Predicting Short Term Traffic Congestion on Urban Motorway Networks

submitted by Taiwo Olubunmi Adetiloye

Report by Nicolas Saunier, Polytechnique Montreal

May 10th 2017

The title of this thesis is "Predicting Short Term Traffic Congestion on Urban Motorway Networks". The general idea of the topic is timely and important, but the work presented does not address it adequately.

Contribution to knowledge (Originality, quality, quantity)

In section 1.3 (research gaps and thesis objectives), the candidate mentions "traffic congestion problems" without any definition. The sentence containing the expression is therefore hard to evaluate as true or not (what are also "classical data mining algorithms"?). What is also striking is the obsession with traffic congestion that seems to be used interchangeably with traffic condition. Congestion is just one of the conditions of traffic, and any work on studying congestion needs to distinguish it from other traffic conditions.

The candidate should provide a definition of hybrid algorithms, which could not be found in the whole document. It is unclear why and how undefined "hybrid and state of the art algorithms" can "improve the computation time", which seems to be the first research objective (the purpose of these algorithms at this point seems to be traffic congestion problems...).

The statement that "this [simulation] is generally a cost-effective and time-saving approach for analyzing and reproducing real-world scenarios while avoiding certain complexities emanating from social traits or the changing weather conditions" is contradictory, as "real-world scenarios" do have complexities emanating from behaviour and changing weather conditions. "Analytics" are undefined. The second objective seems to develop a microscopic traffic simulation, without any specific goal in terms of improvement over the state of the art. No contribution is claimed in this area.

The third objective seems to be the analysis of social media data (only Twitter is mentioned) for traffic condition analysis and its causes. It is more clear.

The last objective is about "data fusion techniques" for several traffic information sources.

Finally, the candidate states that "the objective of the thesis is to propose a multi-data multi-method framework for short-term traffic congestion on urban motorway networks", which obviously lacks a term, probably traffic congestion prediction as in the thesis title. This general objective should be first in section 1.3. This last paragraph is more clear that the rest that should be completely rewritten.

As is described in the next section, there is no contribution to knowledge in the presented document. If any of the work done by the candidate is an actual original contribution, the presentation makes it impossible to properly assess it.

Research and experimental work (adequacy, skill, thoroughness)

Chapter 2

Chapter 2 deals with some machine learning methods for traffic congestion prediction. The candidate cannot state that he "develop[s] various models based on these algorithms" since these algorithms already exist. As most literature reviews in each chapter, it is partial and poorly worded. The choice of specific machine learning techniques is not justified.

As will be shown in the experimental results, many of the tasks presented in the literature review are trivial, for example predicting the congestion level (3 categories) from speed in figure 2.3 (speed defines congestion!).

Sections 2.3, 2.4, etc. on the methods chosen by the candidate to be tested on the experimental data should be subsections of a section on "tested/selected methods". As is, it is easy to miss the end of the literature review. Also, instead of presenting the general problem and the input/output variables, they are repeated for each method without clearly stating they are the same. On pages 24-25, it is unclear why three traffic flow functions/variables are needed, and how they are computed based on (undefined) "traffic flow state, s". For some tested models, the back-propagation neural network (BP-NN) and the random forests (RF), the parameters are provided in the presentation of the model, without any justification, while they are not provided for the other models. How figure 2.8 was produced is not explained (what data was used, the model parameter and the size of the bootstrap sample).

The performance evaluation is poorly presented. In particular, the performance measures can be computed as presented only for a scalar numerical output, while the output is a categorical variable (three levels of congestion). The model sensitivity is completely confused. There is no explicit link between the Spearman rank correlation and the analysis of sensitivity analysis of the model to its input variable. The candidate fails to define for which variables the correlation is computed. The first definition of Spearman's rho is odd with undefined terms ("accurate predictions" for which task? and which "test sets"?), unrelated to the evaluation of the input variables. The second part of the equation is the standard definition of rho, but is unrelated to the first definition (and di is not properly defined).

The experimental results are presented in section 2.8. The test data is traffic data from the UK, published on an open data portal. The presentation of this data by the candidate is misleading from the beginning, as the candidate makes it sound like he obtained data from "multiple sources including Automatic Number Plate Recognition (ANPR) cameras, in-vehicle Global Positioning Systems (GPS)

and inductive loops built into the road surface and captured in real time" and combined these sources himself, while the data is already aggregated and provided as is on the UK data portal. To be clear, the candidate has processed no GPS data himself. The data presentation is further muddled in terms of dimensions (96 intervals per day, but there is "a total sample size of 1048576"), "split per the weekly traffic information" (what does that mean?), the input/output data ("dimension of 9", but the input is "length of traffic lane", average speed and traffic flow and the output is the average journey time)... How the various links are tested (separately or as a whole) is also not described (from p35, it seems only link AL100 is used...). A screenshot of a data mining software is not a proper description of the dataset (figure 2.10). It is unclear how the tests sets are generated and why four test sets are used. Something also must be wrong because the performance on the test sets is superior to the performance on the training set.

