Intensity assessment of pedestrian collisions in city of Mashhad based on fuzzy probabilities

Farzaneh Tatari and Mohammad-R. Akbarzadeh-T.

Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
Email: fa_tatari@yahoo.com, akbarzadeh@ieee.org

Abstract- *Fuzzy probabilities have been widely used in the areas of risk assessment and decision making. Here, we propose a system based on fuzzy probabilities for assessing the intensity of pedestrian collisions and levels of injury. The proposed approach uses possibility probability distributions for demonstrating fuzzy probabilities. A composition of possibility probability distributions (fuzzy probabilities) of collision intensities coming from different environmental variables yields the final fuzzy probability of intensity with higher reliability. The proposed approach is then applied to an actual database of pedestrian collisions from the City of Mashhad for over a period of five years.*
Keywords_ *Collision Intensity, Collision Variables, Fuzzy Probability, Possibility Probability Distribution.*

1 Introduction

Walking is a healthy and natural mean of transposition; it is therefore desirable that it also be safe. However, with the modern society and its rising crowd of transportation vehicles, this aspect of transposition has been sidelined and pedestrian collisions have caused significant social and financial losses every year. The safety of pedestrians during daily transposition is influenced by various variables. An appropriate paradigm for handling the effect of these variables can help insurance companies as well as city planners to better evaluate the safety of pedestrians in different situations and to plan the city environment for their improved safety. Unfortunately, recognizing all of the effective variables in a collision and their precise amount of effectiveness is non-trivial. The available information is highly imprecise and can lead to uncertain conclusions. With respect to the existing uncertainty in the collected information, the first source of data is known as “soft” data [17] that can be represented in linguistic form. This data is from expert opinions based on their background knowledge about the collision's intensity by considering related variables. Due to the linguistic form of this information and imprecision of data in this application, soft computing based methods, and in particular fuzzy-logic based analysis, offer a promising solution paradigm [18] to handle this type of uncertainty. In addition to soft data, the statistical historical data of pedestrian collisions are another valid data source which can be helpful in this case. The statistical data of pedestrian collisions may be incomplete when the number of observed data is not considerably large. Hence the statistical data are also accompanied by uncertainty which is caused by sparsity and insufficiency of data. We have found fuzzy probability framework as a suitable approach which enables us to enhance the reliability of our assessments by employing the both databases of soft data and statistical data at the same time. Fuzzy probabilities were first introduced by Zadeh (1984). Fuzzy probability theory [2] is a fuzzy approach to probability theory and is a generalized form of probability theory. In fuzzy probability, probability theory is complemented with an extra dimension of uncertainty provided by fuzzy set theory [16]. Generally we can divide the applications of fuzzy probabilities to two different main areas. The first is the area of reliability and risk assessment. In this area fuzzy probability has been widely applied in fuzzy fault trees [3,4,8] to assess the fault risk, reliability assessment for pressure piping [20], risk assessment of natural hazards [7] and reliability enhancement by combining expert opinions [14]. The second area is the area of the decision making. In this area, the fuzzy probabilities have been employed for decision making in perception-based theory [19], optimal decision fusion [11,12], inference by aggregation [9], information retrieval [6] and inventory control [1].

At any collision, there are two states for the pedestrian who has been hit by a car, in one state he would be just injured and in the other state he would die. Here, we employ fuzzy probabilities to handle the imprecision in expert opinions as the first dataset to assess the probability of injury and death for the pedestrians of Mashhad. We use fuzzy numbers to show these fuzzy probabilities. The second database is the statistical numeric data of pedestrian's collisions in Mashhad [10]. We combine these two insufficient databases to provide a more reliable result by the use of possibility probability distribution (PPD) which can model fuzzy probabilities in a suitable way.

