Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 14 (2016) 2035 – 2043

Transportation Research Procedia

www.elsevier.com/locate/procedia

6th Transport Research Arena April 18-21, 2016

UNIETD – assessment of third party data as information source for drivers and road operators

Daniel Elias [a,*], Friedrich Nadler [a], Ian Cornwell [b], Susan Grant-Muller [c], Thomas Heinrich [d]

[a]nast consulting ZT GmbH, Lindengasse 38, Vienna 1140, Austria
[b]Mott MacDonald, 1 Atlantic Quay, Broomielaw, Glasgow, United Kingdom
[c]University of Leeds, Faculty of Environment, 1-3 Lifton Place, University of Leeds, Leeds LS2 9JT, United Kingdom
[d]TRANSVER, Maximilianstr. 45, Munich 80538, Germany

Abstract

The paper deals with the assessment of third party data such as crowd sourced/social media and floating vehicle data as information source for road operators in addition to traditional infrastructure-based techniques. For purposes of quality assessment of different types of data and available ground truths existing test/evaluation methodologies have been assessed. A new methodology has been designed for assessment of speeds and travel times using normalized (between 0 and 1) quality indicators that can distinguish between “detection rate” and “false alarm rate” concepts. In terms of harvesting social media the relevance of social media content has been assessed against a range of traffic management requirements. Furthermore the level of content that will be available has been estimated as well as commercial sources and business models for road authorities. Analyses cover unstructured data from Twitter and Facebook both historical data and three months of contemporary data. In addition surveys are conducted in England and Austria to retrieve information from the public in terms of which social media platforms are commonly used to share information about traffic related incidents.

* Corresponding author. Tel.: +43 1 523 47 33 - 28; fax: +43 1 523 47 33 - 31.
E-mail address: elias@nast.at

doi:10.1016/j.trpro.2016.05.171

Keywords: social media; traffic information; data fusion; traffic prediction

1. Introduction

The research project UNIETD (Understanding New and Improving Existing Traffic Data) is currently being carried out as part of the CEDR Transnational Road research Programme Call 2013. The funding for the research is provided by the national road administrations of Belgium-Flanders, Denmark, Finland, Norway, UK, and the Netherlands.

It deals with the assessment of third party data such as crowd sourced/social media and floating vehicle data as alternative information source in addition to traditional infrastructure-based traffic data sources since the quantity and potential of these alternative data sources have increased in the past years. Road administrations are now facing the challenge whether to keep and maintain their own traffic data collection infrastructure or to partially source out the acquisition of traffic data information. Up to now there are no standard methodologies in place for road administrations to verify the quality of external traffic data. Therefore UNIETD's objective is the provision of guidance and validation tools for road administrations in terms of third party data usage.

Preliminary study results of existing data sources showed that information on the quality of the traffic data sources is very limited. So far it seemed unlikely that social media data alone would satisfy the requirements of traffic management although it has great potential as additional data source. Given the fact that different alternative data sources can provide various kinds of data with different characteristics, traffic data fusion appears to be increasingly important for traffic management purposes.

This paper deals with the following preliminary research results:

- development of a software toolkit for quality assessment of traffic data and services
- analysis of social media data as additional data source for traffic management
- potential of crowd sourced traffic information gathering
- preview of the results of the Austrian survey in terms of traffic related information on social media platforms

2. Existing methods for quality assessment of traffic data

This chapter presents existing methods for quality assessment of traffic data from different sources, gives indication about difficulties in terms of interpretation of analysis results and guides towards the new evaluation method that is being developed during UNIETD.

2.1. Existing evaluation methods

The majority of traffic data assessment methods depend on a specific "ground truth" which can be seen as data source whose quality can be treated as highly accurate in order to be used as reference for the assessment of secondary data sources such as data from mobile devices or navigation units.

In the past years different data validation methods have been developed based on the individual characteristics of the corresponding input data that is being assessed. Data sources which are commonly being analysed could therefore be raw data, derived speeds, travel times or level of service values in different levels. Table 1 presents existing data assessment methods for different kinds of reported data and available ground truth data.

Table 1. Existing traffic data analysis methodologies for different kinds of reported data and available ground truth data.

