Evaluating the reliability, coverage, and added value of crowdsourced traffic incident reports from Waze

Corresponding Author:
Mostafa Amin-Naseri
Iowa State University
2711 S. Loop Drive, Suite 4700, Ames, IA 50010-8664
Phone: (720) 334-9798
Fax: (515) 294-3524
aminnas@iastate.edu

Pranamesh Chakraborty
Iowa State University
2711 S. Loop Drive, Suite 4700, Ames, IA 50010-8664
Phone: (515) 357-9076
Fax: (515) 294-0467
pranames@iastate.edu

Anuj Sharma
Iowa State University
351 Town Engineering Building
Iowa State University, Ames, IA, 50011
Phone: (515) 294-3624
Fax: (515) 294-0467
anujs@iastate.edu

Stephen B. Gilbert
Iowa State University
3026, Black Engineering Building
Iowa State University, Ames, IA, 50011
Phone: (515) 294-6782
Fax: (515) 294-3524
gilbert@iastate.edu

Mingyi Hong
University of Minnesota
6-109 Keller Hall,
University of Minnesota, Minneapolis, MN 55455
Phone: (612) 625-3505
Fax: (515) 294-3524
mhong@umn.edu

ABSTRACT

Traffic managers strive to have the most accurate information on road conditions, normally by using sensors and cameras, to act effectively in response to incidents. The prevalence of crowdsourced traffic information that has become available to traffic managers brings hope and yet raises important questions about the proper strategy for allocating resources to monitoring methods. Although many researches have indicated the potential value in crowdsourced data, it is crucial to quantitatively explore its validity and coverage as a new source of data. This research studied crowdsourced data from a smartphone navigation application called Waze to identify the characteristics of this social sensor and provide a comparison with some of the common sources of data in traffic management. Moreover, this work quantifies the potential additional coverage that Waze can provide to existing sources of the Advanced Traffic Management System (ATMS). One year of Waze data was compared with the recorded incidents in the Iowa’s ATMS in the same timeframe. Overall, the findings indicated that the crowdsourced data stream from Waze is an invaluable source of information for traffic monitoring with broad coverage (covering 43.2% of ATMS crash and congestion reports), timely reporting (on average 9.8 minutes earlier than a probe-based alternative), and reasonable geographic accuracy. Waze reports currently make significant contributions to incident detection and were found to have potential for further complementing the ATMS coverage of traffic conditions. In addition to these findings, the crowdsourced data evaluation procedure in this work provides researchers with a flexible framework for data evaluation.

INTRODUCTION

Traffic managers aim for increased mobility and safety on the roads. Real-time information on road conditions is necessary for taking proper actions. However, relying on the sensors and cameras for monitoring traffic conditions at all locations and times is neither possible nor economically justifiable *(1)*. Moreover, many sensors detect incidents based on speed changes, while in less populated areas, a crash may present a high-risk zone for secondary crashes without an immediate significant speed drop. These circumstances point to the insufficiency of the existing means for full road condition monitoring.

Recent research has demonstrated the potential value in leveraging social media to detect traffic incidents *(2)–(5)*. Thus, crowdsourced data, have recently gained attention in traffic management. To this end, many cities and departments of transportation (DOTs) have incorporated data from a crowdsourced smartphone application called Waze into their ATMS. Using crowdsourced data, however, poses several questions to the traffic managers. In this research, a quantitative analysis is implemented to provide data-driven answers to some of the common concerns of traffic managers with regards to Waze data.

Iowa Department of Transportation (IDOT) has used Waze data as a source of incident detection since September 2015. One year of data (2016) was used to address questions in three primary areas.

a. How does Waze compare to existing sources?
    - Are Waze reports reliable?
    - What percentage of the current recorded incidents were detected by Waze?
    - How does Waze compare to other common sources of data collection in the ATMS?
b. What are the characteristics of Waze data?
    - How does Waze coverage compare to other sources?
    - How does Waze coverage vary by time and location?
c. What is the estimated potential additional coverage that Waze can provide to the ATMS?
    - In the locations where ATMS is unable to verify Waze reports, can Waze be trusted?

This last question is a critical topic. In current ATMS settings, crowdsourced data needs validation by a second source before being trusted. This is not available in all locations and times, however. Thus, an estimation of the potential added coverage in Waze provides a ground for justifying allocating resources to developing methods that assess crowdsourced reports using historical data. One of the ultimate goals of studying crowdsourced data is to understand its characteristics profoundly enough to know when and where to rely on crowdsourced reports in locations where there are no other means for validation. Hence, this work seeks answers to the above questions in the process of finding the response to Question c. Moreover, some of the main challenges in utilizing Waze data for traffic monitoring were identified and discussed for future work.

BACKGROUND

Crowdsourced data and social media have been widely used in many areas. For instance, tweets have been used to detect earthquakes in real-time *(6)* or predict influenza outbreaks *(7), (8)*. More closely related to traffic, the Twitter-based Event detection and Analysis System (TEDAS) has been proposed by Li and colleagues *(3)*. Another work utilized Twitter to detect traffic incidents in real time *(4)*. To increase the percentage of useful tweets, Gu et al. have implemented a method to extract geolocation from the text of traffic-related tweets *(2)*. Furthermore, the validity of the traffic information acquired from social media was approved by comparing to the recorded traffic situation in London *(9)*. These applications demonstrate the potential wealth of information in crowdsourced data. Regardless of how the data are collected, however, there are challenges in using crowdsourced data that require consideration.

Although crowdsourced data usually come at a relatively inexpensive price, there are challenges in understanding and interpreting this type of data. The crowdsourced data are reported by users who might be slightly inaccurate in time or location. For users traveling on the roads at the speed of 60 miles per hour, 30 seconds' delay in reporting an incident is a 0.5-mile distance. Moreover, users might falsely assume the causes of irregular congestion and report a crash while simply stuck in traffic.