2 Possibility probability distribution

The idea of representing assessments with PPD has been originally proposed by Haung [5]. This idea was improved by Karimi and Hullermeier [7] which was a new concept of fuzzy probability expressed by PPD. As we mentioned, on one hand, the statistical data in our type of application are typically insufficient, on the other hand background knowledge is often accompanied by imprecision. Therefore we apply the PPD to combine these two databases based on Bayesian approach. In the PPD approach, the prior knowledge (soft data) which is called prior distribution can be specified by a fuzzy number. $\bar{\theta}^{(i)}$ is an estimated probability by an expert for the state of death or injury in any collision where i is the variable that influences this probability and $p_i(\theta)$ is the prior distribution modeled in the form of symmetric, triangular fuzzy number, whose center is given by $\bar{\theta}^{(i)}$ (θ is the variable representing any probability in unit interval of [0,1]). As stated in [7] the support of the fuzzy number is of the form

ISBN: 978-989-95079-6-8

$$[\overline{\theta}^{(i)} - \overline{\theta}^{(i)}(1-\overline{\theta}^{(i)}).s \ ,\overline{\theta}^{(i)} + \overline{\theta}^{(i)}(1-\overline{\theta}^{(i)}).s] \quad (1)$$

$$0 \le s \le \min(1/\overline{\theta}^{(i)}, 1/(1-\overline{\theta}^{(i)})) \quad (2)$$

the so-called uncertainty factor, is a constant that specifies the reliability of the point estimation. For combining prior knowledge with empirical data which is the available statistical numeric data, we first divide the unit interval of probabilities into m points, $\theta_j, j = 1,...,m$, and then calculate the posterior probability for each point θ_j as follows [7].

$$p'_i(\theta_j) = \binom{n}{n_i} \theta_j^{n_i}(1-\theta_j)^{n-n_i}.p_i(\theta_j) \approx$$

$$\frac{1}{c}(\gamma_j.q_j + (1-\gamma_j).q_{j+1}),$$

$$q_j = \theta_j^{n_i}(1-\theta_j)^{n-n_i} p_i(\theta_j) \quad (3)$$

where n_i is the number of observations that satisfies the features of the ith variable among n observations. $\gamma_j \in [0,1)$ is a unique value such that

$$\theta_j = \gamma_j \theta_j + (1-\gamma_j).q_{j+1} \quad (4)$$

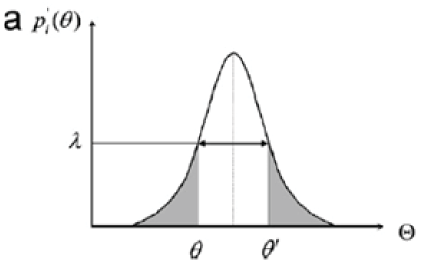
and c is a normalizing constant that guarantees

$$\int_0^1 p_i'(\theta) d\theta = 1 \quad (5)$$

Now is the time of transforming probability distribution to possibility distribution. Suppose a probability measure P on a set *X* is obtained via some statistical experiment. This probability function is a very rich piece of information, if the number of statistical experiments supporting it, is high enough. According to widely accepted consistence principle, the possibility measures should dominate P in the sense that $P(A) \le \pi(A)$ for all events $A \subseteq X$ and the maximally specific possibility distribution that exists must be unique [7]. As illustrated in [7] the most specific possibility distribution that approximates P from above (in the sense that $P \le \pi$) can be derived quite easily: for obtaining a possibility degree for any θ, $\pi_i(\theta)$, first $\lambda = p_i(\theta)$ is computed. Next, the second boundary point θ' that satisfies $p_i(\theta') = \lambda$ is found. Finally, the possibility degree $\pi_i(\theta)$ can be obtained as:

$$\pi_i(\theta) = \pi_i(\theta') = 1 - \left| \int_{\theta}^{\theta'} p'_i(x) dx \right|. \quad (6)$$

This possibility is equal to the area shaded grey in Fig. 1.a.



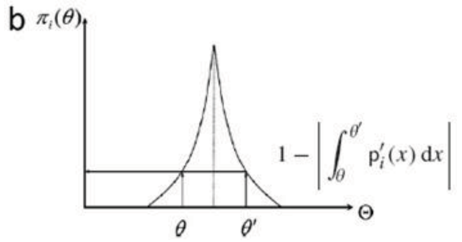


Figure 1: a. probability distribution to be transformed into a possibility distribution; b. transformed possibility distribution

3 Combining experts opinions and statistical data

We collected expert opinions (based on their own background knowledge and their reviews of traffic statics of death and injury in middle east countries [10]) about the percentage of collisions intensity (injury and death) by considering just one certain significant variable. The results of expert opinions are shown in Tables 1~5.

