

# Fuzzy Logic–Based Mapping Algorithm for Improving Animal-Vehicle Collision Data

Yunteng Lao<sup>1</sup>; Yao-Jan Wu, A.M.ASCE<sup>2</sup>; Yin Hai Wang, M.ASCE<sup>3</sup>; and Kelly McAllister<sup>4</sup>

**Abstract:** Animal-vehicle collisions (AVCs) cause hundreds of human and wildlife animal fatalities and tens of thousands of human and wildlife animal injuries in North America. It is estimated that AVCs cause more than \$1 billion in property damage each year in the United States. Further research efforts are needed to identify effective countermeasures against AVCs. Two types of data have been widely used in AVC-related research: collision reported (CRpt) data and carcass removal (CR) data. However, previous studies showed that these two data sets are significantly different, implying the incompleteness in either set of the data. Hence, this study aims at developing an algorithm to combine these two types of data to improve the completeness of data for AVC studies. A fuzzy logic–based data mapping algorithm is proposed to identify matching data from the two data sets so that data are not overcounted when combining the two data sets. The membership functions of the fuzzy logic algorithm are determined by a survey of the Washington State Department of Transportation CR staff. As verified by expert judgment collected through another survey, the accuracy of this algorithm was approximately 90%. Applying this algorithm to the WSDOT data sets identified that approximately 25~35% of the CRpt data records have matching pairs in the CR data. Compared with the original CR data set, the combined data set has 15~22% more records. The proposed algorithm provides an effective means for merging the CRpt data and the CR data. Such a combined data set is more complete for wildlife safety studies and may provide additional insights into understanding the issue of AVCs. DOI: 10.1061/(ASCE)TE.1943-5436.0000351. © 2012 American Society of Civil Engineers.

**CE Database subject headings:** Vehicles; Traffic accidents; Data processing; Traffic safety; Fuzzy sets.

**Author keywords:** Animal-vehicle collision; Carcass removal data; Fuzzy logic; Data mapping; Traffic safety.

## Introduction

The continuing growth in both animal and motor vehicle populations has resulted in more and more animal-vehicle collisions (AVCs) (Curtis and Hedlund 2005). Deer-vehicle collisions are a common type of AVC. In Washington, approximately 3,000 collisions occur yearly with deer and elk on state highways (Wagner and Carey 2006). Romin and Bissonette (1996) estimated that at least 500,000 deer-vehicle collisions occurred nationwide in 1991. It is known that AVCs cause significant damage to human, property, and wildlife. Approximately 200 people are killed and 20,000 people injured each year in AVCs in the United States (Huijser et al. 2007). Property damage alone from AVCs exceeds US\$1 billion annually. Wildlife animals die immediately or shortly in most AVCs (Allen and McCullough 1976). Additionally, AVCs may reduce the population level of some precious species (Van der Zee et al. 1992;

Huijser and Bergers 2000) or even cause a serious decrease in population survival probability (Proctor 2003). All of these statistics imply that effective countermeasures against AVCs are urgently needed to mitigate AVCs.

To make a good use of limited safety improvement resources, it is essential to identify the factors associated with AVCs. Mathematical modeling techniques and statistical analysis are typically adopted to extract factors associated with AVCs from the observed AVC data, traffic data, and roadway geometric and environmental data. The quality of the relevant data sets is crucial for a reliable analysis. One of the major issues traffic safety researchers have been facing is that the AVC data are usually inaccurate or incomplete and may result in erroneous conclusions if such data problems are not properly addressed. Therefore, data quality control is very important for reliable traffic safety analysis and modeling. This research focuses on the issue of incomplete data and tends to provide a more complete data set.

In most AVC studies, two types of AVC data are usually used: collision reported (CRpt) data and carcass removal (CR) data, as emphasized in a National Cooperative Highway Research Program (NCHRP) report by Transportation Research Board (Huijser et al. 2007). In Washington, the CRpt data are collected by police officers or reported by citizens, whereas the CR data are collected by the maintenance team of the Washington State Department of Transportation (WSDOT). Because the two data sets are collected by different agencies using varying methods, data integration and interpretation have been challenging. Therefore, most previous AVC studies used either CRpt data or CR data, treating the two data sets separately. For example, Hubbard et al. (2000), Malo et al. (2004), and Seiler (2005) conducted AVC analyses on the basis of CRpt data, whereas Reilley and Green (1974), Allen and McCullough (1976), Knapp and Yi (2004), and Lao et al. (2011b) employed

<sup>1</sup>Doctoral Student, Dept. of Civil and Environmental Engineering, Univ. of Washington, P.O. Box 352700, Seattle, WA 98195-2700 (corresponding author). E-mail: laoy@u.washington.edu