Creating a clean dataset by reconciling the variation in crowdsourced user reports of the same incident and matching these reports to incidents recorded in the ATMS represents one of the primary challenges. The matching procedures as explored in the literature are known as matching or conflation methods *(10)–(14)*. As summarized Xavier et al. *(13)*, similarity measures for point data (like the incident data in this study) are generally a combination of the following:

Geometric: Distance or area overlap

Semantic: Measures of non-geometric properties.

Context: the special relationship between objects.

Ruiz et al. added the temporal criteria into their categories as well. For point matching, using geographic distance (Euclidian distance is most common) is the most classic approach *(15)–(17)*. Adding extra information about the points when available, such as road names and direction, adds additional power to the matching function. The hybrid approach of geographic and semantic information has shown high accuracy in matching crowdsourced information *(18)*. Considering the problem at hand and the available data in this research, a hybrid approach was used to leverage geographic as well as semantic matching methods.

DATA

Waze Data

Waze is a navigation application that leverages crowdsourced user reports for providing service. Users can report traffic crashes, congestion, hazards, or police traps on the road (www.waze.com/about). The Iowa Department of Transportation (IDOT) joined the Connected Citizen Program (CCP), which is an agreement in which the city or state managers provide Waze with information on road closures and constructions and, in return, Waze provides user reports to the managers. However, since the raw Waze data contain duplicate reports for a single incident and all reports may not have high reliability, data preprocessing is necessary *(19)*. IDOT's ATMS implements stringent acceptance criteria for Waze reports before considering them for validation (filtering criteria: type = crash or reliability>=6 or report rating >=4). The reports that meet the criteria are sent to ATMS operators to verify the incident. If the incident is verified, it will be recorded in the ATMS database.

ATMS Data

The Iowa ATMS records all incidents, hazards, and congestion detected by various sensors and cameras or the reports by the highway helpers or police. The incidents in this dataset are validated by ATMS operators and thus serve as a reference for evaluating other sources of data. However, not all incidents, particularly congestion, are recorded in this dataset.

Incidents Detected from Third-Party Traffic Services Vendors

Third-party traffic services vendors such as INRIX (www.inrix.com) gather anonymized position data, which in turn provide rural and urban system-wide traffic data with reasonable accuracy *(21)*. Iowa ATMS applies a state of the art method for detecting incidents from INRIX data. This method utilizes interquartile range (IQR) of the historical speed data in each timeframe to detect outliers as described by Chakraborty and colleagues *(22)*. Threshold speeds are computed for each segment, day of the week, and 15-minute period of the day utilizing the last 8 weeks of data. More specifically, threshold = (Median - 2 × IQR) is computed for each period and an incident alarm is triggered when the real-time speed is below the corresponding threshold. The data generated from this process are another feed of data to the ATMS and a basis for comparison with Waze data.

Traffic Camera Images

Cameras mounted in various locations across Iowa are one of the main means for traffic monitoring in the ATMS. To estimate false alarms in Waze reports, this study uses screenshots of the camera video feed that are captured every five minutes. Cameras in the Des Moines, Iowa metropolitan area (56 cameras) were selected for manual labelling of road conditions. Since labeling the road conditions (particularly congestion) based on a single image is a subjective decision, the images were labelled “clear” when the road was obviously clear and no congestion or incidents were observed. The labelled road images were used to detect the false alarms in Waze reports.

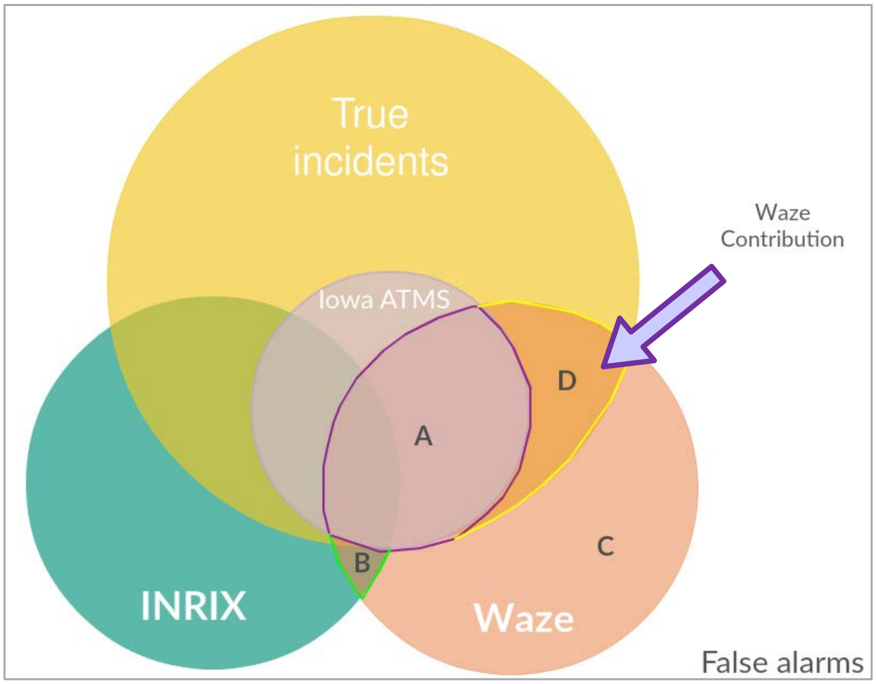


FIGURE 1 Venn diagram of the sources of traffic monitoring data, pointing to region of interest (D), the potential contribution of Waze.