Table 1: Significant variable: hit organ

Injury probability if	Death probability if
Head and neck is hit, is approximately 0.09	Head and neck is hit, is approximately 0.55
Hand and arm is hit, is approximately 0.01	Hand and arm is hit, is approximately 0.001
Chest and stomach is hit, is approximately 0.2	Chest and stomach is hit, is approximately 0.2
Legs and pelvis is hit, is approximately 0.3	Legs and pelvis is hit, is approximately 0.03
More than one organ are hit, is approximately 0.5	More than one organ are hit, is approximately 0.3

Table 2: Significant variable: agility (crossing ability)

Injury probability for a pedestrian	Death probability for a pedestrian
Able to cross is 0.3	Able to cross is 0.4
Unable to cross is 0.6	Unable to cross is 0.6

Table 3: Significant variable: vehicle front shape

Injury probability if vehicle	Death probability if vehicle
Has bonnet is 0.7	Has bonnet is 0.4
Does not have bonnet is 0.2	Does not have bonnet is 0.2

Table 4: Significant variable: collision location

Injury probability if location is	Death probability if location is
In the city is 0.6	In the city is 0.3
Out of city is 0.04	Out of city is 0.05
In the suburb is 0.2	In the suburb is 0.15

Table 5: Significant variable: type of road

Injury probability if the road is a	Death probability if the road is a
Highway is 0.2	Highway is 0.6
Boulevard is 0.5	Boulevard is 0.6
Two_way road is 0.3	Two_way road is 0.2
incidental way is 0.1	incidental way is 0.05

With respect to the resulting experts ideas we obtained the fuzzy numbers for the proposed probability based on the approach illustrated in Section 2 where s is chosen as

$$s = 1/2(\min(1/\overline{\theta}^{(i)}, 1/(1-\overline{\theta}^{(i)}))) \quad (7)$$

It should be mentioned that we normalized $P(\overline{\theta}^{(i)})$ of all

ISBN: 978-989-95079-6-8

the prior probability distributions by factor N, to convert the prior probability distributions to fuzzy numbers.

$$N = \begin{cases} 2/(\bar{\theta}^{(i)}) & \bar{\theta}^{(i)} \le 0.5 \\ 2/(1-\bar{\theta}^{(i)}) & \bar{\theta}^{(i)} \ge 0.5 \end{cases} \quad (8)$$

As an example, through Figs. 2~3 we have shown the result of fuzzy probabilities of injury and death based on hit organ variable.

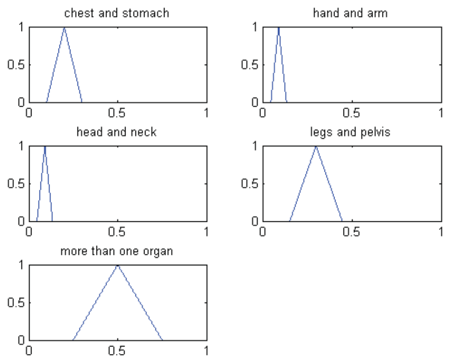


Figure 2: Injury fuzzy probability considering hit organ

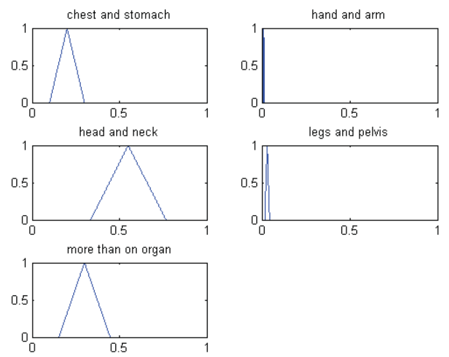


Figure 3: Death fuzzy probability considering hit organ

In the next step we apply (3) to obtain the posterior distribution based on statistical data which was gathered during one month in Shahidkamyab and Emamreza hospitals. In our database among 131 victims 99 of them had just injured and 32 of them had died. While injury and death are two independent states, we have considered the injuries which ends to fatality in the death statistical society. Therefore n=32 for dead pedestrians statistical society and n=99 for injured statistical society, these numbers are distributed through any variable partitions based on the collected data which is available in Tables 6 ~ 10. Among 18 factors which consist of human, vehicle, road and environmental factors, we could find 5 variables (hit organ, agility (crossing ability), vehicle front shape, location of collision and type of road) that had significant relations with collision intensity by the 5 percentage of error using SPSS.

