Australian Journal of Basic and Applied Sciences, 4(2): 314-320, 2010
ISSN 1991-8178

Ranking of the Factors Associated with Road Accidents using Correlation Analysis and Fuzzy TOPSIS

Lazim Abdullah and Nurnadiah Zamri

Department of Mathematics, Faculty Science and Technology, University Malaysia Terengganu

Abstract Road accident is one of the major causes of death and injuries in Malaysia. The increase of road accidents is said to be associated with the factors of rapid growth in population, economic development, and motorization. However no specific literature was found to specify weights and subsequently rank the factors. This paper proposes a ranking of three selected factors associated with road accidents using the correlation analysis and Multi Criteria Decision Making, fuzzy TOPSIS. Statistical accident data issued by Royal Malaysian Police and linguistic judgement data collected from three authorised personnel of three Malaysian Government agencies were considered in analysis. The ranks ane be drawn using the strength of correlation coefficients and the magnitude of closeness coefficients in fuzzy TOPSIS. The results from two analyses indicate that registered vehicles yielded the highest ranking followed by population and road length. This ranking gives rise to concerns about the relevance of the factors in reducing accidents rate.

Key words Road accidents, Pearson correlation, fuzzy sets, fuzzy TOPSIS, linguistic variables

INTRODUCTION

In the presence of motor vehicles on the road, accident has become a central issue for discussions. The increased of road accident has contributed to an increase of people deaths in every year. According to United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP), in 2006, at least an estimated 440,000 persons were killed and more than two millions were injured in accidents on the roads of the countries or areas in the ESCAP region. The road accidents are a growing worldwide problem. Around 1 million deaths and over 23 million injuries per year, and around 85% of these deaths are occurred in developing countries. The problem is particularly urgent in developing countries as the Asia. Malaysia has no exception in facing the death caused by road accidents. Accidents become among ten of the major factors of deaths in Malaysia. The United Nations has ranked Malaysia 30th among countries with the highest number of fatal road accidents, registering an average of 4.5 deaths per 10,000 registered vehicles (Bernama, 2006). It has been reported by the Royal Malaysian Police (2008) that total number of road accidents had increased from 24,581 cases in 1974 to 363,319 cases in 2007. The number of fatalities (death within 30 days after accident) also increased but at slower rate compared to total road accident from 2,303 in 1974 to 6,282 in 2007. The figure suggests that traffic accidents in Malaysia have been increasing at the average rate of 9.7% per annum over the last three decades. Thus, road accidents have become a hot topic of discussions among public and definitely a concern for all countries.

There are many contributing factors and causes to road accidents. A comprehensive study of road safety (Treat *et al*., 1977) found that human error was the sole cause in 57% of all accidents and was a contributing factor in over 90%. In contrast, only 2.4% were due solely to mechanical fault and only 4.7% were caused only by environmental factors. Notwithstanding the human factors, the other factors such as road defects and vehicle defects are also occupied in discussing factors to road accidents. In a report by Highway Planning Unit, Road Safety Section, Malaysian Ministry of Works (2008), it was pointed out that road condition, population and number of registered vehicles have been associated with road accident. The increase of road accidents is said to be linked with the rapid growth in population, economic development, industrialisation and motorisation encountered by the country. Since 1970's, Malaysia had experienced a remarkable growth in these sectors. In fact, there is an increase in Malaysian population from 10.4 million in 1974 to 26.1 million in 2005 at an average growth rate of about 2.1% per year. Law *et al*. (2005) also supported the statement that growth in vehicle ownership was contributed to the increasing number of road accidents. Furthermore, the total length

Corresponding Author: Lazim Abdullah, Department of Mathematics Faculty, Science and Technology, University Malaysia Terengganu
E-mail: lazim_m@umt.edu.my,

of road had also increased from 11,161 km in 1974 to 71,814 km in 2005 to accommodate an increase in numbers of vehicles in Malaysia. This also led to an increase of ownership from 9.6 persons per vehicle in 1974 to 1.7 persons per vehicle in 2005. The total numbers of registered vehicles also increased from 1,090,279 to 15,026,660 vehicles in 2005. However, this report offers no explanation for the magnitude for each factor. Far from that the report also suggests that no single factor was contributed more from the others. This hypothetical statement would have been a far more convincing if the report had considered the intensity of each factors attributed to the accidents occurrence.