		Reported data			
		FCD/GPS	**LOS**	**Speed**	**Travel time**
Ground truth data	**FCD/GPS**	Inherent checks	QFCD	Q-bench	Q-bench
	Stationary detector data	-	QKZ	SIMPE	TTD Index
	Subjective data	-	Subjective Method	-	-

2.2. Evaluation of reported Level of Service

In comparison to the remaining data input formats according to Table 1 the Level of Service (LOS) is based on a binary classification that either indicates a congestion incident or not. During the process of comparing the LOS with the corresponding available ground truth four possibilities do exist:

- true positive – there was a real event and it was reported
- false negative – there was a real event but it was not reported
- true negative – there was no real event and none was reported
- false positive – there was no real event but one was reported

When assessed over a set of time periods these can be used to derive:

- "detection rate" – ratio of event reports to real event states
- "false alarm rate" – ratio of reported events to event-free states

These indicators can easy be understood and can be visualised on a 2D graph. Methods according to Bogenberger (2003) – QKZ – and Bogenberger et al (2009) – QFCD – provide specific formulae to derive variants of these indicators.

Apart from the objective data assessment process it is also possible to compare the reported LOS on a subjective assessment level with survey personnel. This method certainly bears inherent limitations due to the difficulty to assign reported LOS to the actual LOS as perceived by each individual assessor. Furthermore inherent difficulties occur if more providers are being compared at the same time since the number of defined LOS may vary in between the service providers. To ensure a meaningful comparison, explicit measures have to be taken into account to compensate these differences.

2.3. Evaluation of reported speed and travel-time

In contrast to the methods described above Lux (2011) – Q-Bench, Huber et al (2014) – SIMPE – and Huber et al. (2013) – TTD-Index – derive numeric indicators in order to assess the quality of reported speeds or travel times. At the end of the individual analysis process a single summary indicator is presented that give indication in terms of the overall quality. Q-Bench measures the benefit to the service user. The overall Q-Bench metric is the ratio of benefit from the service to benefit from an ideal service which could be seen as perfect knowledge according to the ground truth incorporating all real delays. Although it can be seen as very useful metric it does not distinguish between understating congestion and overstating congestion.

3. The UNIETD approach

Based on the analysis of the existing methodologies discussed in chapter 2 the UNIETD project aimed to develop a new methodology that can be used for the evaluation of reported travel times by means of avoiding problems with different and/or unknown definitions of Level of Service; that delivers normalized (in between 0 and 1) quality indicators that are easy to understand and interpret which are able to distinguish between "detection rate" and "false alarm rate" concepts; that has only a small number of easily understandable parameters.

Within the UNIETD approach, the possible combinations of ground truth travel time tt_{gt} and reported travel time tt_{rep} are distinguished into cases using congestion thresholds derived from free flow speeds or travel times (1):

$$tt_{cong} = \frac{l}{v_{cong}} = \frac{l}{r_{cong} \times v_{ff}} = \frac{1}{r_{cong}} \times tt_{ff} \tag{1}$$
$$r_{cong} = \frac{v_{cong}}{v_{ff}}$$
$$r_{cong} \in \,]0;1]$$

and using travel time tolerance Δtt_{tol} calculated as a share of the ground truth travel time (2):

$$\Delta tt_{tol} = r_{tol} \times tt_{gt} \quad (2)$$
$$r_{tol} = \frac{\Delta tt_{tol}}{tt_{gt}}$$
$$r_{tol} \in [0; 1]$$

The different cases are:

- “Not relevant” for the quality assessment: The ground truth travel time and the reported travel time are both lower than or equal to the congestion threshold travel time (3):

$$tt_{gt} \leq tt_{cong} \cap tt_{rep} \leq tt_{cong} \quad (3)$$

- “Relevant” for the quality assessment: The ground truth travel time or the reported travel time are higher than the congestion threshold travel time (4):

$$tt_{gt} > tt_{cong} \cup tt_{rep} > tt_{cong} \quad (4)$$

The “relevant” case is further divided into:

- “Correctly reported” congestion (crc): The reported travel time is within a travel time tolerance Δtt_{tol} below or above the ground truth travel time (5):

$$tt_{gt} - \Delta tt_{tol} \leq tt_{rep} \leq tt_{gt} + \Delta tt_{tol} \quad (5)$$