Table 6: Distribution of statistical database on Significant variable of hit organ

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Head and neck = 10	Head and neck = 17
Hand and arm = 0	Hand and arm = 4
Chest and stomach = 5	Chest and stomach = 2
Legs and pelvis = 33	Legs and pelvis = 1
More than one point = 51	More than one point = 8

Table 7: Distribution of statistical database on Significant variable of agility (crossing ability)

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Able = 97	Able = 24
Unable =2	Unable = 8

Table 8: Distribution of statistical database on Significant variable of vehicle of front shape

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Has bonnet = 86	Has bonnet = 23
Does not have bonnet = 13	Does not have bonnet = 9

Table 9: Distribution of statistical database on Significant variable of location of the collision

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
In the city =60	In the city =14
Out of city =3	Out of city =6
Suburb =36	Suburb =12

Table 10: Distribution of statistical database on Significant variable vehicle of type of road

Total number of injured pedestrians = 99	Total number of died pedestrians = 32
Highway = 4	Highway = 7
Boulevard = 51	Boulevard = 10
Two way road = 34	Two way road = 11
Incidental way = 10	Incidental way = 4

The resulting posterior distributions of injury and death for the hit organ variable are as below:

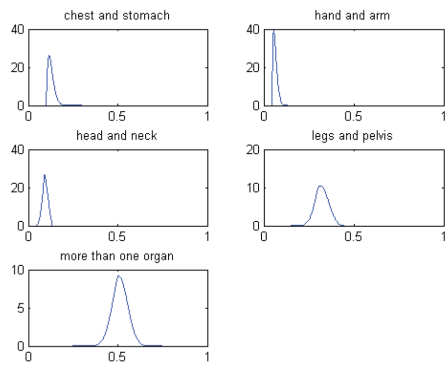


Figure 4: Posterior distribution of injury considering hit organ

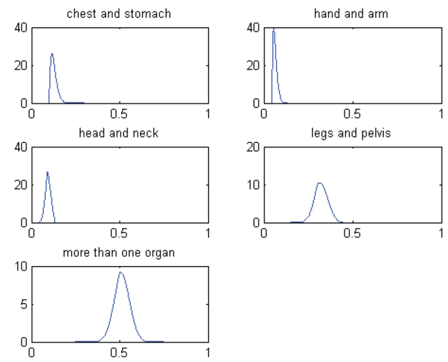


Figure 5: Posterior distribution of death considering hit organ

ISBN: 978-989-95079-6-8

Based on the procedure mentioned for transforming probability distribution to possibility distribution in Section 2, we obtained the possibility distributions of probability distributions. By comparing the prior fuzzy probabilities and possibility-probability distributions (Figs. 6~15), It can be easily interpreted that by combining the information of two databases less ambiguous, i.e. more certain and reliable results, are obtained in PPDs. It is also obvious that the PPD (fuzzy probability) can easily handle the existing uncertainty in the results.

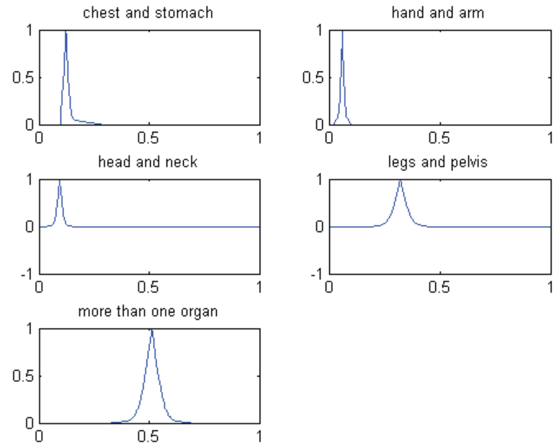


Figure 6: Possibility-probability distribution of injury considering hit organ

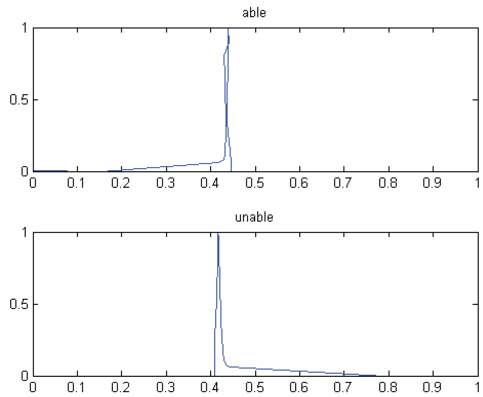


Figure 7: Possibility-probability distribution of injury knowing agility (crossing ability)

