Fuzzy Logic-based Mapping Algorithm for Improving Animal-Vehicle Collision Data

Yunteng Lao[1], Yao-Jan Wu[2], Yinhai Wang[3], Kelly McAllister[4]

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 more than $1 billion in property damage each year in the U.S.. 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 datasets 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 datasets so that data are not over-counted when combining the two data sets. The membership functions of the fuzzy logic algorithm are determined by a survey of Washington State Department of Transportation carcass removal staff. As verified by expert judgment collected through another survey, the accuracy of this algorithm was approximately 90%. Applying this algorithm to Washington State data sets identified that about 25%~35% of

[1] 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)
[2] Assistant Professor, Department of Civil Engineering, Saint Louis University, 3450 Lindell Boulevard, McDonnell Douglas Hall Room 2051, St. Louis, MO 63103. E-mail: yaojan@slu.edu
[3] Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of Washington, P.O. Box 352700, Seattle, WA 98195-2700. E-mail:yinhai@u.washington.edu
[4] Habitat Connectivity Biologist, Washington Department of Transportation, 310 Maple Park SE, Olympia WA 98504-7331. E-mail: mcallke@wsdot.wa.gov

the CRpt data records have matching pairs in the CR data. Compared to the original CR dataset, the combined dataset has 15%~22% more records. The proposed algorithm provides an effective means for merging the CRpt data and the CR data. Such a combined dataset is more complete for wildlife safety studies and may provide additional insights into understanding the issue of AVCs.

Keywords: animal-vehicle collision, carcass removal data, fuzzy logic, data mapping, and 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 AVCs. In Washington State, approximately 3,000 collisions occur yearly with deer and elk on state highways (Wagner and Carey, 2006). Romin and Bissonettee (1996) estimated that at least 500,000 deer-vehicle collisions occurred nationwide in 1991. 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 1 billion U.S. dollars annually. Wildlife animals die immediately or shortly in most AVCs (Allen and McCullough, 1976). AVCs may also reduce the population level of some precious species (e.g., van der Zee et al. 1992; Huijser and Bergers 2000) or even cause a serious decrease in population survival probability (Proctor, 2003). All 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 datasets 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 completed dataset.

In most AVC studies, two types of AVC data are usually used: the Collision Reported (CRpt) data and the carcass removal (CR) data, as emphasized in a National Cooperative Highway Research Program (NCHRP) report by Western Transportation Institute (Huijser etc., 2007). In Washington State, the CRpt are collected by police officers or reported by citizens whereas the CR data are collected by the maintenance team of WSDOT. Since the two datasets are collected by different agencies using varying methods, data integration and interpretation have been a challenging issue. Therefore, most previous AVC studies used either the CRpt data or the CR data, treating the two datasets separately. For example, Hubbard et al., (2000), Malo et al., (2004) and Seiler (2005) conducted their AVC analyses based on the CRpt data, while Reilley and Green (1974), Allen and McCullough (1976), Knapp and Yi (2004) and Lao et al. (2011a) employed the CR data. Only a few studies considered this two kinds of data together (Lao et al., 2011a).

Based on the findings of a survey conducted by this study, carcass removal professionals at the Washington State Department of Transportation (WSDOT) basically agree that over 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 based solely on the CRpt data or the CR data may result in biased results. Based on 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 the CRpt data and CR data together would be useful to provide a more complete dataset, and thus reduce the negative effect on the modeling analysis. Methods are needed to properly merge the two datasets.

Previous studies by Johnson and Walker (1996) and 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 datasets provide users with more information for each record, e.g. medical data combined with and financial outcome information. However, their method only focuses on increasing the number of attributes for each data record rather than increasing the number of record. Merging two incomplete datasets to increase the data size (number of records) and data quality (increasing number of attributes) are both critical and main objectives of this paper. To achieve these objectives, a fuzzy logic-based data mapping algorithm is developed to combine the CR data and the CRpt data.

The paper is organized as follows. The data overview and relevant issues will be addressed in the next section, followed by the 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 earlier, two types of AVC data are commonly used in AVC modeling and analysis. In this study, we will use the two datasets collected in Washington State to demonstrate the fuzzy logic-based data mapping algorithm. CRpt data can be extracted from the Washington State accident file provided by Highway Safety Information System (HSIS). HSIS is operated by the University of North Carolina Highway Safety Research Center and the LENDIS corporation under a contract with Federal Highway Administration (FHWA) (HSIS, 2009). The HSIS collision data of Washington were compiled from the State Trooper filed field reports and citizen reports. Note that the AVC records in the HSIS database have no detailed animal type information other than "domestic" or "non-domestic". 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 dataset contains detailed information about animal species, such as mule deer, white-tail deer, and elk.