The most important issue however is that the task, as presented, is to predict travel time t based on link length l and average speed s: this is trivial since t = d/s! I actually downloaded the data used by the candidate and verified it experimentally. There is no point in using machine learning methods if there is an analytical solution to compute the output from the inputs. Even if the output is discretized, which is unclear, it can still be done exactly from mathematical formulas.

Figure 2.12 has poor quality and is not legible, ie 3D plots are difficult to read when projected on 2D space (a page). The "sensitivity analysis" is mostly presented in terms of the various model performance (it's not clear how the Spearman's rank correlation, between undefined variables, related to model performance). There is only one sentence stating "the two most important variables affecting the sensitivity of the model results are he traffic flow and average speed" without any evidence (these variables are not in figure 2.12 which is the only piece of evidence in this section). It seems also that there is something about time period in the prediction task, but this is the first time this is mentioned, at the end of the section on performance evaluation.

There is clearly no contribution in this chapter, and many errors, missing information.

Chapter 3

The second objective is presented in the third chapter, "Agent-based modeling of traffic congestion". The author choose implicitly from the beginning an agent-based model. The use of traffic simulation is rather correctly stated, except for the emphasis on visualization that is a by-product of a simulation. It feels as if the software used for traffic simulation was chosen first, then came the literature review (and the research problem).

The review of traffic models is limited given the literature in this very active area of transportation research (see for example the (already old) monograph from the Transportation Research Board committee on traffic flow theory). The general description is ok, but vague with obvious statements and some contradictions. For example, agent based model are by definition microscopic as they model each agent (vehicle/driver/user) individually ("multi-agent based simulation offers mainly microscopic modeling of parameters"). It also lacks a clear typology of traffic models, which is easy to find in textbooks and survey papers. The research presented in table 3.1 does not precisely describe the traffic models used in the cited research (and cites traffic simulation software instead of the models implemented in the software).

The proposed model seems confined to single lane traffic, which is in practice very limited. The model is heavily based on the Anylogic software, which makes it difficult to evaluate the candidate's work (Anylogic even includes traffic simulation modules). The elements of the simulation, for example how the space is represented, are not clearly or completely described: is the space continuous? How are roads and road users represented? Continuous 2D space? Individual user attributes? What distributions are used to generate the traffic flows? Where are the origin-destination matrix calculated or where do they come from? The traffic model may be "the car-following within safe distance pattern (Brackstone and McDonald, 1999)", but it is not described further. The "model" presented in figure 3.1 includes only 3 vehicles plus vehicle "x", without any general mechanism of how the vehicles move on specific road segments, from one to another, how vehicles interact with each other...

The results presented in section 3.4 lack too many details to be of any value: the network and demand are not defined, whether several simulations are run is not stated for figure 3.2, the generation model (Poisson?) is not defined, figures 3.3 and 3.4 lack legends and proper description of the phenomena, etc. The candidate fails to properly present and use the Greenshields speed-density relationship (when density reaches jam density, speed is 0, traffic is at a standstill). It is also unclear why one would use last-in first-out for traffic models, as queue priority is obviously first in, first out...

The results presented in section 3.5 are for a "road boundary", apparently made of three lanes ("lane1, lane2 and lane3"). Again, there are no further details about the cases being used for experimental results. The quality of the graphs is extremely poor. Case 1 is simulated over 60 seconds only (to represent the "onset of rush hour"!). The only presented result is the number of vehicles per lane. How this relates to the performance of the model is not explained. In case 2, how the number of vehicules per lane presented in figure 3.7 relates to "the effect of traffic signal controls" is anyone's guess. The candidate does on to present new equations (in the experimental results) purporting to explain how queueing is managed in the model. Equation 3.2 is impossible to understand, where g seems to be a function, but is defined as the "step-size on parameter y_k+1", h is undefined, f is incomprehensibly defined as "a function of the amount of traffic signaling with respect to the delay time", x is defined, but x_k is not... Some accuracy is reached, without stating to what variable it relates and how it is measured (with what real data?). I have not idea what conclusions can be drawn from table 3.3: something was simulated, with some numbers, that's it. The cases are not defined in sufficient details, and there appears to be no real data to compare to.