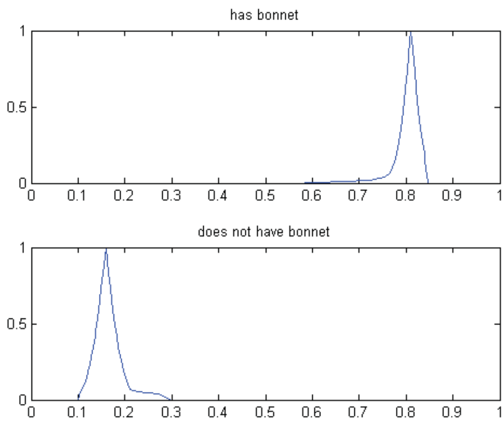


Figure 8: Possibility_probability distribution of injury considering vehicle front shape

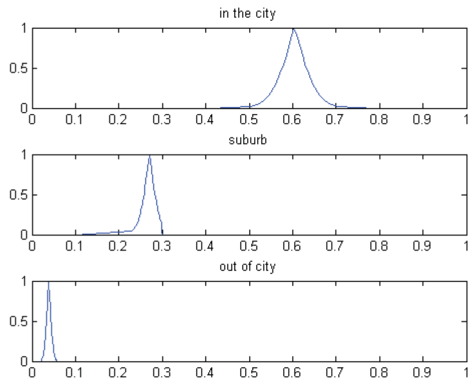


Figure 9: Possibility_probability distribution of injury considering location of collision

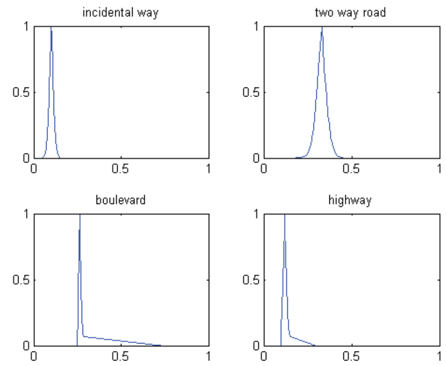


Figure 10: Possibility_probability distribution of injury considering type of road

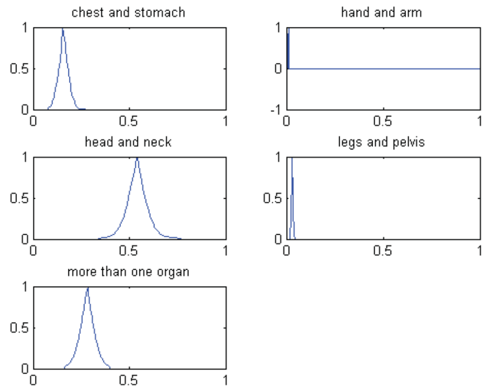


Figure 11: Possibility_probability distribution of death considering hit organ

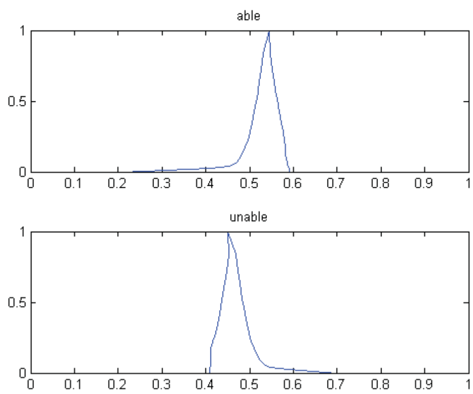


Figure 12: Possibility_probability distribution of death considering agility (crossing ability)