Numerous studies have attempted to explain and discuss road accidents with various approaches and methods. Hejar *et al*. (2005) used descriptive statistics in a cross sectional sampling study to determine the prevalence of road traffic injuries involving motorcar occupants and to find the association between the injured motorcar occupants with related safety factors among the upper six students in Selangor, Malaysia. Law *et. al*., (2005) made a projection of the vehicle ownership rate to the year 2010 and to use this projection to predict road accident deaths in year 2010. They used Gompertz growth model to project vehicle ownership and the prediction of road accident death rate using Autoregressive Integrated Moving Average (ARIMA) model with transfer noise function. Recently, Soma, *et al*. (2007) estimated value of mortality risk reductions in a traffic safety context. These estimates can be used both to calculate the benefits of specific traffic safety improvements and to compute the social cost of traffic crashes. Qirjako et.al, (2008) assessed the prevalence of fatal road traffic accidents in Tirana, Albania, and describe their determinants. This cross-sectional study included all road traffic accidents and used multivariable-adjusted binary logistic regression analysis to assess the predictors of fatal road traffic accidents.

Since accidents can be considered uncertain and vague, thus the fuzzy theory (Zadeh 1965) is germane. There were many researches used fuzzy theory approaches in elucidation of accidents. For example, Hassan and Mohamed (2002) applied fuzzy adaptive resonance theory MAP (fuzzy ARTMAP) neural networks to analyze and predict injury severity for drivers involved in traffic accidents. Fuzzy ARTMAP was used for analyzing driver injury severity for drivers involved in accidents on highways, signalized intersections, and toll plazas. However, all the previously mentioned methods had no discussion about ranking of factors associated with road accidents. Banking on the premises of the significance issue of accidents and the importance of factors associated with road accidents, this paper proposes two rankings of the factors associated with road accidents using two approaches. Specifically the objective of this paper is to establish rankings of the selected factors associated with road accidents using correlation analysis and fuzzy TOPSIS.

In the recent years, there are lots of papers discussed on fuzzy TOPSIS but too little attention has been paid to rank the causes of road accidents. For example, Dagdeviren *et al*. (2008) developed an evaluation model based on the analytic hierarchy process (AHP) and the technique for order performance by similarity to ideal solution (TOPSIS), to help the actors in defence industries for the selection of optimal weapon in a fuzzy environment where the vagueness and subjectivity are handled with linguistic values parameterized by triangular fuzzy numbers. The AHP is used to analyze the structure of the weapon selection problem and to determine weights of the criteria, and fuzzy TOPSIS method was used to obtain final ranking. Another example, Semih and Selin (2008) explored on selection of the appropriate solid waste site requires consideration of multiple alternative solutions and evaluation criteria because of system complexity. Evaluation procedures involve several objectives, and it is often necessary to compromise among possibly conflicting tangible and intangible factors. For these reasons, multiple criteria decision making (MCDM) has been found to be a useful approach to solve this kind of problem. Taken together, the ranking in the present paper can be seen as a bivariate analysis using secondary source of statistical data and also a multi criteria decision making (MCDM) problem using linguistic judgments of experts.

MATERIAL AND METHODS

Two models were employed to establish ranking. Correlation analysis uses statistical data released by the authority where as fuzzy TOPSIS uses linguistic data from expert opinions. Details of the methods are explained as follows.

Correlation Analysis:

Correlation analysis is used to measure the intensity of association observed between any pairs of variables (Glover and Mitchell, 2005). The two variables are usually known as dependent and independent variables. A widely used index of the association of two quantitative variables is the Pearson product-moment correlation coefficient. If two variables X and Y with data then Pearson product-moment correlation coefficient is

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{[\sum X^2 - \frac{(\sum X)^2}{n}][\sum Y^2 - \frac{(\sum Y)^2}{n}]}} \tag{2.1}$$

Value r lies between – 1 and +1 depending on the strength of relationship between X and Y and also directions of relationship.