Ten State Routes (SRs) (US-2, SR-8, US-12, SR-20, I-90, US-97, US-101, US-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 total numbers of records in each data set over a five-year period (2002-2006) on each of the study routes. Note that the CRpt data only contains "non-domestic" animals whereas the CR data only contains "deer" and "elks" in Fig. 1. It is obvious that the CRpt and CR datasets are substantially different. The number of CR records is typically more than that of the CRpt data on each route except for US-101. The CRpt data may likely underestimate the frequency of these types of collisions.

Data Matching Issues

Since 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 a CRpt has a match in the CR dataset is to compare its similarities in occurrence time and location. Generally, the CRpt data is reported at the same day when AVC occurs; however, the carcasses are picked up by the WSDOT maintenance staff depending on when they find the carcass. 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 datasets rarely happens. The record of the same event typically looks different in time and/or location in each dataset. 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 might not be removed for several days or even longer. In essence, these are cases where 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 non-existent.

- WSDOT maintenance staff generally do 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, WSDOT maintenance staff do not patrol the highways every day because they have other priority duties. In this case, a carcass not affecting traffic movement significantly may not be reported or identified immediately and hence might not be picked up in a couple of days.
- In addition, human errors may be introduced to the two datasets when the records were 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 datasets are very likely to underestimate the actual number of AVCs to some extent. Combining the two datasets will make the research data more complete and hence provide a better information base for AVC studies. Specifically, combining these two datasets will extend the data breadth (increase samples).

Methodology

The same AVC captured by both datasets 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. Here, our fuzzy logic based mapping algorithm will be explained following these three steps.

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 “non-domestic” animal type reported in AVC data is matched with the three deer types and elk in CR data. After the animal types had 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 dataset. Note that the date recorded in the CRpt data is usually the same as the day of collision, whereas the date in the CR dataset 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:

$$\text{Date difference} = \text{Date in the CR dataset} - \text{Date in the AVC dataset} \tag{1}$$

Location difference is the milepost difference between the CRpt location and the location where the carcass was collected. The State Route numbers in a data pair are required to be identical before mileposts could be compared. Therefore, the location difference is defined as the absolute value between the milepost in the AVC dataset and the milepost in the CR dataset:

$$\text{Location difference} = |\text{Milepost in the AVC dataset} - \text{Milepost in the CR dataset}| \tag{2}$$

These inputs are then translated into four fuzzy classes based on the level of difference: small, medium, big, and very big (S, M, B, and VB). VB presents the situation in which the input is larger than a critical range. For example, if the location difference is only considered within 3 miles, a 5 mile 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 zero to one. Most research (Taylor et al., 1998; Nikunja, 2006; Naso et al., 2006) has assumed the membership function to be a triangle for simplification and has designed it based on 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 will be described in the algorithm 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 datasets 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 big and the date difference is big, then their matching degree is very low. Rules 10 and 11 deal with 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 (a.k.a. Center-of-Area or gravity methods) (Runkler, 1996; Taylor and Meldrum, 1998) is used to determine the matching degree (*MD*) in this research:

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

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

Application

Determination of Membership Function

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

Forty-eight out of the 54 received responses were considered valid. The six discarded surveys were incomplete in critical areas. From each expert’s inputs, we were able to understand how these experts judge the date and location differences and the threshold values to be used. Fig. 2 illustrates the fuzzification process of an expert. Take above-mentioned Q1 and Q2 for example. For each survey, each expert determines values of “Average” and “Large” based on 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 this 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 this expert’s opinion. Following the same rules, the other two input classes, Big and Very Big can be determined. Hence, the location difference of the same data pair may be categorized as a different input class from different expert’s view. 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 the classes of S, M, B respectively) can be calculated by using the membership function for class A_i. The membership function is constructed as shown in Equation (4) by using the survey inputs from the WSDOT experts.

$$f_i(u) = n_{i,u} / K \tag{4}$$

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

The results for the constructed membership functions of the survey are shown in Fig. 3 to Fig. 5. Fig. 3 shows the membership function for location difference between the AVC and CR datasets. For example, approximately 56% of the staff regarded one mile as a big difference while 38% of staff thought that it was a Medium difference and about 6% of staff regarded it as a Small difference.