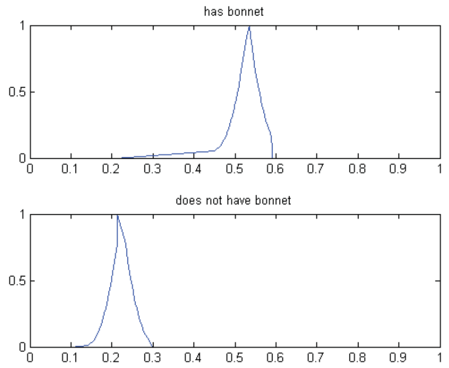


Figure 13: Possibility_probability distribution of death considering vehicle front shape

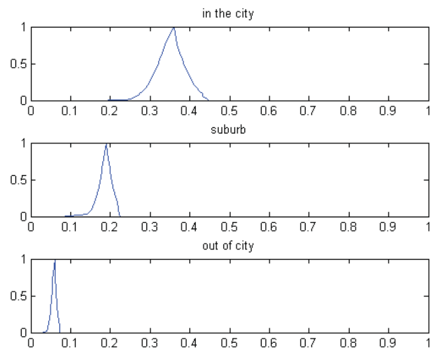


Figure 14: Possibility_probability distribution of death considering location of collision

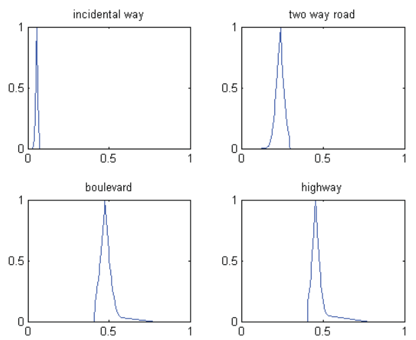


Figure 15: Possibility-probability distribution of death considering type of road

4 Collision intensity assessment

As it was explained before there are various variables which influence the intensity of collisions. If we assess the collision intensity based on one or few known variables the uncertainty of the assessment is so higher in comparison with the time that the assessment is based on more known variables. Therefore we can assess the fuzzy probability of injury or death of the pedestrians based on the combination of all significant variables PPDs. It should be mentioned that the employed method for combining fuzzy probabilities is based on fuzzy arithmetic [2] not Bayesian approach.

Assume that $\bar{A}$ and $\bar{B}$ are fuzzy probabilities of X. By having $\bar{A}$ and $\bar{B}$ which have been calculated based on different variables of X , we want to obtain $\bar{C}$ as a more reliable fuzzy probability of X by considering both variables, as the following

$$P[X(a)] \text{ is } \bar{A} \text{ and } P[X(b)] \text{ is } \bar{B} \equiv P[X(a,b)] \text{ is } \bar{C}.$$

Where a and b are related variables of X which effect assessing the fuzzy probability of X. One of the most common ways to aggregate fuzzy probabilities such as $\bar{A}$ and $\bar{B}$ is to obtain the α _cuts of $\bar{C}$ for any $\alpha \in [0,1]$ as below

$$\bar{A}[\alpha] = [a_1(\alpha), a_2(\alpha)], \bar{B}[\alpha] = [b_1(\alpha), b_2(\alpha)] \Rightarrow$$

$$\bar{C}(\alpha) = [\min(a_1(\alpha), b_1(\alpha)), \max(a_2(\alpha), b_2(\alpha))] \qquad (9)$$

We can extend this procedure to more than two fuzzy probabilities with different variables to obtain a more reliable fuzzy probability of X, $\bar{C}$. It is also notable to know that any α _cut in any fuzzy probability is a sub interval of the unit interval of probabilities. However the mentioned method may exhibit shortcomings, first the max and min operations are noninteractive so they tend to lead to the results that are getting close to 1 and 0 respectively. Second, the result of combination at any level of α is entirely dependent upon the extreme bounds. To overcome these shortcomings we propose to calculate the α _cuts of $\bar{C}$ based on fuzzy arithmetic [2].

$$\bar{A}[\alpha] = [a_1(\alpha), a_2(\alpha)], \bar{B}[\alpha] = [b_1(\alpha), b_2(\alpha)] \Rightarrow$$

$$\bar{C}(\alpha) = [a_1(\alpha) \times b_1(\alpha) \ , \ a_2(\alpha) \times b_2(\alpha)] \qquad (10)$$