Furthermore, no comparison is made with other models to demonstrate how the stated objective to "improve the quality of monitoring traffic situations with a view to reducing the traffic jams in the real-world" is achieved.

Again, this chapter shows no contribution, and the model used is the least well described, with meaningless case studies.

Chapter 4

The candidate is very confused about what constitutes a contribution as the following sentence shows: "The main contribution of this chapter is to review the application of twitter big data using sentiment analysis and cluster classification for traffic flow prediction; to develop a robust model methodology and to perform the comparative evaluation of results" (p 58). Reviewing and applying existing methods is not a contribution.

The main "tweet mining approach" from (Azam et al. 2015) presented is missing many details to be properly understood. There seems to be a confusion between an actual graph (the mathematical object, with the theory that goes with it) and the social network ("social network generation") or "social graph". Several notations are undefined, in particular ti, so the equations cannot be understood.

The candidate confuses the method (section 4.3) with implementation details (use of API and software libraries). The "phrase search algorithm" is based on previous work and is poorly described (p62-63). It is unclear what its inputs are, nor what the k-nearest neighbour, a classifier, is used for. The loop at line 6 to 10 will run only once since it contains a return statement with no condition. The purpose of the "Forward-Position-Intersect algorithm" is not described and it is difficult to tell from the algorithm.

The goal from equation 5.6 seems to classify tweets in "positive, negative or neural" category, but it is unclear what this means and how this could be useful for traffic condition estimation/prediction. There is no mention of localization of the tweets, i.e. one does not know what roads or parts of the network the conditions are related to. How the tfidf is used (p66-67) in this classification is unclear.

In the experimental evaluation, only 121 tweets are used, which is clearly not "big data" as presented in chapter 1 and the chapter introduction (4.1). The classification accuracy can be computed as presented only for binary classification problem, while the one presented has three classes. There are errors in the equations of the classification performance measures. It is unclear why there is no mention of recall along with precision, while FP rate is undefined. Talking of "correctly classifying the traffic congestion" is not an an accurate description of the research as the true traffic condition is not known. It is not even clear where the true labels for the tweets comes from, to be able to compute these performance measures. In the "model validation" section (4.4.3), it is unclear how Latent Dirichlet Allocation (LDA), a clustering method, can be used for classification. Furthermore, figure 4.2 is impossible to interpret with its unlabelled axes. To support the usefulness of the method, the candidate must demonstrate how these "interesting results [...] could help inform decision-making on traffic congestion, incidents and control".

Again, the lack of clear explanation of the model and experimental validation prevents from identifying any contribution in the work presented in this chapter.

Chapter 5

This chapter again starts with a long literature review (about 10 pages) split into sections 5.1 and 5.2, the latter being the one actually titled "related work"! Important sections on the different data fusion architectures are paraphrased from (Castanedo 2013) (https://www.hindawi.com/journals/tswj/2013/704504/) that is cited in this part of the text. This is borderline plagiarism, for example:

- thesis: "The centralized scheme is theoretically optimal if we assume that the central processor’s tasks such as association, filtering, and tracking are performed correctly and data transfer time is insignificant. "

- (Castanedo 2013): " If we assume that data alignment and data association are performed correctly and that the required time to transfer the data is not significant, then the centralized scheme is

theoretically optimal."

- thesis: "The architecture consists of a network of nodes in which is node has its own central processor. Hence, there is no single point of data fusion such that each node fuse its local information with other information it received from its peers. "

- (Castanedo 2013): "a decentralized architecture is composed of a network of nodes in which each node has its own processing capabilities and there is no single point of data fusion. Therefore, each node fuses its local information with the information that is received from its peers. "

- thesis: "In terms of the disadvantages, high communication cost which is $O(n2)$ at each communication steps is expected, where n is the number of nodes. Also, scalability becomes an issue when the number of
nodes is increased where each nodes communicates with all of its peers."

- (Castanedo 2013):"The main disadvantage of this architecture is the communication cost, which is at each communication step, where is the number of nodes; additionally, the extreme case is considered, in which each node communicates with all of its peers. Thus, this type of architecture could suffer from scalability problems when the number of nodes is increased;"

There are at least two types of work presented in table 5.1, the first about traffic condition estimation/prediction and the second about route planning using traffic information, which should be better presented and organized.

The candidate considers the Kalman filter KF and "Mamdani fuzzy-rule based inference (MFRI)" without any reference for this method. The reason for using the extended KF (EKF) are unclear on p85, while it is easy to find that EKF and other extensions are used when the assumptions for the KF are not met, e.g. non-linear state transitions (which the candidate notes on p89).