Anticipated Coverage of Data Sources

In practice, each of these sources cover a portion of the true incidents; they have some overlaps, and may have false alarms as well. The Venn diagram of our data sources depicted in FIGURE 1 illustrates this relationship (circles are not drawn to scale); the characteristics of the overlapping areas are of primary interest. Iowa ATMS captures a subset of the true incidents which is validated and free from false alarms. Waze and INRIX are expected to cover some of the true incidents while having a portion of false alarms. This study is mainly focused on estimating the potential additional contribution of Waze to the ATMS (region D). It is worth noting that the exact findings of this work are applicable to states and locations like Iowa, and that depending on the number of Waze users and penetration rates, the results may vary.

EVALUATION PROCEDURE

Region (D) on the Venn diagram of our data sources (FIGURE 1) marks the potential contribution of Waze to the ATMS. However, since data on true incidents in all locations and times are not available, the existing sources were used to quantify the potential contribution and value in Waze feed. Hence, the estimation of (D) was achieved in four main steps which are explained in this section using notations from FIGURE 1. The four steps are:

1. Match Waze and ATMS incidents (A)
2. Match Waze and INRIX incidents (B)
3. Estimate the false alarms (C)
4. Estimate Waze's contribution D = Waze – $(A \cup B \cup C)$

This study focused on two main type of incidents, congestion and crashes, as the sources that most directly impact traffic. To accomplish these steps, a matching function was necessary, which is described below.

TABLE 1 Event Matching Procedure for Step 1 (ATMS and Waze matching)

Matching Levels	Criterion	Logic	Matching method	Action category
First	Time	Waze reports 20 minutes before the start and after the end time of an ATMS record	Temporal	Preprocessing
Second	Location	Crashes in a 2.5-mile radius, Congestion in 1-mile radius	Geographic	Preprocessing
Third	Road name and direction	Grouped into: Matching both and opposite direction	Semantic	Preprocessing
Fourth	Type of incident	Type, road name, and direction match	Semantic	Full/exact Match
	Type of incident	ATMS event is a crash, Jam reported in Waze, No full match exists	Semantic	Secondary Jam of a crash
	Road direction	Everything matches, Opposite direction, 1-mile radius	Semantic	Opposite direction

Matching Function

For matching incidents between sources, a hybrid method leveraging geographic and semantic matching methods was implemented. In both data sources, the road name and direction, as well as the type of the incident (i.e., crash, congestion, or stalled vehicle) were recorded. TABLE 1 presents the levels of the matching function as well as the criteria and method used in each level. The matching function first selects incidents in the temporal vicinity, then the geographic distance is examined. From spatiotemporal neighboring incidents, semantic information such as road names, direction, and type of the incident were used to mark matching incidents. The matching function introduced for this step (Match Waze and ATMS incidents) is the most comprehensive one. In the next steps, when matching with INRIX data and detecting false alarms, the match function was slightly modified to fit the semantic features of the respective data fields. TABLE 2 provides a summary of the evaluation procedure in this work and the data used in each step.

TABLE 2 Summary of the Waze Evaluation Procedure Steps

Step	Name	Venn diagram segment	Research motivation	Data	
				Time	Location
0	Exploratory analysis	-	Waze and ATMS reports based on: - Time of day - Region - Road type - Etc.	2016 entire year	entire state of Iowa
1	match Waze and ATMS	A	- Waze and ATMS overlap - Redundancies - Influential factors in Waze coverage	2016 entire year	entire state of Iowa
2	match Waze and INRIX	B	- ATMS and INRIX overlap - Waze vs INRIX contribution to ATMS	October 2016	entire state of Iowa
3	Estimate the false alarms in Waze	C	- % of Waze reports when road is clear (False alarms)	October 2016	Des Moines Area
4	Estimate Waze's contribution	D	- The information that Waze can add		

RESULTS

Exploratory Waze Data Analysis

To initiate the evaluation, an exploratory data analysis was performed to better understand the Waze and ATMS data. The exploratory analysis looked into the pure number of reports regardless of the matching percentages or potential duplicates, to provide a high-level understanding of the two sources of data.

Sources of Incident Detection in the ATMS

Waze has been used as a source of incident detection in the IDOT ATMS since September 2015. As depicted in FIGURE 2, part (a), among the 23 sources of detection in the Iowa ATMS, law enforcement (which includes 911 calls, County Sheriff, State Patrol, etc.) contributes the highest number of incidents in the ATMS. Interestingly, Waze reports (detection source for 13.4% of the ATMS records) rank fourth in detection sources, after law enforcement, CCTV, and highway helpers. Comparing the operation and maintenance cost of each of the first three sources, Waze has a considerable contribution as a "free" detection source. However, in the current ATMS settings, the Waze reports need to be verified, usually by one of the top three sources before being trusted.

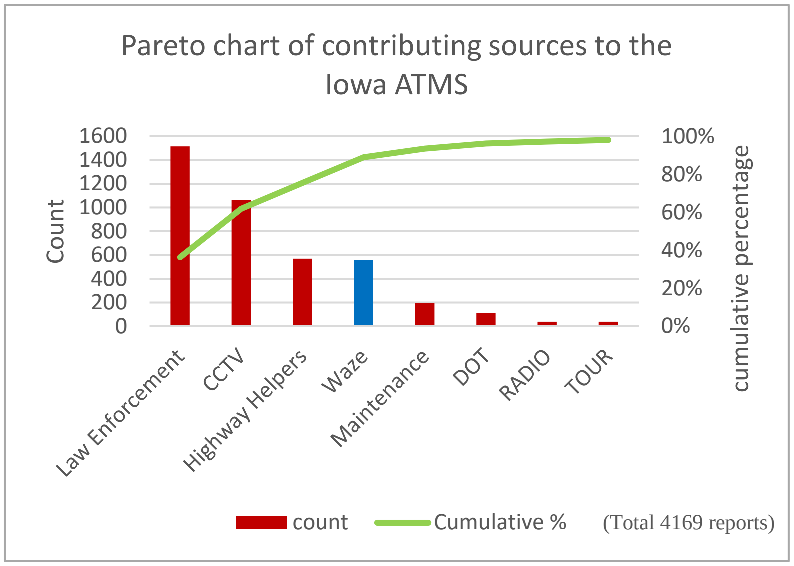
Incident Reports in Distinct Locations and Road Types