  - “Not reported” (nrc): The reported travel time is lower than the ground truth travel time minus the travel time tolerance (6):

$$tt_{gt} - \Delta tt_{tol} > tt_{rep} \quad (6)$$

  - False reported” congestion (frc): The reported travel time is higher than the ground truth travel time plus the travel time tolerance (7):

$$tt_{rep} > tt_{gt} + \Delta tt_{tol} \quad (7)$$

Based on the cases defined above, we define two normalized (to the range 0 to 1) quality indicators similar to detection rate and false alarm rate (as used in LOS methodologies) but based on travel times (not on LOS):

- The not-reported rate nrr is the ratio of the sum (over all segments and time intervals) of the not-reported delays in the not-reported case to sum of the relevant real delays (8):

$$nrr = \frac{sum\ of\ not\ reported\ delays\ at\ not\ reported\ case}{sum\ of\ relevant\ real\ delays} \quad (8)$$
$$nrr = \frac{\sum tt_{gt,nrc} - tt_{rep,nrc}}{\sum tt_{gt,gt>cong} - tt_{ff}}$$

This not reported rate shows the proportion of the relevant real delays that have not been reported by the traffic information service. To obtain a quality indicator similar to true positive rate, a reported rate rr can be calculated as 1 – nrr.

- The false reported rate frr is the ratio of the sum of the false reported delays in the false reported case to sum of the reported delays (9):

$$frr = \frac{sum\ of\ false\ reported\ delays\ at\ false\ reported\ case}{sum\ of\ relevant\ reported\ delays\ at\ all\ cases} \tag{9}$$
$$frr = \frac{\sum tt_{rep,frc} - tt_{gt,frc}}{\sum tt_{rep,rep>cong} - tt_{ff}}$$

This false detection rate shows the proportion of the relevant delays reported by the traffic information service that have actually been false (i.e. not real).

The not-reported rate nrr in combination with the false reported rate frr can be visualised as a single point on a 2D graph. Fig. 1 shows an example of such an assessment which also includes graph background colors. These colours indicate the quality level, with green being best and red being poorest.

The developers of the new methodology consider the UNIETD approach useful to provide information that is easier to understand and interpret than metrics that have previously been discussed. Calibration parameters consist of the so called congestion rate and tolerance rate, which can be changed in order to derive congestion thresholds and travel time tolerances for each segment.

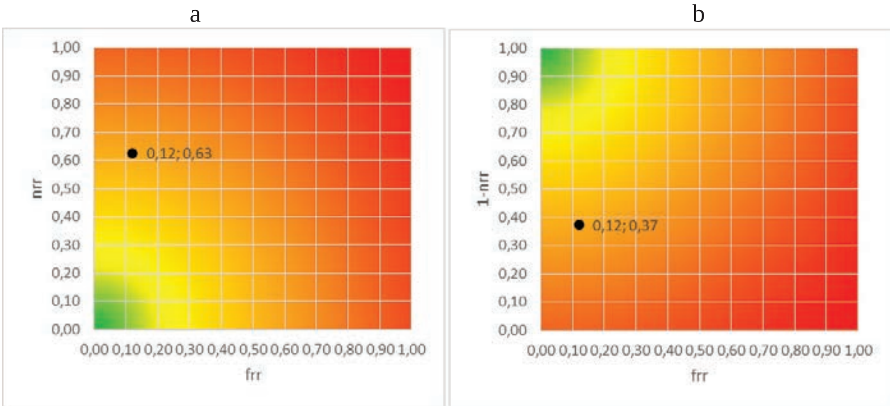

Fig. 1. Quality diagrams applying UNIETD method to a real service – (a) example for not-reported rate nrr; (b) false reported rate frr.

In addition to the development of the algorithms a corresponding software toolkit has been created that implements the new method and several other quality evaluation methodologies. First assessments that have already taken place reviewed data from Flanders.

For a tender of FCD-based traffic information and more specifically only travel times, the Flanders Transportation Center ran a quality evaluation of the external provided data, using the Q-bench method. The candidate service provider delivered the “reported” travel times for two road sections. As ground truth data, bt-equipment was installed at the start and end of each road sections.

The data in respect was provided in order to the newly developed UNIETD approach. As the data respect to ground truth and reported travel times (already map-matched), several of the methodologies described above could already been carried out.

