Minimizing a stochastic maximum-reliability path. *Networks: An International Journal*, 52(3):111–119, 2008.
13. Hesham A Rakha, Ihab El-Shawarby, Mazen Arafeh, and Francois Dion. Estimating path travel-time reliability. In *2006 IEEE Intelligent Transportation Systems Conference*, pages 236–241. IEEE, 2006.
14. Statista. Proportion of population in cities worldwide from 1985 to 2050, 2022.