Fuzzy TOPSIS:

MCDM has proven to be an effective approach for ranking by a finite number of alternatives characterized by multiple criteria. One of the techniques for order preference is called Technique for Order Performance by Similarity to Ideal Solution and abbreviated as TOPSIS. This technique based on fuzzy sets theory which has proven to be a powerful modelling tool for coping with subjectiveness and imprecision in human judgements. Modelling using fuzzy sets has proven to be an effective way for formulating decision problems where the information available is subjective and imprecise (Zimmermann, 1996). Many fuzzy TOPSIS methods have been proposed to handle linguistic decision making (Chen, 2000; Chu, 2002). Herrera and Herrera –Viedma (2000) sustained that linguistic terms are intuitively easier to use when decision makers express the subjectivity and imprecision of their assessment. General steps of fuzzy TOPSIS proposed by Olson (2004) are listed below.

Step1. Establish a decision matrix for ranking. A MCDM problem can be concisely expressed in matrix format as

$$C_1 \quad C_2 \quad \cdots \quad C_n$$

$$D = \begin{array}{c} A_1 \\ A_2 \\ \vdots \\ A_M \end{array} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{3.1}$$

where $A_1, A_2, \ldots, A_m$ are possible alternatives among which decision makers have to choose, $C_1, C_2, \ldots, C_n$ are criteria with which alternative performance are measured, x_{ij} is the rating of alternative A_i with respect to criterion C_j .

Step 2. Calculate the normalized decision matrix. The normalized value { r_{ij} } is calculated as

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{j=1}^{J} x_{ij}^2}} \quad , \; j = 1, \ldots, J; i = 1, \ldots, n. \tag{3.2}$$

Step 3. Calculate the weighted normalized decision matrix. The weighted normalized value v_{ij} is calculated as

$$v_{ij} = w_{ij} \times r_{ij} \quad , \; j = 1, \ldots, J; i = 1, \ldots, n, \tag{3.3}$$

where w_i is the weight if the ith criterion, and $\sum_{i=1}^{n} w_i = 1$.

Step 4. Determine the positive ideal solutions and negative ideal solutions respectively

$$A^* = \{ v_1^*, \ldots, v_n^* \} = \{ (\max_j v_{ij} \mid i \in I'), \ (\min_j v_{ij} \mid i \in I'') \},$$

$$A^- = \{ v_1^-, \ldots, v_n^- \} = \{ (\min_j v_{ij} \mid i \in I'), (\max_j v_{ij} \mid i \in I'') \}, \tag{3.4}$$

where I' is associated with the positive criteria, and I'' is associated with the negative criteria.

Step 5. Calculate the separation measures using the n-dimensional Euclidean distance. The separation of each alternative from the ideal solution is given as

$$D_j^* = \sqrt{\sum_{i=1}^{n} (v_{ij} - v_i^*)^2}, j = 1, \ldots, J. \tag{3.5}$$

Similarly, the separation from the negative-ideal solution is given as

$$D_j^- = \sqrt{\sum_{i=1}^{n} (v_{ij} - v_i^-)^2}, j = 1, \ldots, J. \tag{3.6}$$

Step 6. Calculate the relative closeness to the ideal solution. The relative closeness of the alternative a_i with respect to A^* is defined as

$$C_j^* = \frac{D_j^-}{D_j^* + D_j^-}, j = 1, \ldots, J. \tag{3.7}$$

Step 7. Rank the preference order. A large value of closeness coefficient C_j^* indicates a good performance of the alternative A_i. The best alternative is the one with the greatest relative closeness to the ideal solution.

Research Framework:

Framework of this research considers two models structure. The first model was correlation analysis and the second was fuzzy TOPSIS. Correlation model was a straightforward analysis in which two variables were analyzed. The strength of relationships between first variables i.e. population (A_1), registered vehicles (A_2) and road length (A_3) and second variable road accidents occurrence (RA) were determined. An official data issued by Royal Malaysian Police (2008) was used to obtain coefficients correlation. Data that relevant in the study were statistics of road accidents, vehicle ownership, road length and vehicles registered for the respective years for the Year 1974 to 2005.