The location of each report was mapped to the demographics of the region based on 2010 census data (23). Every county is grouped by their population as either metropolitan (>50,000), micropolitan (10,000-50,000), urban cluster (2,500-50,000), or rural (any non-urban region is considered rural). This analysis provides an insight into the spread and coverage of each source of data. As depicted in FIGURE 2 part (b), the ATMS has recorded almost no congestion incidents (jams) outside of the metro area. This is while there are many congestion incidents reported in Waze from the urban clusters and rural areas (even off the interstates). In addition, the considerably larger numbers of reports on the interstates show the concentration of reports in both sources. This chart indicates the type of incident and locations where Waze could best contribute to the ATMS.

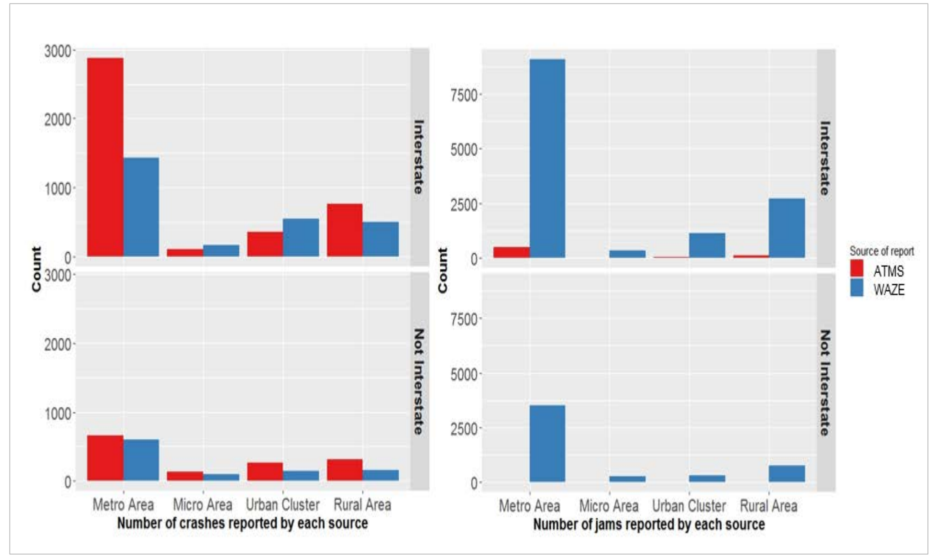
Impact of Time on the Waze Reports

To evaluate how the crowdsourced data reflect the reality on the roads, the number of reports in each hour of the day were compared and it was expected that the crowdsourced data resemble the ATMS records. As observed in FIGURE 2 part (c), both data sources tend to have a higher frequency of crash records during the rush hours. However, between midnight and 6 a.m., although ATMS shows 50-100 crash records, there are less than 10 Waze crash reports in the same time. The proportion of the number of Waze to ATMS crash reports during these hours (mean 9%) showed a statistically significant difference from the same proportion for other hours of the day (mean 37%). This indicates that Waze is not be a reliable detection source during midnight to 6 a.m. This observation aligns with the fact that during these hours there are fewer drivers on the roads and consequently fewer Waze users that might observe and report an incident.

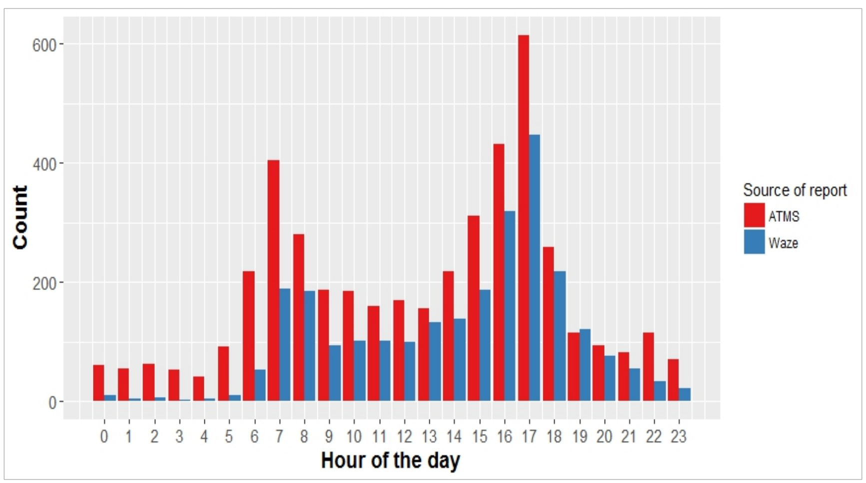
Otherwise, the number of crashes reported in each hour of the day (from 6 a.m. to 11 p.m.) was highly correlated (R^2=0.9) between ATMS and Waze; as depicted in FIGURE 2 part (d). Thus, the number of Waze crash reports during the day follow the reality of the roads.