<sup>2</sup>Assistant Professor, Dept. of Civil Engineering, Saint Louis Univ., 3450 Lindell Boulevard, McDonnell Douglas Hall Room 2051, St. Louis, MO 63103. E-mail: yaojan@slu.edu

<sup>3</sup>Professor, Dept. of Civil and Environmental Engineering, Univ. of Washington, P.O. Box 352700, Seattle, WA 98195-2700. E-mail: yinhai@u.washington.edu

<sup>4</sup>Habitat Connectivity Biologist, Washington Dept. of Transportation, 310 Maple Park SE, Olympia, WA 98504-7331. E-mail: mcallke@wsdot.wa.gov

Note. This manuscript was submitted on November 30, 2010; approved on September 12, 2011; published online on September 14, 2011. Discussion period open until October 1, 2012; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Transportation Engineering*, Vol. 138, No. 5, May 1, 2012. ©ASCE, ISSN 0733-947X/2012/5-520-526/\$25.00.

CR data. Only a few studies considered these two kinds of data together (Lao et al. 2011a).

On the basis of the findings of a survey conducted by this study, CR professionals at WSDOT basically agree that more than 90% of the carcasses removed from the road are likely involved in accidents, (The remaining carcasses could result from natural death.) Thus, these two sets of data should overlap to a large extent. However, previous studies (Romin and Bissonette 1996; Knapp et al. 2007) found that they are significantly different. This implies that the two sets of data complement each other and should be combined to improve the quality of AVC data. Analyses that use solely CRpt data or CR data may result in biased results. On the basis of previous studies (Hauer and Hakkert 1989; Elvik and Mysen 1999; Ye and Lord 2011), crash data are often underreported. This issue could yield significant biases in crash probability prediction (Hauer and Hakkert 1989; Ye and Lord 2011). Combining CRpt data and CR data together would be useful to provide a more complete data set and thus reduce the negative effect on the modeling analysis. Methods are needed to properly merge the two data sets.

Previous studies by Johnson and Walker (1996) and the National Highway Traffic Safety Administration (NHTSA) (2010) attempted to merge crash reports with the data retrieved in the Crash Outcome Data Evaluation System (CODES) by using probabilistic linkage. The merged data sets provide users with more information for each record (e.g., medical data combined with and financial outcome information). However, this method focuses only on increasing the number of attributes for each data record rather than increasing the number of records. Merging two incomplete data sets to increase the data size (number of records) and data quality (number of attributes) is critical and is the main objective of this paper. To achieve these objectives, a fuzzy logic-based data mapping algorithm is developed to combine CR data and CRpt data.

The paper presents a data overview and relevant issues in the next section, followed by an introduction to the fuzzy logic algorithm. Next, a case study is conducted to illustrate the decision-making process using the fuzzy logic-based approach. Then, the proposed methodology will be verified using the expert judgment data collected from a survey in WSDOT, followed by conclusions.

## Research Data

As mentioned, two types of AVC data are commonly used in AVC modeling and analysis. This study uses two data sets collected in Washington to demonstrate the fuzzy logic-based data mapping algorithm. The CRpt data can be extracted from the Washington accident file provided by the Highway Safety Information System (HSIS), which is operated by the University of North Carolina Highway Safety Research Center and the LENDIS Corporation under a contract with the Federal Highway Administration (FHWA) (U.S. Dept. of Transportation FHWA 2009). The HSIS collision data of Washington were compiled from state trooper-filed field reports and citizen reports. The AVC records in the HSIS database have no detailed animal type information other than “domestic” or “nondomestic.” However, they do have other detailed information, such as collision time and weather. The CR data used in this study were provided by the maintenance team of WSDOT. This data set contains detailed information about animal species, such as mule deer, white-tail deer, and elk.

Ten state routes (SRs) (U.S. Route 2, SR 8, U.S. Route 12, SR 20, Interstate 90, U.S. Route 97, U.S. Route 101, U.S. Route 395, SR 525, and SR 970) with relatively high AVC rates in the past several years were chosen as the study routes for this case study following the recommendation from WSDOT. Fig. 1 shows the

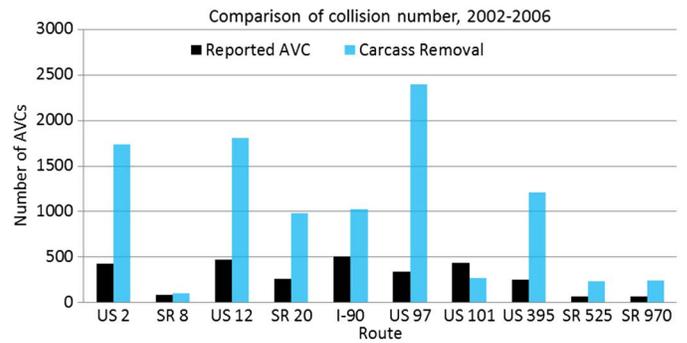


Fig. 1. Comparisons of total number of records between two data sets for each study route during 2002–2006

total numbers of records in each data set over a 5-year period (2002–2006) on each of the study routes. The CRpt data contain only “nondomestic” animals, whereas the CR data contain only “deer” and “elks” in Fig. 1. It is obvious that the CRpt and CR data sets are substantially different. The number of CR records is typically more than that of the CRpt data on each route except for U.S. Route 101. The CRpt data may likely underestimate the frequency of these types of collisions.