Fig. 4 and Fig. 5 show the membership function for 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 three days a Big difference for weekdays but fewer staff (38%) considered the same period of time as a Big difference for weekends.

Mapping Results

The fuzzy logic based mapping algorithm was used to combine the five-year (2002-2006) CRpt data and CR data for the ten study routes 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 dataset of matched records has variables combined from both datasets. The union of the two datasets can improve the data completeness. Compared to the original CR dataset, the new union dataset has about 15%~22% more records, as shown in the Improved Percentage column. Moreover, the intersection of the two datasets can improve the richness of the dataset because the combined dataset will have more attributes (or variables) for each pair of matched AVC record. The non-matched data in the CR data could be identified as the unreported AVCs in the CRpt data. With this more complete collision dataset, many modeling analyses can be conducted. Examples can be found in (Wang et al, 2010) where the authors identify the contributing factors of AVCs using the combined datasets.

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 dataset. However, because no ground-truth AVC data is available, it is nearly impossible to validate the performance of the algorithm by using the existing datasets. Therefore, another expert survey was also conducted from Mar. 5 to Mar 23, 2009 for verification purposes. Again, the survey participants are WSDOT employees who had collected CR data for more than three 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.

A total of 13 data pairs included in the survey questionnaire were extracted from the AVC dataset and the CR dataset. These data pairs are considered representative of both the day and location differences between the two datasets. As shown in Table 4, information about State Route, Milepost, Weekday, Month, and Day from the data pairs was also provided on the survey questionnaire. Experienced WSDOT staff were invited to fill out the questionnaire. They were asked to determine whether the data pairs match or not. Matching degree for each of the 13 listed data pairs was computed based on 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 matching degrees from both the survey results and the fuzzy logic based mapping algorithm, as well as the percentage of the errors between survey and the results of the proposed algorithm. In the Matching Degree column, the gray cells indicate that the data pair should refer to the same collision; the clear cells indicate that the data pair does not match (In this study, the matching degree of a data pair should be 50% or higher to be marked as a match.).

The table shows that the survey and algorithm results agree in all cases except data pair No. 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:

$$AR = N_{accurate} / N_{total} = 12/13 = 92.3\% \tag{5}$$

where $N_{accurate}$ is the number of data pairs correctly matched by the algorithm; N_{total} is the 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 (Morris, 1986), were used as the error indicator. The MAE of the proposed algorithm can be calculated by using Equation (6) :

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

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

Summary

This paper presented a fuzzy logic-based data mapping algorithm that aims to improve animal-vehicle collision (AVC) data by combining two types of data commonly used in AVC analysis: the CRpt data and carcass removal data. Two datasets collected from ten study routes in Washington State were used in this study.

The membership functions used in the fuzzy logic based mapping algorithm were formulated based on the 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 the decision similar to the decision made by experts.

Using the proposed mapping algorithm, the carcass removal and the CRpt datasets can be combined to produce a more complete set of data. Through the use of this mapping algorithm, intersections of the two datasets can be identified as well. Records in the intersection of the two datasets contain more variables on the same accidents and can be used to support more detailed analysis of AVCs. About 25%~35% of the CRpt data can be matched to the CR data. The union of the two datasets can significantly increase the number of samples for AVC studies and hence expand breadth of data. Compared to the original CR dataset, the new union dataset increases the number of records by

15%~22%. In contrast, if compared to the original CRpt dataset, the new union dataset increases the number of records by 300%~390% compared to the original CRpt data.

The proposed algorithm was verified by the 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 dataset will benefit wildlife safety studies by providing more completed datasets. Since the design of the membership functions is adaptive in nature, the fuzzy logic based mapping algorithm introduced in this paper can also be easily transferred for applications in other areas. Although the algorithm was calibrated specifically for Washington State, it can be easily extended to other states. In future applications, online survey could be used to collect expert's opinions in a more cost efficient manner.

Acknowledgments

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

References

Allen, R.E. and McCullough D.R., (1976). Deer-Car Accidents in Southern Michigan. Journal of Wildlife Management, Volume 40, pp. 317–325.