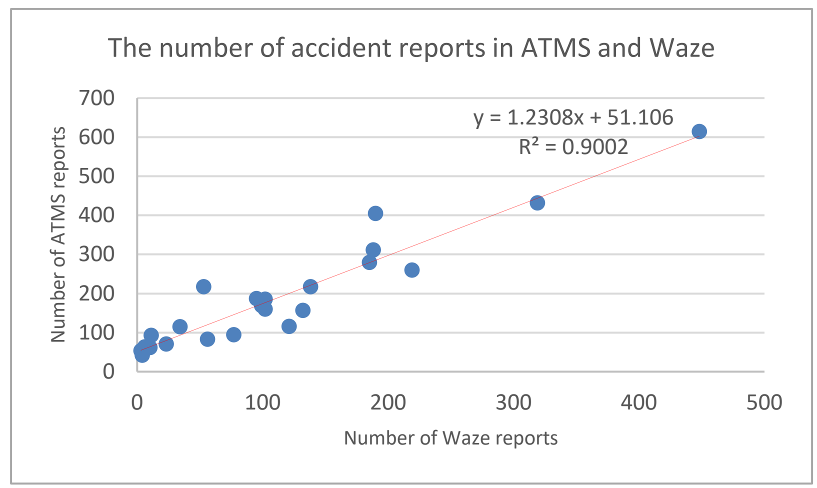
(a)



(b)



(c) Total crash counts in all weekdays of year 2016 per hour of the day



(d)

FIGURE 2 Exploratory data analysis results, comparing number of reports in Waze and ATMS. All data are from 2016

Evaluation and Comparison

Step 1: The ATMS incidents that were reported in Waze (Estimating A: Waze ∩ ATMS)

This step compares Waze reports to the ATMS reports as source of validated events. The percentage of matching incidents in both sources answers questions regarding the reliability of Waze reports, while leading to the estimation of the potential contribution of Waze.

Using the described matching function, overall the congestion and crashes reported in Waze covered 43.2% of the ATMS records. The matching percentage by each type of incident is presented in TABLE 3.

In Iowa, similar to many other Midwestern U.S. states, traffic is not a daily concern for most people, and thus fewer people are familiar and active users of Waze, compared to more populated cities and states. Yet, the number of matched reports are interesting, considering a single crowdsourced feed of data has captured 43.2% of ATMS records.

TABLE 3 ATMS-Waze Matching Percentage by Report Type

Type of incident	Total reports in ATMS	% matched with Waze
Crashes	3713	42.1 %
Congestion	456	58.5 %
Stalled vehicles	12552	43.0 %

What factors contribute to an incident being reported in Waze in the Metro area?

To find the variables which have a statistically significant influence in determining whether an ATMS incident is reported in Waze, a binomial logistic regression was conducted. The binomial logistic regression was performed to ascertain the effects of day of the week, hour of the day, incident type, and the road type on the likelihood that an event covered by an ATMS record would be covered by Waze as well. The logistic regression model was statistically significant, $\chi 2(31) = 450.2$, $p << .001$. The model explained 20.0% (Nagelkerke R^2) of the variance in the matched instances and correctly classified 63.6% of cases. Of the thirty-one predictor variables (factors converted to dummy variables), the statistically significant ones were related to time and road type (as shown in TABLE 4). The incident type did not indicate a significant impact in this model.

Since the road type turned out to be a significant contributing variable to the model, another logit model was tested using the interstate road names (9 variables) in the metro area as new variables, to investigate if a certain road significantly impacts the chance of an ATMS report being covered in Waze. None of the major interstates indicated a significant impact.

TABLE 4 Significant Influencers in ATMS-Waze Matching (indicates significance level of 0.001)**

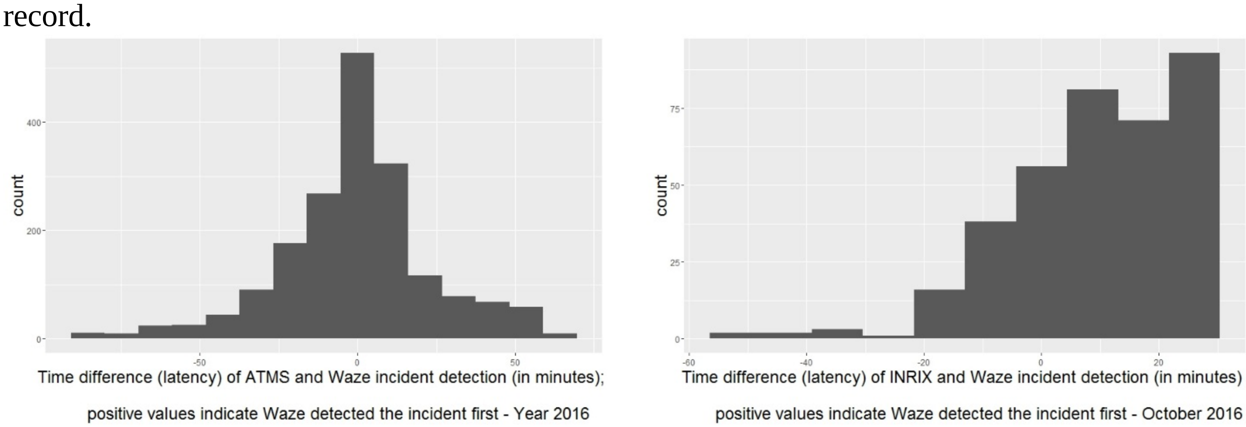
Variable group	Variable	Estimate	P-Value	**Variable definition**	
Time of the Day	07:00-08:00	1.5444	< .0001 **	07:00 – 09:00	Morning rush hour
	08:00-09:00	0.9143	.0003		
	11:00-12:00	0.6435	.0267	11:00 – 13:00	Lunch time
	12:00-13:00	0.7137	.0135		
	14:00-15:00	0.9792	.0004	14:00 – 19:00	Afternoon
	15:00-16:00	0.8815	.0006		
	16:00-17:00	1.5484	< .0001 **		
	17:00-18:00	1.5602	< .0001 **		
	18:00-19:00	0.8350	.0015		
	20:00-21:00	0.7376	.0333	20:00 – 21:00	Evening
Road type	Interstate or not	0.9083	< .0001 **	Interstate/Freeway or not**	

What Percentage of Waze Was Covered in ATMS? And Were There Redundant Reports?

