

Increasing Efficiency of an Intelligent Semaphore by Implementing an ID3 Approach in Decision-Making

J. A. Castán Rocha¹, S. Ibarra Martínez^{1*}, J. Laria Menchaca¹, J. D. Terán Villanueva¹, M. G. Treviño Berrones¹, J. Pérez Cobos¹, E. Castán Rocha²

¹Faculty of Engineering, Autonomous University of Tamaulipas, Ciudad Victoria, Mexico ²TecNM/Instituto Tecnológico de Ciudad Madero, Ciudad Madero, Mexico Email: jacastan@uat.edu.mx, *sibarram@uat.edu.mx

How to cite this paper: Rocha, J.A.C., Martínez, S.I., Menchaca, J.L., Villanueva, J.D.T., Berrones, M.G.T., Cobos, J.P. and Rocha, E.C. (2018) Increasing Efficiency of an Intelligent Semaphore by Implementing an ID3 Approach in Decision-Making. *Journal of Computer and Communications*, **6**, 316-324.

https://doi.org/10.4236/jcc.2018.61030

Received: November 22, 2017 Accepted: January 27, 2018 Published: January 30, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

CC O Open Access

Abstract

The main objective of a semaphore is to provide a correct and fluent vehicular mobility. Many countries around the world are using such devices in urban areas. However, the traditional semaphore operative ways are outdated. We report in a previous contribution the development of a methodology capable of improving the vehicular mobility by proposing a new green light interval based on road conditions with a CBR approach. However, this proposal should include whether it is needed to modify such light duration. To do this, the paper proposes the adaptation of an inductive decision tree helping to decide when the semaphore should try to fix the green light interval according to specific road requirements. Some experiments are conducted in a simulated environment to evaluate the pertinence of implementing a decision-making before the CBR methodology. For example, using the ID3 approach the decisions of the system improve almost 8% in a set of 10,000 experiments. Finally, some conclusions are drawn to emphasize the benefits of including this technique in a methodology to implement intelligent semaphores.

Keywords

Urban Traffic Control, Intelligent Transport System, ID3 Approach, Vehicular Mobility

1. Introduction

The decision-making process of human beings is quite complex. Such complexity increases as the amount of information to be considered grows as well as the $\frac{1}{5}$. Ibarra Martínez is the corresponding author.

reliability of said information. To reduce the risk of these factors within the decision-making structure, in an intelligent system (e.g., an autonomous semaphore) there are different techniques such as fuzzy logic, neural networks or decision trees, which facilitate decision making. To do this, these techniques consider the parameters, attributes, constraints, and conjectures involved in generating the problem, so that the decisions are reached as humans do.

A traffic infrastructure is a dynamic and open scenario shaped by different kind of information which is used to make trustworthy and reliable decisions. In Mexico, the traditional way of controlling a junction is by using a traffic light signal. Such control was functional and suitable for the past. The number of cars and other relevant factors that affect the uses of the road are representing a technological challenge for researchers. In previous work [1] presented a methodology that makes an evaluation of the green light time based on the volume of cars to modify the duration of such light in the next phase of the semaphore. However, it was observed that sometimes the change in the green light duration was made and this change was not really significant to increase the level of service of the road. As a first conclusion, the fact of incurring in a non-adequate recommendation could cause both noise in the previously synchronized phase of the junction and increase the computational cost associated to the procedure of the system. The relevance of improving the way of adjusting the traffic light signals is quite needed due to current road constraints. For that reason, the paper proposes to include a prior review of some road aspects in order to decide if a modification of the green light interval could improve the service level of the junction. An ID3 method is evaluated in order to corroborate the above statement.