Chen, G. and Pham, T. T., (2001). Introduction to Fuzzy Sets, Fuzzy Logic, and Fuzzy Control Systems. CRC Press, pp.147-150.

Curtis, P.D. and Hedlund J.H., (2005). Reducing Deer-Vehicle Crashes. Wildlife Damage Management Fact Sheet Series. Cornell Cooperative Extension, Ithaca, N.Y.

Elvik, R. and Mysen A.B.. (1999). Incomplete Accident Reporting—Meta-analysis of Studies Made in 13 Countries. Transportation Research Record, No.1665, Transportation Research Board, National Research Council, Washington, D.C., pp.133-140.

Hauer, E. and Hakkert A.S., (1989). Extent and Some Implications of Incomplete Accident Reporting. Transportation Research Record, No.1185, Transportation Research Board, National Research Council, Washington, D.C., pp.1-10.

HSIS., (2009). http://www.hsisinfo.org/. Accessed June 15, 2009

Hubbard, M.W. B.J., (2000). Danielson and R.A. Schmitz. Factors influencing the location of deer-vehicle accidents in Iowa. Journal of Wildlife Management, Volume 64, pp. 707–712

Huijser, M.P., Fuller J., and Wagner M.E., A. Hardy, and A.P. Clevenger., (2007). Animal-vehicle Collision Data Collection: A Synthesis of Highway Practice. National Cooperative Highway Research Board Program: Synthesis 370. Transportation Research Board, Washington, DC.

Huijser, M.P. and P.J.M. Bergers, (2000). The Effect of Roads and Traffic on Hedgehog (Erinaceus europaeus) Populations. Biological Conservation, Vol. 95, pp. 111–116.

Knapp K.K. and Yi X., (2004). Deer-Vehicle Crash Patterns and Proposed Warning Sign Installation Guidelines. Transportation Research Board, Washington, DC.

Knapp, K.K., Lyon C., Witte A., and Kienert C.., (2007). Crash or Carcass Data: A Critical Definition and Evaluation Choice. In Transportation Research Record: Journal of the Transportation Research Board, No. 2019, Transportation Research Board of the National Academies, Washington, D.C., pp. 189-196.

Johnson S. W. and Walker J., (1996). The crash outcome data evaluation system (CODES) NHTSA Technical Report. DOT HS 808 338.

Lao Y., Zhang G., Wu Y., Wang Y., (2011a). Modeling animal–vehicle collisions considering animal–vehicle interactions. Accid. Anal. Prev. Accepted for publication. doi:10.1016/j.aap.2011.05.017

Lao Y., Wu Y., Corey J., Wang Y., (2011b). Modeling animal-vehicle collisions using diagonal inflated bivariate Poisson regression. Accid. Anal. Prev., 43(1), 220-227.

Li H.X., Yen V.C., (1995). Fuzzy sets and fuzzy decision-making, Boca Raton: CRC Press, pp. 93-99.

Malo, J.E., Suarez F., and Diez A., (2004). Can we mitigate animal–vehicle accidents using predictive models?. Journal of Applied Ecology, Volume 41. pp. 701–710.

Mohammed A., Noland, Robert B. and Ochieng, Washington Y., (2006). A High Accuracy Fuzzy Logic Based Map Matching Algorithm for Road Transport. Journal of Intelligent Transportation Systems, 10(3), pp.103-115.

Morris H. DeGroot., (1986). Probability and statistics. Addison-Wesley Pub. Co., pp. 209.

Naso, D., Scalera, A., Aurisicchio, G., and Turchiano, B., (2006). Removing spike noise from railway geometry measures with a fuzzy filter. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions. Volume 36, Issue 4, pp. 485–494.

National Highway Traffic Safety Administration (NHTSA) (2010) The Crash Outcome Data Evaluation System (CODES) and Applications to Improve Traffic Safety Decision-Making. DOT HS 811 181. http://www-nrd.nhtsa.dot.gov/Pubs/811181.pdf.

Nikunja K. Swain, (2006). A Survey of Application of Fuzzy Logic in Intelligent Transportation Systems (ITS) and Rural ITS. Southeast Con, Proceedings of IEEE, pp. 85-89.