4. Social media as additional source of traffic information

In addition to the software toolkit for quality assessment which has been presented in the previous chapter another UNIETD objective deals with the analysis of social media data as additional data source for traffic management purposes. In terms of this objective social media with geographical information was analysed in order to find out if it can be used to estimate the speed of traffic; furthermore textual content of social media messages and posts was evaluated to obtain exemplary information on traffic flows, delays, infrastructure, and environment--related traffic issues.

4.1. Speed and direction estimation from Twitter tweets

The University of Leeds collected more than 2.8 million geotagged Twitter tweets sent by over 60,000 users for the time period between June 2011 and April 2013. Area of interest was the city of Leeds in the UK. The selection was restricted to geotagged messages in order to estimate traffic speeds and compare them with the corresponding section of the A1(M) road between junctions 42 and 46. Geometric pre-processing selected 14,000 tweets within a 500 m corridor around the A1(M) motorway including a number of tweets sent from residential areas close to the A1(M) road.

In general the analysis results showed an increasing trend in the volume of tweets through the period. The volume of tweets sent was, on average, higher towards the PM peak and variable across the day with a minimum of tweets in early morning hours. A similar number of tweets were sent from Monday to Wednesday, with an increase on Thursday, Friday, and over the weekend.

The geotagged tweets were used to calculate speeds from pairs of tweets, sent by the same user, less than 5 minutes apart. In contrast to usual map-matching algorithms straight-line distances were used to calculate approximate speeds since the A1(M) is relatively straight in the study area bearing in mind the inherent underestimation of travelling speeds. The direction in which the Twitter user was travelling was also calculated in order to assign the speeds to either the north- or south-bound direction.

Once the analysis of the pairwise tweets was conducted a total of approximately 400 users were found that showed realistic travelling speeds. These speeds, ranging from 10 km/h to 200 km/h were then compared to the average speed data derived from the corresponding inductive loop data of the same road section, direction and time time stamp of the user. The mean speeds for several sections of the A1(M) are summarised in Table 2 together with correlation coefficients.

Table 2. Mean speed comparison from data derived from tweets and inductive loops along the A1(M).

A1(M) section	Length	Mean loop speed	Mean Twitter speed	No of Data	Corr coef
J42 to J43 north	14.6 km	103.4 km/hr	91.4 km/hr	35	0.618
J43 to J44 north	10.5 km	101.0 km/hr	103.6 km/hr	75	0.439
J44 to J45 north	11.8 km	109.5 km/hr	86.5 km/hr	138	0.230
J45 to J46 north	4.7 km	104.7 km/hr	75.1 km/hr	24	0.200

A reasonable match between the mean speeds obtained from processed tweets and traffic detection loops could be identified. The correlation coefficients between individual speeds calculated from the Twitter data and the corresponding mean loop speed was still relatively poor, but this might be expected as the speeds obtained from the inductive loops are averages for vehicles passing the detectors over a 15 minute period, while speeds obtained from the processed tweets came from individual users/vehicles. An additional reason for the poor correlation between the two sets of speeds lies in the fact that the locations derived from geotags are not always accurate, leading to high speeds calculated for several users. If a mobile device has GPS enabled, the locations, and therefore speeds, should be reasonably accurate.

4.2. Analysis of textual content

The relevance of tweets as well as a corresponding classification was derived by means of designing ontology for textual content relevant to traffic management. Previous studies on ontology such as Gal-Tzur et al (2014) and ISO TR25100:2012 (2012) were considered but rejected since they had been derived for different purposes rather than road traffic information collection.

Tweets from transport authorities were removed since they were dealing with already known matters that would not be needed to be identified in terms of the proposed objectives. The remaining tweets were filtered to produce a set mentioning relevant road numbers in the study area. By means of manual search relevant messages containing traffic related information were identified. As an example '*Love the M62 me #stationary*' would be considered relevant, if not particularly useful, and a tweet such as '*Either I've taken a wrong turning on the way home or there's a New York Taxi on the M62 ?!!?*' would be considered irrelevant.