For fuzzy TOPSIS, data for criteria and alternatives must be identified as part of a MCDM problem. The four criteria were Motorcycle (C_1), Car (C_2), Bus (C_3) and Lorry (C_4). These four criteria were selected based on the number of vehicles involved in accidents. Furthermore, a committee of three decision-makers or experts, D_1, D_2 and D_3 were identified to seek reliable data over the accidents. Three decision-makers were an Assistant Enforcement Officer from Road Transport Department of Kuala Terengganu (D_1), a Traffic Police Inspector from Police Traffic Department of Kuala Terengganu (D_2) and the third expert was an Assistant Superintendent of Fire Brigade Department of Kuala Terengganu (D_3). Data in form of linguistic variables were collected via interviewing of the three authorised personnel. The interview was conducted in three separated sessions to elicit the information about vehicles that regularly involved in accident. The questions of road accidents were mainly focused on opinion of the experts regarding rating of the vehicles prone to accidents based on the identified criteria. The experts were asked to specify rating of association of the criteria for each vehicle with linguistic expression varying from 'very poor' (VP), 'poor' (P), 'medium poor' (MP), 'fair' (F), 'medium good' (MG), 'good' (G), to 'very good' (VG). The weights of importance for each criterion were specified by experts with linguistics expression varying from 'very low' (VL), 'low' (L), 'medium low' (ML), 'medium' (M), 'medium high' (MH), to 'very high' (VH). These score were later aggregated to calculate the rating as a triangular fuzzy number for each criterion.

In this experiment, the rating x_{ij} of alternative A_i and the weights w_j of criteria C_j are assessed in linguistics terms represented by triangular fuzzy numbers as shown in Table 1 and Table 2.

Table 1: Linguistic Variables for the ratings x_{ij} of the vehicles A_j.

Linguistic Variables for The Ratings of The Vehicles	
Very Poor (VP)	(0, 0, 1)
Poor (P)	(0, 1, 3)
Medium Poor (MP)	(1, 3, 5)
Fair (F)	(3, 5, 7)
Medium Good (MG)	(5, 7, 9)
Good (G)	(7, 9, 10)
Very Good (VG)	(9, 10, 10)

Table 2: Linguistic Variables for the Weight w_j of criteria C_j

Linguistic Variables for The Ratings of The Vehicles	
Very Low (VL)	(0, 0, 0.1)
Low (L)	(0, 0.1, 0.3)
Medium Low (ML)	(0.1, 0.3, 0.5)
Medium (M)	(0.3, 0.5, 0.7)
Medium High (MH)	(0.5, 0.7, 0.9)
High (H)	(0.7, 0.9, 1.0)
Very High (VH)	(0.9, 1.0, 1.0)

The framework of this experiment can be seen in Fig.1.

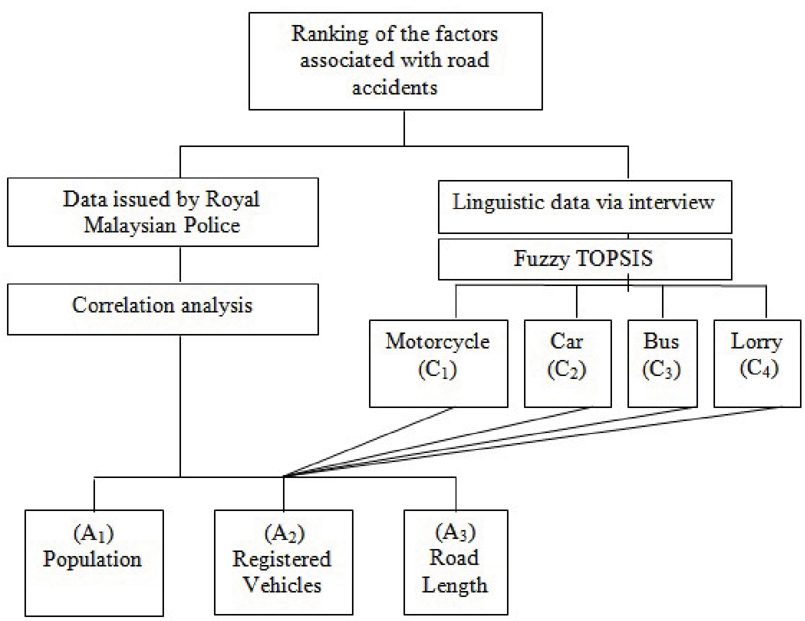

Fig. 1: Research Framework of the Decision Problem.