Proctor, M.F., (2003). Genetic Analysis of Movement, Dispersal, and Population Fragmentation of Grizzly Bears in Southwestern Canada, Ph.D. dissertation, The University of Calgary, Calgary, AB, Canada.

Rao, D. H., and Saraf S. S., (1995). Study of Defuzzification Methods of Fuzzy Logic Controller for Speed Control of a DC Motor. IEEE Transactions, pp. 782–787.

Reilley, R.E. and Green H.E., (1974). Deer Mortality on a Michigan Interstate Highway. Journal of Wildlife Management, Volume 38, pp. 16–19.

Romin, L. A., and Bissonette J. A., (1996). Deer–Vehicle Collisions: Status of State Monitoring Activities and Mitigation Efforts. Wildlife Society Bulletin, Volume. 24, No. 2, pp. 276–283.

Runkler, T. A.., (1996). Extended Defuzzification Methods and Their Properties. IEEE Transactions. Volume 1, pp. 694–700.

Seiler A., (2005). Predicting locations of moose–vehicle collisions in Sweden. J Appl Ecol 42: 371–382.

Syed, S., and Cannon, M. E., (2004). Fuzzy logic-based map matching algorithm for vehicle navigation system in urban canyons. Proceedings of the Institute of Navigation (ION) National Technical Meeting, San Diego, CA. pp. 26–28.

Taylor, C. and Meldrum D., (1998). Fuzzy Ramp Metering - Design Overview and Simulation Results. In Transportation Research Record: Journal of the Transportation Research Board, No. 1634, Transportation Research Board of the National Academies, Washington, D.C., pp. 10-18.

van der Zee, F.F., J. Wiertz, C.J.F. ter Braak, R.C. van Apeldoorn, and J. Vink, (1992). Landscape Change as a Possible Cause of the Badger Meles meles L. Decline in The Netherlands. Biological Conservation, Volume 61, pp. 17–22.

Wagner, P. and Carey M., (2007). Connecting Habitats and Improving Safety. WSDOT Transportation Research Project Problem Statements.

Wang Y., Lao Y., Wu Y., and Corey J.. (2010) Identifying High Risk Locations of Animal-Vehicle Collisions on Washington State Highways. Washington State Department of Transportation, Washington State Transportation Center (TRAC) and Transportation Northwest Regional Center X (TransNow). WSDOT Research Report. WA-RD 752.1.

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.(in press).*

Zhao, Y., (1997). Vehicle location and navigation system. MA: Artech House, Norwood, MA.