The only drawback of fuzzy arithmetic method is that it results in low fuzzy probabilities which do not conform the reality (statistical data) but we can solve this problem by normalizing the obtained fuzzy probabilities of all the existing states. This normalization is in the way that for each $\alpha -$ cut , $\alpha \in [0,1]$, the summation of $\alpha -$ cut boundary points (probabilities) must be equal to one. In the specific application that we focus on it, we have two states of "X= death, Y= injury" in any collision, therefore we must normalize the final fuzzy probabilities in the way that

$$P'_{X1\alpha} = \frac{P_{X1\alpha}}{(P_{X1\alpha} + P_{Y2\alpha})}, \quad P'_{Y2\alpha} = \frac{P_{Y2\alpha}}{(P_{X1\alpha} + P_{Y2\alpha})},$$

$$P'_{X2\alpha} = \frac{P_{X2\alpha}}{(P_{X2\alpha} + P_{Y1\alpha})}, \quad P'_{Y1\alpha} = \frac{P_{Y1\alpha}}{(P_{X2\alpha} + P_{Y1\alpha})}$$

$$\text{then } P'_{X1\alpha} + P'_{Y2\alpha} = 1 \text{ and } P'_{X2\alpha} + P'_{Y1\alpha} = 1 \quad \forall \ \alpha \in [0,1] \qquad (11)$$

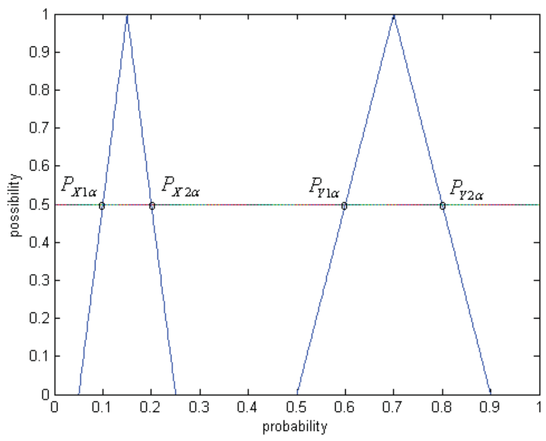


Figure 16: The probabilities which must be normalized at any α level

ISBN: 978-989-95079-6-8

We continue by an example where we used the fuzzy arithmetic procedure to obtain a more reliable fuzzy probability of injury and death based on the obtained results. In the following example for combining the fuzzy probabilities of injury and death which were assessed based on 5 significant variables, we use seven α _cuts (α = 0, 0.1, 0.3, 0.5, 0.7, 0.8, 1) of all fuzzy probabilities (PPDs) to do the combination.

Example: A pedestrian who does not have the agility (crossing ability) is crashed by a mini-bus while crossing a two-way road in city of Mashhad. His chest and stomach are hit by the vehicle. (a) What is the fuzzy probability of his injury? (b) What is the fuzzy probability of his death?

Hit organ: chest and stomach, agility (crossing ability): unable to cross, Vehicle front shape: does not have bonnet, Collision location: in the city, Type of road: two-way road. By combining the PPDs of the related variables we can obtain the final fuzzy probability of injury as below:

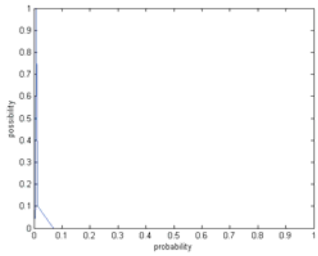


Figure 17: Fuzzy probability of injury before normalization

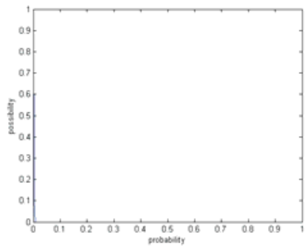


Figure 18: Fuzzy probability of death before normalization

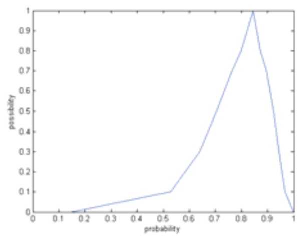


Figure 19: Normalized fuzzy probability of injury

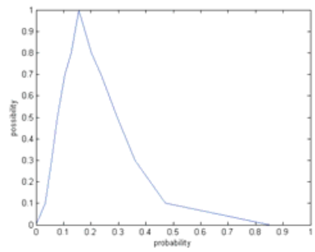


Figure 20: Normalized fuzzy probability of death

The normalized fuzzy probabilities are closer to statistical data and reality and the fuzzy probability of death and injury are complement of each other. The obtained result is more reliable while it has considered 5 significant variables at the same time instead of one.