RESULTS AND DISCUSSION

Pearson correlation coefficients are calculated using Equation (2.1) and the results are shown in Table 3. According to Table 2, the relationship between road accident (RA) and population (A_1) is 0.987049. It is a strong positive correlation. The relationship between road accidents (RA) and registered vehicles (A_2) is also a strong positive correlation with different degree of correlations. Out of the three variables, registered vehicles achieved the highest positive correlation at 0.989428. The second highest positive correlation is the variable of population with 0.987049 correlation measure. The third place in correlation intensity is the variable of road length.

Table 3: Pearson correlation coefficients of road accidents model variables

	RA	A_1	A_2	A_3
RA	1.00	0.987049	0.989428	0.962852

From the strength relationships, it can be proposed that the ranking of the indentified factors associated with road accidents was $A_2 \prec A_1 \prec A_3$. The notation '$\prec$' represents one factor is highly associated to another.

The second analysis is conducted based on linguistic rating variables (see Table 1) supplied by the experts to evaluate rating of the alternatives with respect to each criterion and vehicles in form of decision matrix (Equation 3.1). The weight for each criterion is also translated into fuzzy weight based on definition in Table 2. These results are presented in Table 4 and Table 5.

Table 4: Decision Matrix

	C_1	C_2	C_3	C_4
A_1	(7,8.667,9.667)	(7.667,9.333,10)	(6.333,8.333,9.667)	(7,8.667,9.667)
A_2	(9,10,10)	(9,10,10)	(8.333,9.667,10)	(7.667,9.333,10)
A_3	(3.667,5.667,7.667)	(3,5,7)	(3,5,7)	(3,5,7)

Table 5: Weights for Criteria

	C_1	C_2	C_3	C_4
Weight	(0.833,0.967,1.0)	(0.7,0.867,0.967)	(0.3,0.5,0.7)	(0.233,0.433,0.633)

The equations (3.2) and (3.3) are applied respectively to yield the fuzzy normalized decision matrix and fuzzy weighted normalise decision matrix. These results are presented in Table 6 and Table 7.

Table 6: Fuzzy Normalized Decision Matrix

	C_1	C_2	C_3	C_4
A_1	(0.7,0.867,0.967)	(0.767,0.933,1.0)	(0.633,0.833,0.967)	(0.7,0.867,0.967)
A_2	(0.9,1.0,1.0)	(0.9,1.0,1.0)	(0.833,0.967,1.0)	(0.767,0.933,1.0)
A_3	(0.367,0.567,0.767)	(0.3,0.5,0.7)	(0.3,0.5,0.7)	(0.3,0.5,0.7)

Table 7: Fuzzy Weighted Normalized Decision Matrix

	C_1	C_2	C_3	C_4
A_1	(0.583,0.838,0.967)	(0.537,0.809,0.967)	(0.19,0.417,0.677)	(0.163,0.375,0.612)
A_2	(0.75,0.967,1)	(0.63,0.867,0.967)	(0.25,0.484,0.7)	(0.179,0.404,0.633)
A_3	(0.306,0.548,0.767)	(0.21,0.434,0.677)	(0.09,0.25,0.49)	(0.07,0.217,0.447)

After considering the equations (3.4), (3.5), (3.6) and (3.7), the final results of the fuzzy TOPSIS method is presented in Table 8. Due to limited space, the detailed results are not shown in this paper.

Table 8: Final ranking order

	D^*	D^-	Closeness Coefficient, C^*_j	Ranking order
A_1	1.798	2.500	0.582	2
A_2	1.55	2.703	0.636	1
A_3	2.601	1.664	0.390	3

The results of this investigation show that ranking of the factors link to road accidents is $A_2 \prec A_1 \prec A_3$. The notation '$\prec$' represents one factor is highly associated to another.

Two analyses show that registered vehicles rank in the first followed by population. The third place in ranking is road length.