Only 14.6% of the total Waze reports were matched with incidents in the ATMS records (36.8% for the crashes and 10.0% of the congestion). Thus, it is critical to investigate the unmatched Waze data to estimate the potential added coverage of Waze.

It was also found that on average, each ATMS report matched to 1.9 Waze reports, indicating the redundancy rate in Waze data. The median is 1 report, mean is 1.9, and 80% of the reports have two or fewer matches in Waze.

To examine the accuracy of the matching in distance, the 95% confidence interval for the distance between the matched Waze report and the ATMS record was calculated as .36 to .39 miles. Evaluating the time accuracy of the matches, the time difference (latency of the reports) was calculated. As depicted in FIGURE 3 (a), the time difference forms a bell-shaped distribution around –0.22 minutes (95% CI, –1.3 to .8 minutes), which is slightly skewed to left. Slightly more than half of the matched incidents were detected earlier in Waze than the ATMS record.



(a) (b)

FIGURE 3 Waze incident detection time compared with ATMS and INRIX.

Step 2: Estimating the Common Incidents in INRIX and Waze (B)

Although the INRIX reports are not all validated, the overlap of Waze and INRIX reports increases the plausibility of an actual incident occurrence in the same time and location. To control for weather effects in our results, one month with relatively stable weather and about average matching percentages from Waze and ATMS incidents, was desired. October fulfilled the desired properties; therefore, October 2016 data was used for this part. Having applied incident detection method in Iowa ATMS, as described by Chakraborty et al. *(22)*, the incidents were detected from INRIX.

Using the described matching function in TABLE 1, 48% of Region A of FIGURE 1 (Waze ∩ ATMS) was also matched with INRIX. This result implies that the INRIX feed had detected about half of the common incidents in Waze and ATMS, adding to the validity of the INRIX detected incidents.

To estimate Region B on the Venn diagram, the overlap of the Waze reports with the INRIX data was evaluated. The results indicated 16.8% of Waze reports were matched to INRIX. The time difference between Waze reports and matched incidents demonstrated that on average, INRIX reports were detected 9.8 minutes later (95% CI, 8.25 to 11.36) than Waze reports (FIGURE 3 (b)).

Step 3: Estimating the False Alarms in the Metro Area (C)

Region C of FIGURE 1 represents false alarms from Waze, i.e., reports of incidents that did not actually exist. To estimate the number of false alarms in Waze, manually labelled images from IDOT cameras in the Des Moines metro area were used. The results indicated that overall, only one of the 319 Waze reports in October 2016 and locations was a false alarm. This accounts for 0.3% of the reports.

Although our false alarm definition is not strict (a false alarm is when the road is visibly clear and there is a Waze incident report), the false alarm rate is interestingly lower than expectations. It is worth mentioning a great portion of Waze reports are congestion reports that DOT is not particularly interested in recording. Yet, this is an important finding to understand the validity of these crowdsourced Waze reports.

Step 4: Estimating the Waze Contribution (D)

The final step in the process is to estimate the Waze contribution, or Region D on the Venn diagram of FIGURE 1. Based on the following calculations, 68.3% of the Waze incidents were estimated to be the additional information that Waze can contribute. Once accounting for the number of redundant reports (1.9 redundant reports was rounded up to 2.0 for a more conservative estimation), 34.1% of the Waze's crash and congestion reports (7387 instances which are mainly congestion reports) were potential incidents that were not recorded by the current sources of the ATMS.

$$D = (A \cup B \cup C)' = 100\% - (14.6\% + \ 16.8\% + 0.3\%) = 68.3\%$$

$$Accounting\ for\ redundancies: \frac{68.3\%}{2} = 34.1\%$$

$$Number\ of\ incidents: 34.1\%\ (total\ congestion\ and\ crash\ reports\ in\ Waze) = .341\ \times 21662 \cong 7387$$

To further estimate the potential additional crash coverage in Waze data, the proportion of crash reports among Region D within the 2016 data was 12% of all Region D incidents. Assuming this percentage is uniform in the unmatched Waze reports, this yields about 904 crashes in year 2016 (12% × 7387 reports) which are either potentially missed or recorded with different labels by the ATMS. These numbers provide an estimate of Waze's potential contribution to traffic coverage in the state of Iowa.

Note that the Waze congestion reports don't come with the recurring or non-recurring labels. Thus, many of the congestion reports might be recurring traffic patterns. Although the ATMS operators are not concerned with the recurrent congestions, the Waze reports still provide invaluable information about the traffic conditions. Moreover, records on all types of traffic incidents provide training data for classification models that can distinguish recurring and non-recurring congestion.

Comparing Waze with Findings about Twitter

Now that the contribution of Waze has been estimated, it is worth examining its performance with other data sources of data. The work of Gu et al. *(2)* provided information about traffic incidents extracted from Twitter in Pennsylvania. Comparing some of the findings about Twitter with Waze was insightful. Like the present results with Waze, Gu et al.'s analysis showed Twitter to be less reliable during night hours. Also, most of the tweets were during the peak traffic hours. Gu et al. reported an average of 1.6 Twitter-reported incidents per unique incident. This number was estimated 1.9 reports for Waze, indicating that redundant reports are a common challenge in other crowdsourced data feeds.