## Data-Matching Issues

Because the two sets of data overlap to a certain extent, attention must be paid to avoid duplicating the same accident records. One of the most effective ways to determine if CRpt data match data in the CR data set is to compare any similarities in occurrence time and location. Generally, CRpt data are reported on the same day when an AVC occurs; however, the carcasses are picked up by the WSDOT maintenance staff depending on when the carcass is found. Theoretically, the carcass pickup day should be the same as the day when the AVC is reported. In reality, however, a perfect match between two data sets rarely happens. The record of the same event typically looks different in time and/or location in each data set. Such differences can be explained as follows:

- Dead animals in a ditch or in the tall grass may not be spotted for several days. This is because animals that die off the roadway or far away from any residences may not be removed for several days or even longer. In essence, these are cases in which the dead animal is not an immediate hazard to motorists and/or not an obvious and unpleasant sight. Therefore, reporting and/or response can be delayed or nonexistent.
- The WSDOT maintenance staff generally does not remove carcasses over weekends, except during the winter. During the winter months, the WSDOT maintenance team patrols several times every day and night so the carcasses can be spotted sooner. However, heavy snowfalls may completely hide carcasses and delay the removal process for multiple months. During the summer months, the WSDOT maintenance staff does not patrol the highways every day because of other priority duties. In this case, a carcass not affecting traffic movement significantly may not be reported or identified immediately and hence may not be picked up in a couple of days.
- In addition, human errors may be introduced to the two data sets when the records are input manually.

In summary, not all animal carcasses were removed and reported by transportation agencies. Meanwhile, not all AVCs were properly reported and recorded. Therefore, both data sets are very likely to underestimate the actual number of AVCs to some extent. Combining the two data sets will make the research data more complete and

hence provide a better information base for AVC studies. Specifically, combining these two data sets will extend the data breadth (increase samples).

## Methodology

The same AVC captured by both data sets may have different values for date and milepost for several reasons. This variability may not be solved by a precise quantitative matching technique. Rather, it requires qualitative inferences in addition to quantitative analyses to determine matching data. The fuzzy logic-based data mapping algorithm has proven to be an effective way to deal with such problems related to linguistic vagueness and human factors (Zhao 1997). Fuzzy logic mapping algorithms have been widely used in various fields of transportation engineering, such as ramp metering (Taylor and Meldrum 1998), speed control systems (Rao and Saraf 1995), and map-matching issues (Syed and Cannon 2004; Mohammed et al. 2006). Generally, the fuzzy logic mapping algorithm involves three major steps (Chen and Pham 2001): (1) fuzzification: converting the quantitative inputs into natural language variables; (2) rule evaluation: implementing the mapping logic; and (3) defuzzification: converting the qualitative rule outcomes into a numerical output. This paper presents a fuzzy logic-based mapping algorithm using these three steps as follows.

### Fuzzification

Three attributes are used in the data mapping process: animal type, date, location. The animal categories for CRpt data and CR data are a little different. The “nondomestic” animal type reported in AVC data is matched with the three deer types and elk in CR data. After the animal type has been matched, this algorithm will consider only “date difference” and “location difference” as the inputs.

Date difference refers to the difference between the date when the carcass was collected and the date when the collision was recorded in the CRpt data set. The date recorded in the CRpt data is usually the same as the day of collision, whereas the date in the CR data set is the same or later than the date of the collision because a carcass cannot be collected until after the collision has happened. Therefore, the date difference is mathematically defined as

$$\text{Date difference} = \text{Date in the CR data set} - \text{Date in the AVC data set} \quad (1)$$

Location difference is the milepost difference between the CRpt location and the location where the carcass was collected.

The SR numbers in a data pair are required to be identical before mileposts can be compared. Therefore, the location difference is defined as the absolute value between the milepost in the AVC data set and the milepost in the CR data set:

$$\text{Location difference} = |\text{Milepost in the AVC data set} - \text{Milepost in the CR data set}| \quad (2)$$

These inputs are then translated into four fuzzy classes on the basis of the level of difference: small, medium, big, and very big (S, M, B, and VB). A difference of VB presents the situation in which the input is larger than a critical range. For example, if the location difference is considered only within 4.83 km (3 mi), a 8.05-km (5-mi) difference will be marked as VB.