Figures

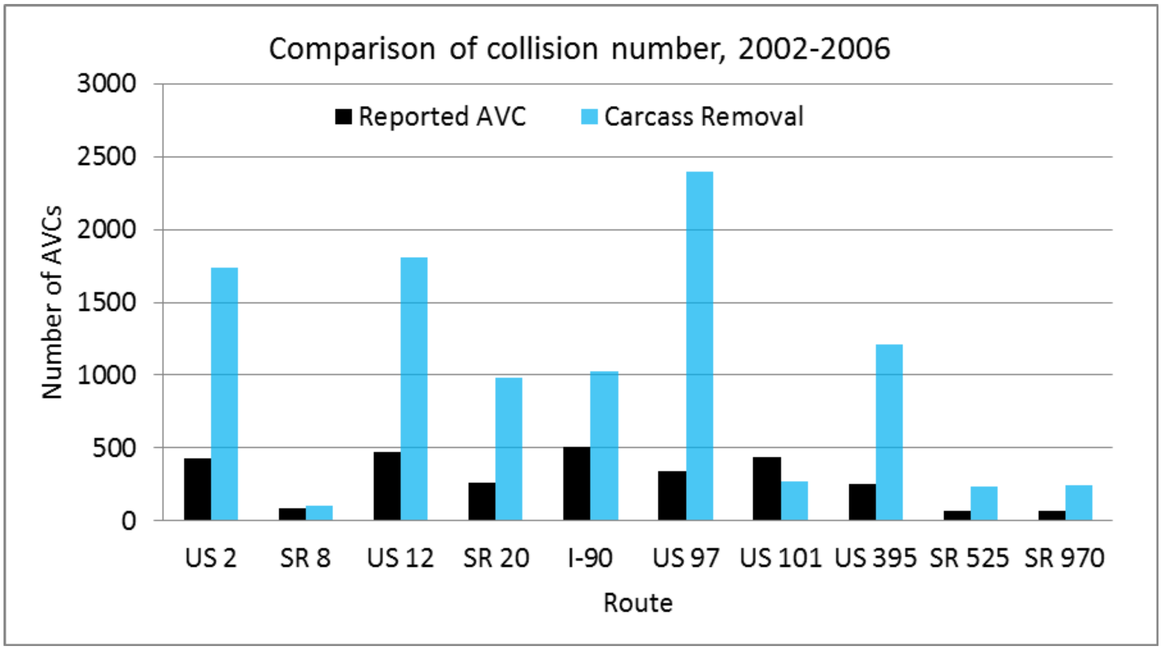

Fig. 1 Comparisons of total number of records bewteen two datasets for each study route during 2002-2006.

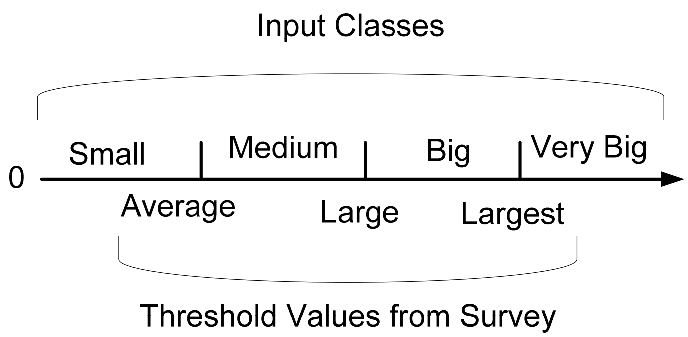

Fig. 2 Determination of fuzzy classes.

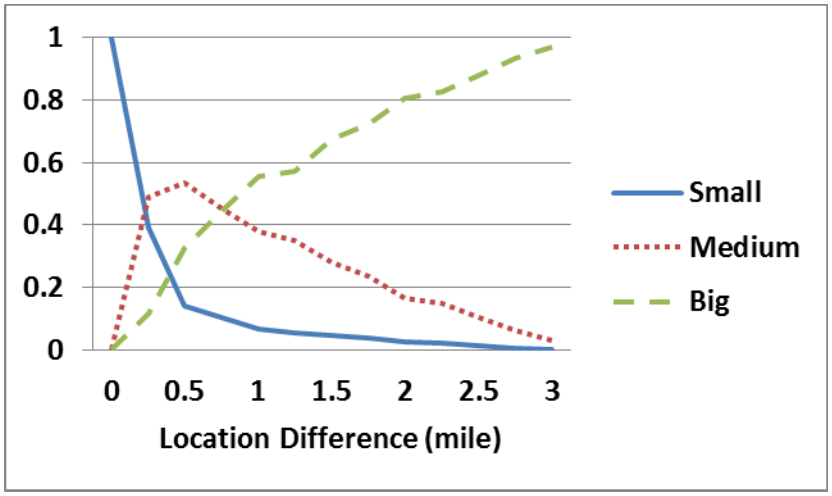

Fig. 3 Membership function for location difference.

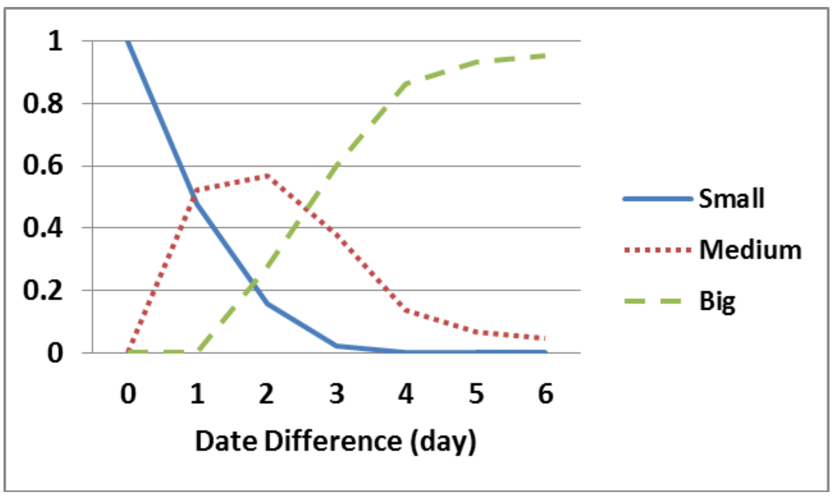

Fig. 4 Membership function for time difference on weekdays.