Summary of the Findings

Based on the quantitative analysis of Waze data, FIGURE 4 is an updated view of the Venn diagram that better illustrates the relationship and overlap of the three sources of data. In this another aspect of the challenge is demonstrated. Although there exists a set of true incidents (the yellow circle), not all of them are known through the existing means. Thus, when evaluating the potential of Waze this challenge should be acknowledged. Note that the (D) region in the figure is now split into sections [3] and [4]. The overlap of (D) and Verifiable incidents [3] shows the incidents that are verifiable through other existing means (particularly CCTV cameras). Part [4] in region (D) are reports that can potentially be valid incidents, and there are currently no cameras or other means to verify their accuracy. Based on this work, it is believed that a considerable percentage of the potential incidents in (D) provide invaluable information to the ATMS.

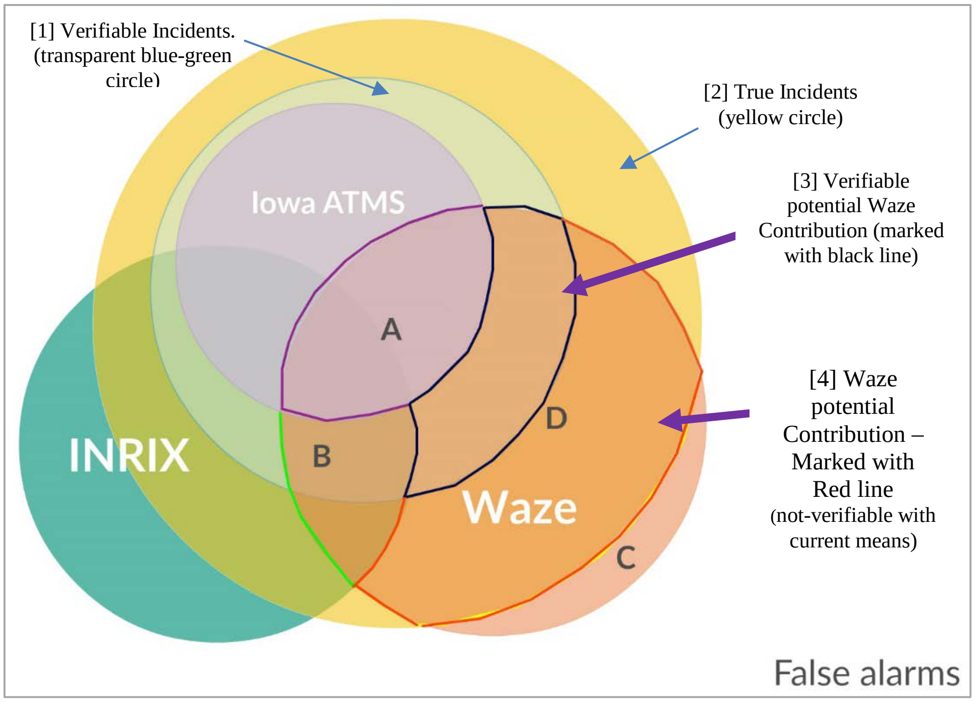


FIGURE 4 Updated Venn diagram based on the analysis, the regions are drawn closer to scale. Region D, the estimated contribution of Waze to the ATMS, is divided into verifiable and non-verifiable regions.

DISCUSSION AND CONCLUSION

This research evaluated crowdsourced traffic incident reports from Waze, to study its characteristics as a data source. This section provides a summary of the findings.

How does Waze compare to the existing sources?

The reliability of crowdsourced incident reports from Waze was affirmed with the matching percentages between Waze and validated ATMS (42.3% of ATMS records) and INRIX data. In the Iowa ATMS, 13.4% of the recorded congestion and crashes were initially detected by Waze reports, making it the fourth most contributing source of incident detection. These findings indicate the reliability and competent coverage of crowdsourced traffic incident reports like Waze.

What are the characteristics of Waze data?

Waze incident reports indicated a wide spread coverage of instances in most locations and road types, particularly for reported congestion. The quality of the reports did not depend on the day of week or a specific roadway. On the other hand, the analysis indicated in the less crowded hours of the day (12 a.m. to 6 a.m.), Waze reports are not a reliable source for monitoring road conditions.

What is the estimated potential additional coverage that Waze can provide to the ATMS?

The potential additional coverage that Waze can provide to the ATMS was estimated to be 34.1% of Waze reports, which accounts for 7387 incidents per year (from which 904 were estimated to be crash reports), making it a valuable source for traffic managers to invest.

Overall, it can be concluded that crowdsourced reports like Waze are invaluable sources of information for traffic monitoring with broad coverage, timely response time, and reasonable accuracy. Integrating this source of data into the ATMS feeds provides significant contributions to the traffic monitoring coverage.

However, there are challenges in working with this crowd-based data, including redundancies, inaccuracies, and mismatches in report types, as well as the need for report reliability estimation. Therefore, preprocessing and validating such data is necessary and requires resource investment. The crowdsourced data, on the other hand, are typically provided freely (or at a low cost) to the ATMS managers. Compared to the immense cost of installation and maintenance of other data sources (sensors, third party probe data, or even law enforcement reports), raw Waze data is available for free. This analysis indicated potential valuable incident information from cleaned and processed Waze data. Therefore, a short-term investment in human resources to establish an infrastructure for eliciting valuable information from Waze data seems economically justifiable. This infrastructure would include models to address the redundancy issue and to automatically estimate the reliability of the reports, which are directions for future work.

Although the exact value of Waze data would vary for different regions and over time, these numbers in a less congested U.S. state seem impressive, and the techniques used in this research for Waze data evaluation could be applied to any region. Moreover, knowing the number of active Waze users in different regions would add a valuable basis for comparative across multiple regions.