The open source program Hall (2009) – Weka – was used for text mining, and a range of classification algorithms provided by the program were chosen for this analysis. One set of manually classified tweets was used to train the classifiers, further sets were used for verification and testing. Of the 10 different learning algorithms used, 7 correctly classified more than 75% of tweets, and the best performing pair (SGD and SMO algorithms) correctly classified over 80% of the tweets. In further test on a second dataset of 650,000 more recent tweets from the same geographic area, results improved slightly with 6 out of 10 algorithms correctly classifying over 80% and the best performing pair this time correctly classifying 85% and 84% of the tweets. Weightings were then manually assigned to each word of each tweet in the training set according to relevance to road travel, and the experiments repeated, but this did not produce any significant improvement in results.

In terms of classifying tweets several classification layers were introduced according to the upper layers of the designed ontology including the following:

- Infrastructure
- Environment and meteorology
- Traffic conditions
- Driver or user experience

The analysis conducted used only the layer mentioned above due to the relatively low quantity of data available. One set of tweets was used for training automatic classifiers while further separate sets were used for verification and testing. Correct classification was lower than for the simple relevant/non-relevant case: 7 out of 10 classifiers achieved more than 50% correct classification and the best classifier achieved 67% correct.

The training tweets were then manually assigned a weighting representing perceived relevance to highways traffic management, by a transport domain expert, and this information was provided to the classifiers along with the tweets. This generally improved classification results by a few percent, although not for all classifiers. The best performing algorithm (Decision Table) correctly classified 72% of the tweets.

The results of the textual analysis are shown in Table 3. From the authors point of view the results suggest that classification of tweets could be useful for example to improve a service offering filtered information, for on-top manual inspection by traffic management operators or data fusion methodologies.

Table 3. Tweet classification success rates from best classifiers.

Experiment	Best classifier	Mean of top 3 classifiers
Classifying relevant/non-relevant	85%	84%
Classifying category	72%	67%

An increased number of relevant tweets could improve the classification success rates as well as further tuning of parameters. However, care would be required to avoid over-fitting to training data.

5. Potential of crowd sourced traffic information gathering using the example of Waze

There is an increasing number of initiatives for crowd-sourcing traffic data gathering based on services that deliver real-time navigation system using different kinds of background maps. Using the example of Waze which provides an international community based traffic and navigation application users can automatically provide location and speed, and can also manually provide further traffic-related information. From these data Waze can provide traffic information to its users.

The content available in typical data feeds show that these growing platforms could be used to support a number of use cases that are relevant to traffic management such as but not limited to:

- Detecting stationary vehicles
- Detecting congestion
- Detecting accident-prone locations
- Analysing network performance
- Analysing user experience
- Analysing the effect of police speed checkpoints
- Weather information
- Improving road works information

Future works should look into a more detailed coverage analysis of specified target areas in different European countries to determine whether the coverage supports the desired use cases.

6. Austrian surveys focusing on traffic related information on social media platforms

At the time of the paper submission surveys were being conducted in England and Austria regarding traffic related posting behaviours of social media platform users. Apart from general participation at several different social media platforms their pro-active and indirect participation about traffic related information was inquired. This chapter consequently focuses on preliminary results of the ongoing survey in Austria.

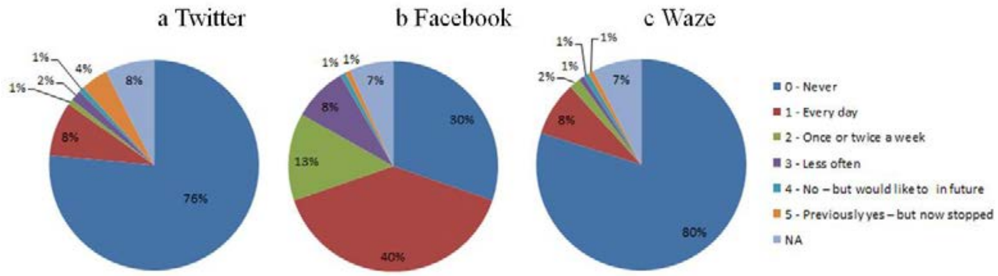

Fig. 2. Preliminary results (N=119) of the Austrian survey about the usage of different social media platforms (a) Twitter usage; (b) Facebook usage; (c) Waze usage.