Inductive decision trees appear to be an adequate manner to response under great and inconstant data. One of the most used inductive learning methods in various test scenarios is the decision trees. The main advantage of these models is that they use both the positive and negative experience to approximate a concrete response. For example, reference [2] predicts traffic conditions taking into account data collected in AMAP, a system used in China for the collection of online maps. In order to achieve this, the authors use an analysis based on vector comparison, where the features of each situation are characterized and the decisions are made based on a criterion of division, such as the gain of information, the relationship of the gain and the Gini index proposed by Jiawei in [3]. Offering an adequate level of service in vehicular traffic is currently one of the most relevant aspects within the organization of a city. Not only to increase the productive level of the different sectors that converge in that a population, but also to create a healthier environment and conducive for citizens to find the comfort and security they so long for.

Reference [4] proposes a random search through the whole space of attributes in tree building, and they call the resulting model randomly selected decision tree (RSDT). Their model significantly reduces the total test cost and maintains a higher classification accuracy in comparison with other selected models. The experimental results on 36 UCI datasets validate the effectiveness of the proposed RSDT model over some of the compared models, the model closest to RSDT is C4.5. The authors believe that the use of more sophisticated multi-objective methods could further improve the performance of RSDT and make its advantages stronger.

The approach [5] describes a framework for learning Terminological Decision Trees from examples described in an ontological knowledge base to be used for performing instance classifications. To accomplish that, a suitable pruning strategies and a new prediction procedure are proposed. An extension of the framework named Terminological Random Forests is performed to tackle the class-imbalance distribution problem. In specific cases, the employment of pruning algorithms has proven its effectiveness to reduce the overfitting, especially for ontologies with a large number of primitive concepts. Experiments have shown that a REP and PEP pruning can be used for simplifying terminological decision tree models.

Reference [6] proposes an attribute selection method for nodes on the basis of the concept model of decision trees in purpose of avoiding the heuristic bias of attribute measurement and improving the performance of decision trees. The probabilistic statistics form is used to define and express the concept model extracted from the given data of things and created by associated certainty of classes distribution and branches distribution to fulfil certainty description of tree. And class constraint uncertainty (CCE) is used as a heuristic measure in the induction of tree to select the split attribute while the processing of the missing branch as an auxiliary leaf measure to construct a novel algorithm of decision tree learning. Experimental findings show that CCE is effective as a heuristic measure to avoid the bias in the selection of the multi-value attribute to all datasets and improve the performance and stability of the decision trees.

Based on a literature overviewed, there is evidence of many works oriented to provide effective approaches to solve well-detected lacks in transport area. The main conclusion of these works aims to continue the evaluation of studies on three specific subjects: 1) the interaction among autonomous semaphores; 2) the capability of the system to be robust under uncertain data; and 3) the reliability on the semaphore decisions. About such statement, the paper argues that the implementation of the support decision technique studied appears to be a suitable and robust way to decide when a semaphore must attempt to change the duration of a green light. Different variables can be implicated in this important and dynamic decision. However, ID3 method arises as a proper solution to be a filter which avoids that semaphore performs unnecessary actions. Therefore, the motivation of this paper is to strengthen the operatively of a novel methodology for controlling semaphores in an autonomous way using a CBR approach. The remainder of this paper is organized as follows: Section 2 introduces the specific aspects of our proposal. Section 3 presents the results obtained in the experiments. Finally, Section 4 emphasizes the main contributions of the paper and defines some ideas for further progress on this interesting domain.

2. Generation of the ID3 Approach

Modern cities are looking for offering a suitable traffic service level from different perspectives (*i.e.*, urbanism, town planning or computer science). The idea is not only to increase the productiveness of the city but also it is to reach a more comfortable, secure and healthy environment for all the citizens. Among all the particularities of this utopia, roads better administered are one of the most studied topics in the actuality. In this sense, the authors present in [1], a general methodology capable of evaluating some aspects on the road in order to modify the green light interval of a semaphore using a CBR approach. However, one data among the results demonstrate that in occasions modifying such interval is not the appropriate response of the system.