A membership function (Li and Yen 1995) for each class needs to be determined during the fuzzification step. A membership function describes the membership degree, defined as the truth extent to the respondent class, and its value ranges from 0–1. Most research (Taylor and Meldrum 1998; Swain 2006; Naso et al. 2006) has assumed the membership function to be a triangle for simplification and has designed it on the basis of subjective experiences. However, the triangular membership functions may be too simple to accurately reflect the reality. Therefore, this study adopted a survey-based method (Li and Yen 1995) to determine the membership functions for the fuzzy classes. Details about the membership function determination process are described in the “Application” section.

### Rule Design

Fuzzy logic rules are needed for mapping inputs to outcomes. Eleven rules, shown in Table 1, are designed for this algorithm. The default rule weights reflect the relative importance of the rules. As mentioned earlier, the two inputs are milepost difference and date difference. The matching output between the AVC and the CR data sets is the outcome, which is represented by six fuzzy classes: very very low (VVL), very low (VL), low (L), medium (M), high (H), and very high (VH). For example, VVL presents the situation in which the output class is very close to zero. In other words, the candidate data pair is too different to be a possible matching pair.

The output class decreases with the increase of milepost difference and/or date difference. Rules 1 through 9 cover normal matching conditions. For example, Rule 9 could be interpreted as follows: If the milepost difference is B and the date difference is B, then their matching degree (MD) is VL. Rules 10 and 11 deal with

**Table 1.** Rule Base for Fuzzy Mapping Algorithm

Rule	Default rule weight	Input class		Output class
		Milepost difference	Date difference	
1	1	Small	Small	Very high
2	1	Small	Medium	High
3	1	Small	Big	Medium
4	1	Medium	Small	High
5	1	Medium	Medium	Medium
6	1	Medium	Big	Low
7	1	Big	Small	Medium
8	1	Big	Medium	Low
9	1	Big	Big	Very low
10	1	Very big	Any input class	Very very low
11	1	Any input class	Very big	Very very low

**Table 2.** Centroid Value for Output Classes

	Very high	High	Medium	Low	Very low	Very very low
$c_i$	1	0.8	0.6	0.4	0.2	0

the situations that the output class will become VVL if either of the inputs is out of the range limits.

### Defuzzification

The defuzzification process converts the qualitative rule outcome into a numerical output. The centroid defuzzification method (also known as the center-of-area or gravity method) (Runkler 1996; Taylor and Meldrum 1998) is used to determine the MD in this research:

$$MD = \frac{\sum_{i=1}^n w_i c_i I_i}{\sum_{i=1}^n w_i I_i} \quad (3)$$

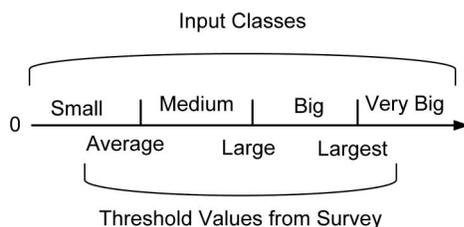
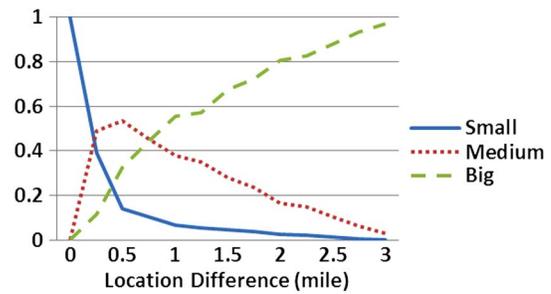
where  $w_i$  = rule weight representing the importance of the  $i^{th}$  rule;  $c_i$  = centroid of the output class  $i$ ; and  $I_i$  = implicated area of the output class  $i$ . The “implicated area” refers to the polygon area of the corresponding class (Taylor and Meldrum 1998). The centroid of each output class is shown in Table 2. If the output classes include VVL, the output MD is set to zero. The MD is calculated for all possible data pair. In this study, a data pair is regarded as a match if  $MD \geq 0.5$ . If multiple matches are found, then the match with the largest MD will be selected.

## Application

### Determination of Membership Function

Before applying the fuzzy logic-based mapping algorithm, the membership functions need to be determined. To make the membership functions objective, an expert survey was conducted to collect necessary information to set them up properly.