ACKNOWLEDGEMENTS

This material is based on the work funded by Iowa Department of Transportation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Iowa Department of transportation.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: Amin-Naseri, Sharma, Gilbert, and Hong; data collection: Amin-Naseri and Chakraborty; analysis and interpretation of results: Amin-Naseri, Sharma, Gilbert, and Hong; draft manuscript preparation: Amin-Naseri, Gilbert, and Chakraborty. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. J. Yoon, B. Noble, and M. Liu, “Surface street traffic estimation,” *Proc. 5th Int. Conf. Mob. Syst. Appl. Serv. - MobiSys ’07*, p. 220, 2007.
2. Y. Gu, Z. (Sean) Qian, and F. Chen, “From Twitter to detector: Real-time traffic incident

detection using social media data," *Transp. Res. Part C Emerg. Technol.*, vol. 67, pp. 321–342, 2016.
3. R. Li, K. H. Lei, R. Khadiwala, and K. C. C. Chang, "TEDAS: A twitter-based event detection and analysis system," *Proc. - Int. Conf. Data Eng.*, pp. 1273–1276, 2012.
4. E. D'Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni, "Real-Time Detection of Traffic from Twitter Stream Analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2269–2283, 2015.
5. C. Seeger, C. Lillehoj, S. Wilson, and A. Jensen, "Facilitated-VGI, Smartphones and Geodesign: Building a Coalition while Mapping Community Infrastructure," in *Digital Landscape Architecture 2014 - Landscape Architecture and Planning: Developing Digital Methods in GeoDesign. Zurich, Switzerland*, 2014, pp. 300–308.
6. T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors," *Proc. 19th Int. Conf. World Wide Web*, pp. 851–860, 2010.
7. E. Aramaki, "Twitter Catches The Flu : Detecting Influenza Epidemics using Twitter The University of Tokyo The University of Tokyo National Institute of," *Comput. Linguist.*, vol. 2011, pp. 1568–1576, 2011.
8. A. Signorini, A. M. Segre, and P. M. Polgreen, "The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic," *PLoS One*, vol. 6, no. 5, 2011.
9. E. Steiger, B. Resch, J. P. de Albuquerque, and A. Zipf, "Mining and correlating traffic events from human sensor observations with official transport data using self-organizing-maps," *Transp. Res. Part C Emerg. Technol.*, vol. 73, pp. 91–104, 2016.
10. C. Y. Goh, J. Dauwels, N. Mitrovic, M. T. Asif, A. Oran, and P. Jaillet, "Online map-matching based on hidden markov model for real-time traffic sensing applications," *Itsc'12*, vol. 117543, 2012.
11. M. Sester, J. J. Arsanjani, R. Klammer, D. Burghardt, and J.-H. Haunert, "Integrating and Generalising Volunteered Geographic Information," in *Abstracting geographic information in a data rich world*, D. Burghardt, C. Duchene, and W. Mackaness, Eds. Springer International Publishing, 2014, pp. 119–155.
12. B. Yang and Y. Zhang, "Pattern-mining approach for conflating crowdsourcing road networks with POIs," *Int. J. Geogr. Inf. Sci.*, vol. 8816, no. June, pp. 1–20, 2015.
13. E. M. A. Xavier, F. J. Ariza-López, and M. A. Ureña-Cámara, "A Survey of Measures and Methods for Matching Geospatial Vector Datasets," *ACM Comput. Surv.*, vol. 49, no. 2, p. Article 39, 2016.
14. J. J. Ruiz, F. J. Ariza, M. a. Ureña, and E. B. Blázquez, "Digital map conflation: a review of the process and a proposal for classification," *Int. J. Geogr. Inf. Sci.*, vol. 25, no. 9, pp. 1439–1466, 2011.
15. C. Beeri, Y. Kanza, E. Safra, and Y. Sagiv, "Object fusion in geographic information systems," in *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, 2004, pp. 816–827.
16. E. Safra, Y. Kanza, Y. Sagiv, C. Beeri, and Y. Doytsher, "Location-based algorithms for finding sets of corresponding objects over several geo-spatial data sets," *Int. J. Geogr. Inf. Sci.*, vol. 24, no. 1, pp. 69–106, 2010.
17. C. Beeri, Y. Doytsher, Y. Kanza, E. Safra, and Y. Sagiv, "Finding corresponding objects when integrating several geo-spatial datasets," in *Proceedings of the 13th annual ACM*

international workshop on Geographic information systems, 2005, pp. 87–96.

18. G. McKenzie, K. Janowicz, and B. Adams, “A weighted multi-attribute method for matching user-generated Points of Interest,” *Cartogr. Geogr. Inf. Sci.*, vol. 41, no. 2, pp. 125–137, 2014.
19. B. M. Pack and N. Ivanov, “Are You Going My Waze ? Practical Advice for Working,” *ITEjournal*, vol. 87, no. 2, pp. 28–35, 2017.
20. H. Li, S. M. Remias, C. M. Day, M. M. Mekker, J. R. Sturdevant, and D. M. Bullock, “Shock Wave Boundary Identification Using Cloud-Based Probe Data,” *Transp. Res. Rec. J. Transp. Res. Board*, no. 2526, pp. 51–60, 2015.
21. A. Haghani, M. Hamedi, and K. F. Sadabadi, “I-95 Corridor coalition vehicle probe project: Validation of INRIX data.” I-95 Corridor Coalition, 2009.
22. P. Chakraborty, J. R. Hess, A. Sharma, and S. Knickerbocker, “Outlier Mining Based Traffic Incident Detection Using Big Data Analytics,” in *Transportation Research Board 96th Annual Meeting*, 2017.
23. United States Census Bureau, “Urban Area relationship files,” 2010. (Online). Available: https://www.census.gov/geo/maps-data/data/ua_rel_download.html. (Accessed: 01-Jan-2017).