The proposed method is missing many details to be understood. There seems to be a "phase 1" dealing with "data obtained from the GPS, and sensor". There is however no information about which "sensor" data is used, nor GPS data (what kind of vehicle? what frequency?) and what is the purpose (output) or their classification "using the ML classification algorithms" (which one among "NN, RF, NF and DL"?). This part is not further presented. It is unclear what alignment is done at several steps.

In "phase 2", the output of GPS and sensor classification is fused with a "computer based model" without further information either using the EKF. The presentation of the KF seems broadly based on the webpage http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/WELCH/kalman.1.html (same notations), but the candidate cannot even present Kalman filters without adding errors of his own, for example on the dimensions of vectors and matrices. Maybe more importantly, the KF and EKF are standard and well-known techniques and their application does not constitute sufficient work for a PhD.

In fact, even if KF and EKF are standard techniques, the candidate cannot even apply them correctly. The application is done in section 5.3.2. Important information about the system are not provided: the dimension of the state vector, the functions f and h. Instead of having xk the system state at time k, the candidate describes only x1, x2 and x3, defined respectively as the amount of delay time, the number of

vehicle agents and the average traffic flow. Given this way, the system is defined only at steps 1, 2 and 3, with different variables. The story is similar with the control inputs u2 and u3 (why no u1?). This is a farce.

It actually becomes even funnier (or more sad?) when the candidate adds to the EKF equations the "twitter normalized outputs", which have never been clearly defined, but seem to be categorical, while the variables x1 to x3 seem to be numerical. The MFRI system consists in 6 rules and it is unclear how they were devised.

Given all these shortcomings, it is impossible to analyze the experimental results, also given the lack of information on the data used, in particular the tweets and their relation to the traffic data from the UK. The z-dimension in figure 5.7 should be the system output, or the error on the output, but is labeled "Correct"... It is unclear how the whole method can be faster than one of its component (the ML classification in phase 1) (" the average run time was lesser with a value of 0.58s, considered better than the best average time obtained by RF").

It is unclear what "the attribute and relational inferencing recommended by Steinberg and Washburn(1995)" is and how it relates to the traffic data fusion. It is described as a "data fusion performance estimation" method, but cannot be understood from the description (what "node X"?). The number of observations N seems like an important parameter for performance, but has not been defined in the method (in chapter 5 at least). It is unclear which method is the best (whether the performance metric should be high or low).

To sum up, none of the methods presented and (maybe) used by the candidate are new. It is also clear from the presentation of the KF that he does not understand and cannot apply some of them.

Also, the data used to validate experimentally the proposed methods is poorly described and the main task, predicting traffic condition as travel time from other variables such as average speed, is trivial.

Overall, the candidate has shown no evidence of an original research contribution, and clearly demonstrated a lack of understanding of the scientific method, in particular the clear presentation of one’s research so that the reader can understand and reproduce it. The presentation of the work is incomplete, filled with inconsistencies and errors, and the experimental results cannot be interpreted in many cases, and, when they can be, are meaningless because the task is trivial. The lack of rigour is breathtaking.

Understanding of the subject (review of previous work, evaluation, validity of conclusions)

From the previous observations, it should be clear that the review of previous work is at best vague and muddled, sometimes clearly derived from the cited sources, with added inconsistencies and errors. The candidate lacks an overall perspective of the previous work in the relevant areas. The candidate has clearly not spent enough time doing the literature review.

Most conclusions are not supported by the evidence provided in the document. The final chapter is very short and lacks a clear statement of the original research contributions, as well as a discussion of the

limitations.

Presentation (organization, tables, figures, style)

The document is poorly presented. The general document organization is ok, but the writing is poor, unclear, with long sentences, some missing a verb. Key terms are often not defined, or vaguely defined, then a similar term is also defined, without clearly relating it to the first. Tables are ok, but figures are often of very poor quality (pixelated instead of vector format) and difficult to interpret because of a poor choice of the type and characteristics of the visualization, for example the 3D figure 2.12.

Quality of Typescript

As stated above, the overall quality of the manuscript is poor, barely readable at times, with very poor quality of the figures. Detailed comments can be found in the annotated document (https://www.dropbox.com/s/xjqgss6f29ebt97/Adetiloye_PhD_F2017.pdf?dl=0).

Conclusion

It should be clear from the previous observations that this work is terrible. It can only be added that it is impossible to understand how anyone could think the quality of this work might be sufficient for a PhD.