The survey was conducted from February 5 to March 3, 2009. The CR and CRpt data sets differ significantly and have different sources, so it is difficult to find people familiar with both data sets. Because the CRpt data are more precise in location and date and are more physically and directly tied to incident location, the CRpt data were chosen as a baseline for comparison with the application of fuzzy logic to the CR data. Therefore, survey subjects are WSDOT staff members who have been working on CR data collection for more than 3 years. The survey questionnaire contains eight questions, including four questions directly related to the determination of the fuzzy membership function. Questions included: “Based on your experience, how far away do you expect to find the carcass from the location where the actual collision took place?” (Q1); and “What is the greatest discrepancy in distance you would expect to find between the actual and reported locations for a carcass removal report?” (Q2). Similar questions about the date difference were also included.

**Fig. 2.** Determination of fuzzy classes**Fig. 3.** Membership function for location difference

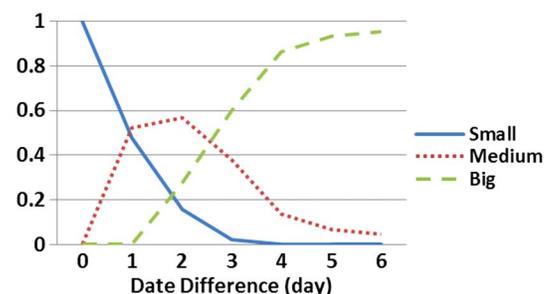
Forty-eight of the 54 received responses were considered valid. The six discarded surveys were incomplete in critical areas. From each expert’s inputs, it was possible to understand how these experts judge the date and location differences and the threshold values to be used. Fig. 2 shows the fuzzification process of an expert. Take the previously mentioned Q1 and Q2, for example. For each survey, the expert determines values of “Average” and “Large” on the basis of the answers to Q1 and Q2, respectively. Next, the value of “Largest” is defined as the largest value among all the “Large” values from all survey responses. If a location difference is smaller than the expert’s expected location difference (the value of “Average”), then the current data pair’s location difference is “Small,” in the expert’s opinion; if the location difference is smaller than the expert’s large difference (the value of “Large”) and larger than the expert’s expected location difference (the value of “Average”), then the current data pair’s location difference is “Medium,” in the expert’s opinion. Following the same rules, the other two input classes, B and VB, can be determined. Hence, the location difference of the same data pair may be categorized as a different input class from different experts’ views. These measured differences in experts’ judgments offer a solid foundation to build up the membership functions.

The degree of membership of input value  $u$  (milepost difference or date difference) in fuzzy class  $A_i$  ( $i = 1, 2, 3$  representing classes S, M, B, respectively) can be calculated by using the membership function for class  $A_i$ . The membership function is constructed, as shown in Eq. (4), by using the survey inputs from WSDOT experts:

$$f_i(u) = \frac{n_{i,u}}{K} \quad (4)$$

where  $n_{i,u}$  = number of observations of  $u \in A_i$  for class  $i$ ; and  $K$  = total number of observations (valid responses received from the survey) for all classes ( $K = 48$  in this study).

The results for the constructed membership functions of the survey are shown in Figs. 3–5. Fig. 3 shows the membership function for the location difference between the AVC and CR data sets. For example, approximately 56% of the staff regarded 1.61 km

**Fig. 4.** Membership function for time difference on weekdays

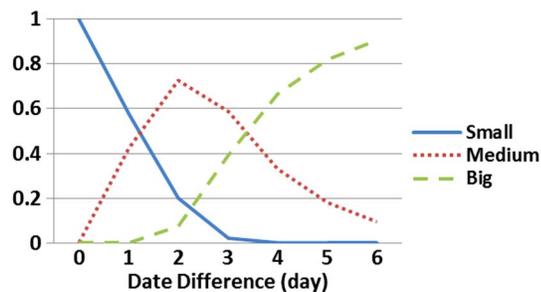


Fig. 5. Membership function for time difference on weekends

(1 mi) as a difference of B, whereas 38% of staff thought that it was a difference of M, and approximately 6% of staff regarded it as a difference of S.

Figs. 4 and 5 show the membership function for the date difference on weekdays and weekends, respectively. When an AVC happens during a weekend, the carcass is often collected on the following Monday or Tuesday. The date difference on weekends is slightly larger than on weekdays. For example, approximately 60% of staff considered 3 days a difference of B for weekdays, but fewer staff (38%) considered the same period of time as a difference of B for weekends.

### Mapping Results

The fuzzy logic-based mapping algorithm was used to combine the 5-year (2002–2006) CRpt data and CR data for the 10 SRs mentioned in the Research Data section.

As shown in Table 3, the fuzzy logic-based mapping algorithm identified a matching percentage between 25~35% for each year.

This new data set of matched records has variables combined from both data sets. The union of the two data sets can improve the data completeness. Compared with the original CR data set, the new union data set has approximately 15~22% more records, as shown in the improved percentage column of Table 3. Moreover, the intersection of the two data set can improve the richness of the data set because the combined data set will have more attributes (or variables) for each pair of matched AVC records. The nonmatched data in the CR data could be identified as the unreported AVCs in the CRpt data. With this more complete collision data set, many modeling analyses can be conducted. Examples can be found in Wang et al. (2010), who identify the contributing factors of AVCs using the combined data sets.