For example, in some cases, the system recommends aggregating 2 or 3 seconds to the green light interval and this action does not reflect an increment in the service level of the analyzed line. In spite of this, it is necessary to evaluate some aspects before of initializing the processes of the methodology. To do this, the authors propose to include an ID3 technique to define if it is suitable to do a modification in the green light duration of a determined semaphore. This method will allow the system to validate a set of attributes to determine, prior to the start of a green light, if according to the conditions of the road, it will be necessary modify the duration of such light. The attributes that were considered in this study are shown in **Table 1**.

Based on information presented in Table 1, and results previously presented in [1], 1000 combinations were extracted. Table 2 shows 5 examples about how can be established the relationship between the attributes and the decision. In this sense, taking into consideration the obtained data, the generation of the decision tree was continued, selecting as discrimination model the measure of entropy (see Equation (1)), known as the gain of the information (see Equation (2)), introduced in [3]. Each attribute generates its entropy value and this information allows ordering the attributes in a descending and interconnected way, in order to generate a suitable decision path for each situation. This making it clear that both positive and negative experiences allow increasing the efficiency of the process when discerning whether or not to modify the period of time assigned to the green light of a particular semaphore. As a reference, Table 3 shows the application of the entropy measure in the "number of cars" attribute. Besides, Figure 1 illustrates the scheme of this node. The blue circle represents the attribute of the tree and black and red boxes represent the variables of the attribute and the answer offered by the tree, respectively. A complete view of the reached decision tree is presented in Figure 2 which is generated by using WEKA 3.8.1 program. To do that, 10,000-reused data were processed.

$$S(E) = \inf(E) = -\sum_{j=1}^{k} p_j * \log_2(p_j)$$
(1)

$$gain(E, X) = info(E) - info_{attrib}(E, X)$$
(2)



Figure 1. Representation of the entropy of the "number of cars" attribute.

ID	ATTRIBUTES	VALUES
		0 - 2
aNC	Number of Core	3 - 15
anc	Number of Cars	15 - 30
		30 or more
		20 - 25
.CI	Crosse Light Time	26 - 30
aGL	Green Light Time	31 - 35
		36 or more
		Early Mourning (EM)
		Mourning (M)
aDT	Day Time	Midday (Md)
		Afternoon (A)
		Night (N)
- 77 D	T	Avenue (A)
alk	i ype of Road	Street (S)
		Week (W)
aTD	Type of Day	Weekend (Wd)
		Holiday (H)

Table 1. Attributes and values used in the model ID3.

Table 2. Example of 5 cases in the attribute-decision relation	ation
--	-------

No.	aNC	aGL	aTR	aDT	aTD	aHP	Answer
1	44	21	А	Md	W	Yes	YES
2	16	26	А	А	W	Yes	NO
3	49	20	S	А	Wd	No	YES
4	52	26	А	М	W	No	YES
5	29	30	А	М	W	No	NO



Figure 2. Decision tree ID3 using DIA program.

Node aNC	Events	Measure
None	1	0.00312
A few	6	0.00211
Some	7	0.98522
A lot	1	0.91829
Info_aNC		0.62112
Entropy_aNC		0.31295

Table 3. Application of ID3 in the quantity of cars attribute.

3. Experimental Results

All the experiments were developed on a Macbook Pro with the following specification:

- macOS Sierra OS10.12.6 version.
- Intel Core i7 2.90 GHz Processor.
- 8 GB 1600 MHz DDR3 memory.
- 1 TB SSD storage.

To validate the efficiency of the generated tree, a phase of 10,000 control experiments were performed each, taking as reference the representation of an infrastructure of four roads, one of the most used in Mexico (see **Figure 3**) with the following characteristics:

- Each road has movements in both directions.
- It can turn left (according to the traffic laws that are regulated in Mexico).
- Continuous right turn (*i.e.*, not necessarily the semaphore should have green light, simply the driver must be cautious).
- Each road has three lanes.
- Average speed of the car 50 km/h.
- Different types of cars (*i.e.*, sizes and shapes).