5 Conclusion

Each year, too many pedestrians die all over the world with significant social and financial losses. The intensity of these collisions is of great importance for city planners and insurance companies alike. Here, we propose the use of fuzzy set theory to complement the probability theory with an additional dimension of uncertainty which leads to fuzzy probability characterized in terms of possibility probability distribution. We assessed the intensity of pedestrian collisions in Mashhad based on the variables that had significant relation with the intensity by use of fuzzy probability. The results indicate that by considering more significant variables in our assessment, we can decrease the amount of uncertainty and reach a more reliable result.

References

[1] J.J.Buckley et al., Fuzzy Mathematics in Economics and Engineering, Physica-Verlag, Heidelberg, Germany, 2002.

[2] J.J.Buckley, Fuzzy probabilities, new approach and applications, Springer, 2005.

[3] James P. Dunyak and Donald Wunsch, Fuzzy Probability for System Reliability, Proceedings of the 37th IEEE Conference on Decision & Control Tampa, 1998.

[4] James P. Dunyak et al., A Theory of Independent Fuzzy Probability for System Reliability, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 7, NO. 2, 1999.

[5] C. Huang, Fuzzy risk assessment of urban natural hazards, Fuzzy Sets and Systems 83 , 271–282, 1995.

[6] Z. Huang and T. D. Gedeon, Information retrieval estimation via fuzzy probability, World Automation Congress (WAC), 2006.

[7] I. Karimi and E. Hüllermeier, Risk assessment system of natural hazards: A new approach based on fuzzy probability, Fuzzy Sets and Systems 158, 987 – 999, 2007.

[8] R. Kenarangui, Event-Tree Analysis by Fuzzy Probability, IEEE TRANSACTIONS ON RELIABILITY, VOL. 40, NO. 1, 1991.

[9] A. L. Ralescu , D. A. Ralescu, Yoko Yamakata , Inference by aggregation of evidence with applications to fuzzy probabilities, Information Sciences 177, 378–387, 2007.

[10] M. Rezaei, estimation of collisions intensity for pedestrians in Mashhad using logistic regression, pp.92_107, 2004.

[11] G. Wang , S. Mao and Y. He, Optimal Decision Fusion With Fuzzy Priori Probabilities, Proceedings of IBP, 1998.

[12] G.Wang, S.Mao and Y. He, Distributed detection fusion with fuzzy priori probabilities and fuzzy cost function, IEEE INTERNATIONAL RADAR CONFERENCE, 2000.

[13] Ronald R. Yager, Decision Making with Fuzzy Probability Assessments, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 7, NO. 4, 1999.

[14] D. Yu and W. S. Park, Combination and evaluation of expert opinions characterized in terms of fuzzy probabilities, Annals of Nuclear Energy 27, 713-726, 2000.

[15] L. A. Zadeh, Fuzzy probabilities, Information processing and management 20(3), pp. 363_372, 198.

[16] L.A. Zadeh, Fuzzy sets , Information and control, 8(3), pp. 338_353, 1965.

[17] L.A. Zadeh, Posibility theory and soft data analysis, In: Cobb, L. and R. M. Thrall, eds., mathematical frontiers of the social and policy sciences. Westview press, Boulder, Colorado, pp.69_129, 1981.

[18] L.A. Zadeh, Soft computing and fuzzy logic, IEEE Softw are, 11 (6), pp.48_56, 1994.

[19] L. A. Zadeh, Toward a perception_based theory of probabilistic reasoning with imprecise probabilities, Journal of statistical planning and inference 105, 233-264, 2002.

[20] J. Zhou, Reliability assessment method for pressure piping containing circumferential defects based on fuzzy probability, International Journal of Pressure Vessels and Piping 82, 669–678, 2005.

ISBN: 978-989-95079-6-8