### Algorithm Verification

After the proposed algorithm has been implemented, a major step is to verify whether the algorithm is able to reasonably imitate the experts' decision process and produce a combined quality data set. However, because no ground-truth AVC data are available, it is nearly impossible to validate the performance of the algorithm by using the existing data sets. Therefore, another expert survey was also conducted from March 5–23, 2009 for verification purposes. Again, the survey participants are WSDOT employees who had collected CR data for more than 3 years. Each survey subject was asked to judge whether the data pairs listed on the questionnaire match. The disparity between the experts' results and the algorithm results can be a measure for the credibility of the proposed algorithm.

Table 3. Data Mapping Results for Study Routes in Five Years (2002–2006)

Year	Total number of records		Matched data pairs	Matching percentage (%)	Union data sets	Improved percentage (%)
	Reported animal-vehicle collisions	Carcass removal				
2002	529	1,876	152	28.7	2,253	20.0
2003	508	1,771	151	29.7	2,128	20.2
2004	529	1,702	139	26.3	2,092	22.9
2005	544	2,290	186	34.2	2,648	15.6
2006	533	1,944	144	27.0	2,333	20.0

Table 4. Survey and Algorithm Matching Percentage for Different Data Pairs

Number	Route	Reported Animal-vehicle collision data				Carcass removal data				Matching degree (%)		$e_i^a$
		Milepost	Weekday	Month	Day	Milepost	Weekday	Month	Day	Survey	Algorithm	
1	2	302.1	Thursday	October	20	302	Thursday	October	20	100	96	0.04
2	2	327.2	Wednesday	May	25	325	Monday	June	20	8	25	0.17
3	12	118.14	Monday	February	14	118	Tuesday	February	15	88	86	0.02
4	20	24.77	Wednesday	October	26	24.1	Wednesday	October	26	58	74	0.16
5	20	8.1	Thursday	November	10	5.5	Friday	November	18	0	24	0.24
6	90	257.27	Sunday	September	25	257	Thursday	September	29	69	51	0.18
7	90	55.2	Sunday	July	31	56	Monday	August	1	88	64	0.24
8	90	32.88	Thursday	March	31	34	Saturday	April	2	50	52	0.02
9	97	25.5	Wednesday	July	20	24	Monday	July	25	46	31	0.15
10	97	299.02	Sunday	September	10	299.7	Monday	October	3	35	35	0
11	195	84.53	Monday	November	14	83	Thursday	November	17	54	40	0.14
12	395	231.44	Friday	April	29	233.8	Thursday	May	12	12	24	0.12
13	970	2.21	Tuesday	November	22	2	Wednesday	November	23	96	82	0.14

<sup>a</sup>Absolute percentage error between the matching results.

A total of 13 data pairs included in the survey questionnaire were extracted from the AVC data set and the CR data set. These data pairs are considered representative of both the day and location differences between the two data sets. As shown in Table 4, information about SR milepost, weekday, month, and day from the data pairs was also provided on the survey questionnaire. Experienced WSDOT staff members were invited to fill out the questionnaire. They were asked to determine whether the data pairs match or not. The MD for each of the 13 listed data pairs was computed on the basis of expert inputs. The computational results are then compared with the fuzzy logic-based mapping algorithm outputs. The last three columns of Table 4 show the MDs from both the survey results and the fuzzy logic-based mapping algorithm, in addition to the percentages of the errors between survey and the results of the proposed algorithm. In the Matching Degree column, the bold values indicate that the data pair should refer to the same collision; the remaining cells indicate that the data pair does not match. (In this study, the MD of a data pair should be 50% or more to be marked as a match).

Table 4 shows that the survey and algorithm results agree in all cases except data pair 11, which experts concluded was a match, but the algorithm rejected. If the survey results are assumed accurate, then the accuracy rate (AR) for the proposed algorithm is

$$AR = \frac{N_{\text{accurate}}}{N_{\text{total}}} = \frac{12}{13} = 92.3\% \quad (5)$$

where  $N_{\text{accurate}}$  = number of data pairs correctly matched by the algorithm; and  $N_{\text{total}}$  = total number of the data pairs evaluated. The matching rate of 92.3% is considered to be a very encouraging result, given the complexity of this issue.

Mean absolute error (MAE), a quantity used to measure how close forecasts or predictions are to the eventual outcomes (DeGroot 1986), was used as the error indicator. The MAE of the proposed algorithm can be calculated by using Eq. (6):

$$MAE = \frac{1}{n} \sum_{i=1}^n |(f_i - y_i)| = \frac{1}{n} \sum_{i=1}^n |e_i| = 12\% \quad (6)$$

where  $f_i$  = result estimated by the fuzzy logic-based data mapping algorithm;  $y_i$  = ground-truth MD values calculated from the survey results; and  $e_i$  = MAE between the algorithm result and the survey result. The calculated error for each surveyed data pair is shown in the last column of Table 4.