Concluding Remarks:

The paper has designed to determine ranking of the selected factors associated with road accidents. Pearson Correlation and fuzzy TOPSIS have been accounted in this study. The correlation analysis proposed strength of relationships between the selected factors and road accidents. The fuzzy TOPSIS methods dealt with multiple factors and multiple motor vehicles involved in accidents. The statistics released by Royal Malaysian Police and the linguistic judgments from three authorized personnel were considered as sources of data. The two methods show that registered vehicles was the first factor associated with road accidents. Interestingly these two approaches produced the same ranking. These findings give insight to the contribution of each factor towards road accidents. Further work needs to be done to establish an accident index by considering more factors associated with accidents. The evidence from this study suggests that the three factors need to be thoroughly investigated in order to reduce and accidents. The rankingh based on linguistic judgment and strength of relationship was successfully made after taking into account the multiple factors.

ACKNOWLEDGMENT

We wish to acknowledge the support given to the second author by the three authorised personnel for voluntarily offering linguistic data.

REFERENCES

Bernama, 2006. Fatal Road Accident, http://www.bernama.com.my/bernama/v3/news.php?id=213066. retrieved on April 2006.

Chen, C.T., 2000. Extension of the TOPSIS for group decision making under fuzzy environment. Fuzzy Sets and Systems, 114(1): 1-9.

Chu, T.C., 2002. Selecting plant location via a fuzzy TOPSIS approach. International Journal of Advance Manufacturing Technology, 20: 859- 864.

Dagdeviren, M., S. Yavuz and N. Kilinc, 2009. Weapon Selection Using the AHP and TOPSIS Methods under Fuzzy Environment. Journal of Expert Systems with Applications: An International Journal, 36(4): 8143-8151.

Glover, T. and K. Mitchell, 2005. An Introduction to Biostatistics, Singapore Mc Graw Hill.

Hejar, A.R., M.Z., Afiah, S. Kulantayan and T.H. Law, 2005. Car Occupants Accidents and Injuries among Adolescents in a State in Malaysia. Proceedings of the Eastern Asia Society for Transportation Studies, 5: 1867-1874.

Hassan, T.A. and A.A. Mohamed, 2002. Investigating Driver Injury Severity in Traffic Accidents Using Fuzzy ARTMAP. Journal of Computer-Aided Civil and Infrastructure Engineering, 17(6): 398-408.

Herrera, F. and E. Herrera –Viedma, 2000. Linguistic decision analysis steps for solving decision problems under linguistic information. Fuzzy sets and Systems, 115(1): 67-82.

Law, T.H., R.S. Radin Umar and S.V. Wong, 2005. The Malaysian Government's Road Accident Death Reduction Target for Year 2010. Journal of IATSS Research, 29(1): 42-49.

Malaysian Ministry of Works, 2008. Overview of Current Road Safety Situation in Malaysia, Highway Planning Unit, Road Safety Section Ministry of Works. Retrieved on 2008. http://www.unescap.org/ttdw/roadsafety/Reports2006/Malaysia_RSpaper.pdf

Olson, D.L., 2004. Comparison of weights in TOPSIS model, Mathematical and Computer Modeling, 40(7): 721-727.

Qirjako, G., H. Burazeri, B. Hysa and E. Roshi, 2008. Factors Associated with Fatal Accidents in Tirana Albania: Cross-sectional Study. Croatian Medical Journal, 49(6): 734-740.

Royal Malaysia Police, 2008. Statistics of Road Accident and Death. At http://www.rmp.gov.my/rmp. Retrieved on April 08, 2008.

Semih, Ö. and S. Selin, 2008. Transshipment Site Selection Using the AHP and TOPSIS Approaches Under Fuzzy Environment. Journal of Waste Management, 28: 1552-1559.

Soma, B., A. Anna and C. Maureen, 2007. The value of mortality risk reductions in Delhi, India. Journal of Risk and Uncertainty, 34(1): 21-47.

Treat, J.R., N.S. Tumbas, S.T. McDonald, D. Shinar, R.D. Hume, R.E. Mayer, R.L. Stanisfer and N.J. Castellan, 1977. Tri-level study of the causes of traffic accidents. Report No. DOT-HS-034-3-535-77 (TAC).

UNESCAP: UNESCAP report Status of Road Safety in Asia, 2006: http://www.iatss.or.jp/english/research/30-1/pdf/30-1-12.pdf. Retrieved on 22 August 2008.

Zadeh, L.A., 1965. Fuzzy Sets, Information and Control, 8: 338-353.

Zimmermann, H.J., 1996. Fuzzy sets theory and its applications. Boston, Kluwer Academic Publishers.