Bearing in mind the preliminary status of the results of the ongoing survey in Austria it could be found out that according to Fig. 2 the majority of the inquired users prefer Facebook for their (daily) social media activities while Twitter and Waze are only used by approximately 8% of the users (on a daily base). This confirms the initial assessment in chapter 5 and highlights the big differences in terms of social media user behaviour across different European countries.

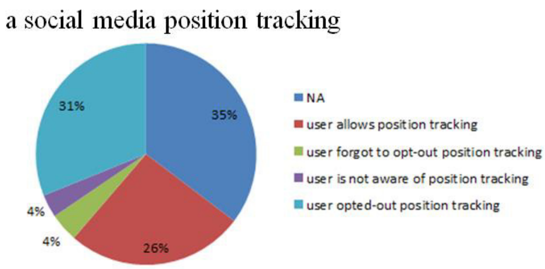

Fig. 3. Preliminary results (N=119) of the Austrian survey about position tracking settingsattempted by social media platforms.

Fig. 3 shows the preliminary results of the position tracking settings of individual users regarding social media platforms. These results were chosen as one of the most interesting results amongst others for presentation reasons because they indicate the potential limitation of data acquisition from social media sources. Apart from the high percentage of NA answers (Not Applicable answers correspond to participants that selected the option that no social media platform was used by them) the high amount of users that selected "opting out position tracking" is a very restraining fact that needs to addressed in order to increase the potential of these data sources for traffic information gathering.

7. Summary and outlook

This paper has described research results from the project UNIETD that currently deals with the development of a software toolkit for quality assessment of traffic data and services as well as the analysis of different kinds of social media data as additional data source for traffic management.

The newly developed UNIETD approach for quality assessment of reported traffic speeds and travel times delivers straightforward quality indicators that can distinguish between real delays and the tendency to falsely report delays. The software is easy to use since there are only two parameters (congestion rate and tolerance rate) that can be modified.

Analysis of the textual content of Twitter data has shown that tweets relevant to traffic management can be classified with reasonable success rates. However the results presented are indicative only as they related to a single geographic area and the relatively low proportions of tweets that are freely available. Further work in UNIETD will consider the implications of sourcing a much larger and potentially more viable data set.

Further works within the UNIETD project will address the following topics:

- consideration of how established traffic data fusion approaches such as those used by Highways England are impacted by the data quality results produced by the project;
- real-time traffic predictions, to enable selection of road maintenance working windows and other traffic management decision support. The outputs of van Vuren et al (2013) will be updated by consideration of the implications of the availability of new or improved data sources for the current state-of-the-art prediction methods and
- comparison of the survey results from England and Austria to determine which social media platforms are commonly used to share information about traffic related incidents; consequences and strategies derived from the results of the surveys.

Acknowledgement

The research presented in this paper was carried out as part of the CEDR Transnational Road research Programme Call 2013. The funding for the research was provided by the national road administrations of Belgium--Flanders, Denmark, Finland, Norway, UK, and the Netherlands.

References

Bogenberger, K., 2003. Qualität von Verkehrsinformationen. Straßenverkehrstechnik, Volume 47, Issue 10.

Bogenberger, K., M. Hauschild 2009. QFCD – A microscopic model for measuring the individual quality of traffic information. ITS World Congress 2009. Stockholm, Sweden.

Gal-Tzur, A., Grant-Muller, S.M., Kuflik, T., Minkov, E., Nocera, S., Shoor, I., 2014. The potential of social media in delivering transport policy goals. Transport Policy, 32, 115–123.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: An update. SIGKDD Explorations, 11(1).

Huber, G., Bogenberger, K., Bertini, R., 2014. New Methods for Quality Assessment of Real-Time Traffic Information, Paper Number 14-2918. 93rd Annual Meeting of Transportation Research Board.

Huber, G., Bogenberger, K., 2013. A Quality Evaluation Model for Real-Time-Traffic-Information. ITSC 2013, (S. 2126-2131).

ISO TR25100:2012 Intelligent transport systems – Systems architecture – Harmonization of ITS data concepts.

Lux, C., 2011. QBench – Evaluation of Traffic Flow Quality. Proceedings of BASt colloquium "Quality of on-trip road traffic information".

van Vuren, T., Cornwell, I., Finer, M., van Hinsbergen, C., Zuurbier, F., 2013. Using Short Term Traffic Predictors in Traffic Management Centres, 9th ITS European Congress, Dublin.