## Summary

This paper presents a fuzzy logic-based data mapping algorithm that aims to improve AVC data by combining two types of data commonly used in AVC analysis: CRpt data and CR data. Two data sets collected from 10 SRs in Washington were used in this study.

The membership functions used in the fuzzy logic-based mapping algorithm were formulated on the basis of survey responses from WSDOT experts who have been working in AVC-related work for years. Unlike predefined deterministic membership functions, the modified membership functions can truly make a decision similar to the decision made by experts.

Using the proposed mapping algorithm, the CR and CRpt data sets can be combined to produce a more complete set of data. Through the use of this mapping algorithm, intersections of the two data sets can also be identified. Records in the intersection of the two data sets contain more variables on the same accidents and can be used to support more detailed analysis of AVCs. Approximately 25~35% of the CRpt data can be matched to the CR

data. The union of the two data sets can significantly increase the number of samples for AVC studies and hence expand the breadth of data. Compared with the original CR data set, the new union data set increases the number of records by 15~22%. In contrast, if compared with the original CRpt data set, the new union data set increases the number of records by 300~390%.

The proposed algorithm was verified by expert judgment data on the surveyed AVC data pairs collected through a survey. The verification results showed that the accuracy of the proposed algorithm is approximately 90% for the limited pairs of data included in the survey. The fuzzy mapping algorithm proved appropriate to increase the quality and quantity of the AVC data. The improved data set will benefit wildlife safety studies by providing more completed data sets. Because the design of the membership functions is adaptive in nature, the fuzzy logic-based mapping algorithm presented in this paper can also be easily transferred for applications in other areas. Although the algorithm was calibrated specifically for Washington, it can be easily extended to other states. In future applications, online surveys could be used to collect experts' opinions in a more cost-efficient manner.

## Acknowledgments

The authors would like to appreciate the funding supports from both WSDOT and the Transportation Northwest (TransNow), USDOT University Transportation Center for Federal Region 10. Special thanks also go to those WSDOT experts who helped with the two surveys described in paper. The authors are also grateful for data support from HSIS and WSDOT.

## References

- Allen, R. E., and McCullough, D. R. (1976). "Deer-car accidents in southern Michigan." *J. Wildl. Manage.*, 40(2), 317-325.
- Chen, G., and Pham, T. T. (2001). *Introduction to fuzzy sets, fuzzy logic, and fuzzy control systems*, CRC Press, Boca Raton, FL.
- Curtis, P. D., and Hedlund, J. H. (2005). "Reducing deer-vehicle crashes." *Wildlife damage management fact sheet series*, Cornell Cooperative Extension, Ithaca, NY.
- DeGroot, M. H., (1986). *Probability and statistics*, Addison-Wesley, New York, 209.
- Elvik, R., and Mysen, A. B. (1999). "Incomplete accident reporting: Meta-analysis of studies made in 13 countries." *Transportation Research Record 1665*, Transportation Research Board, Washington, DC, 133-140.
- Hauer, E., and Hakkert, A.S. (1989). "Extent and some implications of incomplete accident reporting." *Transportation Research Record 1185*, Transportation Research Board, Washington, DC, 1-10.
- Hubbard, M. W., Danielson, B. J., and Schmitz, R. A., (2000). "Factors influencing the location of deer-vehicle accidents in Iowa." *J. Wildl. Manage.*, 64(3), 707-712
- Huijser, M. P., and Bergers, P. J. M. (2000). "The effect of roads and traffic on hedgehog (*Erinaceus europaeus*) populations." *Biol. Conserv.*, 95(1), 111-116.
- Huijser, M. P., Fuller, J., Wagner M. E., Hardy A., and Clevenger, A. P. (2007). "Animal-vehicle collision data collection: A synthesis of highway practice." *National Cooperative Highway Research Board Program: Synthesis 370*, Transportation Research Board, Washington, DC.
- Johnson, S. W., and Walker, J. (1996). "The Crash Outcome Data Evaluation System (CODES)." *NHTSA Technical Rep. DOT HS 808 338*, National Highway Traffic Safety Administration, Washington, DC.
- Knapp, K. K., Lyon, C., Witte, A., and Kienert, C. (2007). "Crash or carcass data: A critical definition and evaluation choice." *Transportation Research Record 2019*, Transportation Research Board, Washington, DC, 189-196.