The results are compared against the performance obtained from the system when it modifies the green light time at all turns of the semaphore cycle. In this sense, if the system uses a decision support technique, such as the ID3 technique, its performance increases by almost 8% as shown in **Figure 4**.



Figure 3. The representation of an infrastructure of four roads used in the test.



Figure 4. Performance obtained when the system uses an ID3 decision tree to decide whether or not to modify the green light time.

Another interesting point of view to be noted is the number of experiments each one of the proposed approaches needed to reach stability in the overall system performance. In other words, this data reflects the number of situations that system requires to get sufficient information into their data base in order to reach trustworthy and reliable decisions. For instance, in **Figure 5** the computed curve using a sliding window up to the current trial with a smooth order 3 method for the performance of the system under different solution perspectives is presented. As can be noted, the system requires almost 2000 cases to reach a more stable behavior in order to ensure a suitable and constant mobility level.

4. Conclusions and Future Work

Modern computing provides robust and suitable decisions in an endless number of scenarios. Road control and administration is one of the most pursuit challenges in our days. In this light, the present paper introduces the use of an inductive decision tree to evaluate when it will be opportune to modify the duration of the green light in a certain semaphore to increasing the time of service offered in vehicular mobility.

The results show that using a decision support technique, such as the decision tree by induction technique, the system performance increases in almost 8%. Therefore, it is possible to identify that the use of this technique allows the system to modify the duration of the green light, improving its decision-making process which is reflected in an improvement of its operative capacity, its reliability and certainty.

It is important to remark that for experimental tests, the human factor had a vital role because an expert in traffic control recommends the variables consider in the structure of the tree. For future approaches, it is necessarily considered to expand the study about which data should be taken into account in order to reach a suitable and trustworthy set of real variables.





A semaphore can improve its service level using proper data collections obtained from real junctions. Therefore, it is very transcendental to go deeper into studies that facilitate the consideration and inclusion of more real aspect which can affect the decision process. For instance, humans sometimes do things arbitrarily and against all odds but in a controlled environment, the variable can be manipulated to avoid randomness. Indeed, people are different and react modifying their behavior pattern, considering variables that have not impact on the situation. Drivers are the most important source of data. At the moment, the literature has not reported a system robust enough to satisfy the road constraints of the cities. Intelligent transport systems must be closely designed according to the human beings. In this light, it is necessary to develop inference models based on human who must consider emotions, beliefs, politics, manners, etc.

References

- Ibarra, S., Castán, J.A. and Laria, J. (2014) Optimizing Urban Traffic Control Using a Rational Agent. *Journal of Zhejiang University Science C*, 15, 1123-1137. https://doi.org/10.1631/jzus.C1400037
- [2] Chen, Y. and Lv, Y. (2016) Analysis and Forecasting of Urban Traffic Conditions Based on Categorical Data. *IEEE International Conference on Service Operation* and Logistics and Informatics, Beijing, 10-12 July 2016, 113-118. https://doi.org/10.1109/SOLI.2016.7551671
- [3] Jiawei, H., Kamber, M. and Pei, J. (2006) Data Mining: Concepts and Techniques. Morgan Kaufmann, Burlington, 284-304.
- [4] Qiu, C., Jiang, L. and Li, C. (2017) Randomly Selected Decision Tree for Test-Cost Sensitive Learning. *Applied Soft Computing*, 53, 27-33. <u>https://doi.org/10.1016/j.asoc.2016.12.047</u>
- [5] Rizzo, G., d'Amato, C., Fanizzi, N. and Esposito F. (2017) Tree-Based Models for Inductive Classification on the Web of Data. Web Semantics: Science, Services and Agents on the World Wide Web, 45, 1-22. https://doi.org/10.1016/j.websem.2017.05.001
- [6] Sun, H. and Hu, X. (2017) Attribute Selection for Decision Tree Learning with Class Constraints. *Chemometrics and Intelligent Laboratory Systems*, 163, 16-23. <u>https://doi.org/10.1016/j.chemolab.2017.02.004</u>