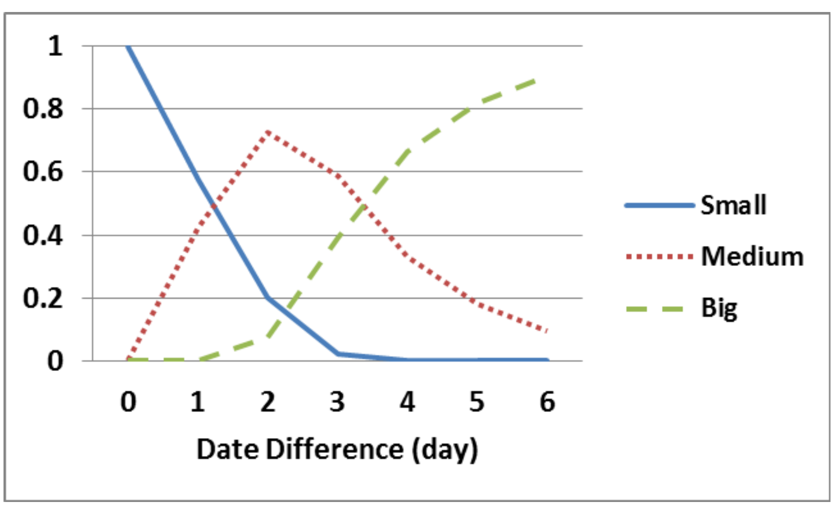

Fig. 5 Membership function for time difference on weekends.

Table 1. Rule Base for Fuzzy Mapping Algorithm

Rule	Default Rule Weight	Input Classes: Milepost difference	Input Classes: Date difference	Output Classes
1	1	S	S	VH
2	1	S	M	H
3	1	S	B	M
4	1	M	S	H
5	1	M	M	M
6	1	M	B	L
7	1	B	S	M
8	1	B	M	L
9	1	B	B	VL
10	1	VB	– *	VVL
11	1	–	VB	VVL

* “–” means any input classes

Table 2. Centroid Value for Output Classes

	VH	H	M	L	VL	VVL
c_i	1	0.8	0.6	0.4	0.2	0

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

Year	Total Number of Records: Reported AVC	Total Number of Records: Carcass Removal	Matched Data Pairs	Matching Percentage	Union Datasets	Improved Percentage
2002	529	1876	152	28.7%	2253	20.0%
2003	508	1771	151	29.7%	2128	20.2%
2004	529	1702	139	26.3%	2092	22.9%
2005	544	2290	186	34.2%	2648	15.6%
2006	533	1944	144	27.0%	2333	20.0%

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

No	Route	Reported AVC Data				Carcass Removal Data				Matching Degree (%)		e_i*
		Milepost	Weekday	Month	Day	Milepost	Weekday	Month	Day	Survey	Algorithm	
1	2	302.1	Thu	Oct.	20	302	Thu	Oct.	20	100	96	0.04
2	2	327.2	Wed	May	25	325	Mon	Jun.	20	8	25	0.17
3	12	118.14	Mon	Feb.	14	118	Tue	Feb.	15	88	86	0.02
4	20	24.77	Wed	Oct.	26	24.1	Wed	Oct.	26	58	74	0.16
5	20	8.1	Thu	Nov.	10	5.5	Fri	Nov.	18	0	24	0.24
6	90	257.27	Sun	Sep.	25	257	Thu	Sep.	29	69	51	0.18
7	90	55.2	Sun	Jul.	31	56	Mon	Aug.	1	88	64	0.24
8	90	32.88	Thu	Mar.	31	34	Sat	Apr.	2	50	52	0.02
9	97	25.5	Wed	Jul.	20	24	Mon	Jul.	25	46	31	0.15
10	97	299.02	Sun	Sep.	10	299.7	Mon	Oct.	3	35	35	0
11	195	84.53	Mon	Nov.	14	83	Thu	Nov.	17	54	40	0.14
12	395	231.44	Fri	Apr.	29	233.8	Thu	May	12	12	24	0.12
13	970	2.21	Tue	Nov.	22	2	Wed	Nov.	23	96	82	0.14

* e_i is the absolute percentage error between the matching results