- Knapp, K. K., and Yi, X. (2004). *Deer-vehicle crash patterns and proposed warning sign installation guidelines*, Transportation Research Board, Washington, DC.
- Lao, Y., Wu, Y., Corey, J., and Wang, Y. (2011a). "Modeling animal-vehicle collisions using diagonal inflated bivariate poisson regression." *Accid. Anal. Prev.*, 43(1), 220–227.
- Lao, Y., Zhang, G., Wu, Y. and Wang Y. (2011b). "Modeling animal-vehicle collisions considering animal-vehicle interactions." *Accid. Anal. Prev.*, 43(6), 1991–1998.
- Li, H. X., and Yen, V. C. (1995). *Fuzzy sets and fuzzy decision-making*, CRC Press, Boca Raton, FL.
- Malo, J. E., Suarez, F., and Diez, A. (2004). "Can we mitigate animal-vehicle accidents using predictive models?" *J. Appl. Ecol.*, 41(4), 701–710.
- Mohammed, A., Noland, R. B. and Ochieng, W. Y. (2006). "A high accuracy fuzzy logic based map matching algorithm for road transport." *J. Intell. Transp. Syst.*, 10(3), 103–115.
- Naso, D., Scaleria, A., Aurisicchio, G., and Turchiano, B. (2006). "Removing spike noise from railway geometry measures with a fuzzy filter." *IEEE Trans. Syst. Man Cybern., Part C Appl. Rev.*, 36(4), 485–494.
- National Highway Traffic Safety Administration (NHTSA). (2010). "The Crash Outcome Data Evaluation System (CODES) and applications to improve traffic safety decision-making." *NHTSA Technical Rep. DOT HS 811 118*, NHTSA, Washington, DC.
- Proctor, M. F. (2003). "Genetic analysis of movement, dispersal, and population fragmentation of grizzly bears in southwestern Canada." Ph.D. dissertation, Univ. of Calgary, Calgary, Alberta, Canada.
- Rao, D. H., and Saraf, S. S. (1995). "Study of defuzzification methods of fuzzy logic controller for speed control of a DC motor." *Power Electronics, Drives and Energy Systems for Industrial Growth Meeting*, IEEE, New York, Vol. 2, 782–787.
- Reilley, R. E., and Green, H. E. (1974). "Deer mortality on a Michigan interstate highway." *J. Wildl. Manage.*, 38(1), 16–19.
- Romin, L. A., and Bissonette, J. A. (1996). "Deer-vehicle collisions: Status of state monitoring activities and mitigation efforts." *Wildl. Soc. Bull.*, 24(2), 276–283.
- Runkler, T. A. (1996). "Extended defuzzification methods and their properties." *Fifth IEEE Int. Conf. on Fuzzy Systems*, IEEE, New York, Vol. 1, 694–700.
- Seiler, A. (2005). "Predicting locations of moose-vehicle collisions in Sweden." *J. Appl. Ecol.*, 42(2), 371–382.
- Swain, N. K. (2006). "A survey of application of fuzzy logic in intelligent transportation systems (ITS) and rural ITS." *Proc., IEEE Southeastcon*, IEEE, New York, 85–89.
- Syed, S., and Cannon, M. E. (2004). "Fuzzy logic-based map matching algorithm for vehicle navigation system in urban canyons." *Proc., 2004 National Technical Meeting of the Institute of Navigation (ION)*, San Diego, CA, 982–993.
- Taylor, C., and Meldrum, D. (1998). "Fuzzy ramp metering: Design overview and simulation results." *Transportation Research Record 1634*, Transportation Research Board, Washington, DC, 10–18.
- U.S. Department of Transportation Federal Highway Administration. (2009). "Highway Safety Information System (HSIS)." (<http://www.hsisinfo.org/>) (Jun. 15, 2009).
- Van der Zee, F., Wiertz, J., Ter Braak, C., Apeldoorn, R., and Vink, J. (1992). "Landscape change as a possible cause of the badger *Meles meles* L. decline in The Netherlands." *Biol. Conserv.*, 61(1), 17–22.
- Wagner, P., and Carey, M. (2006). "Connecting habitats and improving safety." *WSDOT transportation research project problem statements*, Washington State Dept. of Transportation, Olympia, WA.
- Wang, Y., Lao, Y., Wu, Y., and Corey, J. (2010). "Identifying high risk locations of animal-vehicle collisions on Washington state highways." *WSDOT Research Rep. WA-RD 752.1*, Washington State Dept. of Transportation, Olympia, WA.
- Ye, F., and Lord, D. (2011). "Investigating the effects of underreporting of crash data on three commonly used traffic crash severity models: Multinomial logit, ordered probit and mixed logit models." *Transportation Research Record 2241*, Transportation Research Board, Washington, DC, 51–58.
- Zhao, Y. (1997). *Vehicle location and navigation system*, Artech House, Norwood, MA.