FUZZY SETS IN THE SOCIAL SCIENCES: AN OVERVIEW OF RELATED RESEARCHES

MOHD LAZIM ABDULLAH¹, WAN SALIHIN WONG ABDULLAH² & ABU OSMAN MD TAP³

Abstract. A mathematical modelling based on the fuzzy sets theory is synonymous with the fields of engineering and computer science. Since the phenomena of fuzziness in the real world is also closely related with human beings, this paper discusses the application of fuzzy sets in social sciences. An overview of related research is presented, which focuses on an example of social research using a fuzzy model to measure the fuzzy data from a questionnaire. This paper also discusses the questionnaire that is used to collect data by many researchers in social sciences and the elements of fuzziness in it. The vague attributes and labels of a linguistic variable from the questionnaire would be transformed into standard mathematical values. This fuzzy model offers a substantial improvement in data analyses and gives more accurate conclusions. In short, this paper gives a new insight of the vast contribution of fuzzy sets in social sciences.

Keywords: Fuzzy sets, social sciences, fuzzy model, linguistic variables, fuzzy conjoint

INTRODUCTION

In everyday conversation, the use of imprecise and vague terms like “It is a humid day” or “You are good” or “The tall guy over there” are very commonly heard. The words such as humid, good, and tall have no clear boundaries. Each individual has their own interpretations of the meaning. Thus, for some people, good in examination implies a total mark above 80 percent while for others, it may be 50 percent. Another
example, for some people, a height of 1.57 meter is considered tall but not for others. The words such as humid, good, tall and many others are vague and fuzzy. Kantardzic (2003) tossed a slightly different example of vague phenomena in the real world. Rain is a common natural phenomenon that is difficult to describe precisely since it can rain with varying intensity, anywhere from a light shower to heavy downpour. Since the word rain does not adequately or precisely describe the wide variations in the amount and intensity of any rain event, rain is considered a vague phenomena. Very often, the concepts in the human brain for perceiving, recognising, and categorising natural phenomena are also vague and imprecise. The boundaries of these concepts are not clearly defined. Therefore, the judging and reasoning that emerge from them also become vague. In struggling to find a way of expressing succinctly the idea of vagueness in life, Zadeh (1965) proposed the idea of fuzzy sets.

2.0 THE DEFINITION OF FUZZY SETS

In order to explain the concepts of fuzzy sets, the basic idea in classical set theory must be understood. In mathematics, the concept of classical set is very simple. A set is a collection of well-defined objects. These objects cover almost anything that can either belong or do not belong to the set.

The classical set $A$ in the universe $U$, $A \subset U$ is normally characterised by the function $\mu_A(x)$, which take value 1 or 0, indicating whether or not $x \in U$ is a member of $A$:

$$\mu_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{for } x \in A \end{cases}$$

(1)

Hence, $\mu_A(x) \in \{0, 1\}$. The function $\mu_A(x)$ as equation (1) takes only the value 1 or 0.

Assume that the function $\mu_A(x)$ may take values in the interval $[0, 1]$. In this way, the concept of membership is not any more crisp, but become fuzzy in the sense of representing partial belonging or degree of membership (Bojadziev, 1999). A fuzzy set $R$ is defined by:

$$R = \{(x, \mu_R(x)) / x \in A, \mu_R(x) \in [0,1]\}$$

(2)

where $\mu_R(x)$ is a function called membership function; $\mu_R(x)$ specifies the grade or degree to which any element in $A$ belongs to the fuzzy set $R$.

Ragin (2000) had a very simple explanation about fuzzy sets. He iterated that the basic idea behind fuzzy sets is to permit the scaling of membership scores and this allows partial or fuzzy membership. A membership score of 1 indicates full membership in a set; a score close to 1 (e.g., 0.8 or 0.9) indicates strong but partial membership in a set; scores less than 0.5 but greater than 0 (e.g., 0.2 and 0.3) indicate that objects are more “out” than “in” a set, but still weak members of the set;
FUZZY SETS IN THE SOCIAL SCIENCES: AN OVERVIEW OF RELATED RESEARCHES

A score of 0 indicates full non-membership in the set. Thus, fuzzy sets combine qualitative and quantitative assessment.

The fuzzy set theory can represent the uncertainty or vagueness inherent in the definition of linguistic variables (Zadeh, 1975). Age is one of the examples of a linguistic variable whose values are words like very young, middle age, and very old. The situation of fuzziness and the role of linguistic variables will be explained further in this paper. In short, fuzzy set theory is becoming an alternative way in explaining the fuzzy phenomena in the real world.

3.0 FUZZINESS: FOR WHOM?

In describing the target for fuzziness, it is very useful to look at this question: which one have an uncertain criteria, engineering and science or social sciences and the humanities? In answering this question, Mukaidono (2001) stated that social sciences and humanities are blooming in the world of fuzziness and uncertainties. On the other hand, the target in science and engineering is objects.

The fuzzy theory was pioneered by an electrical engineering professor, L.A. Zadeh and the theory was developed for practical use in engineering. Mukaidono (2001) stressed that the original effectiveness should be in social science and humanities since the target is human, for which uncertainty is inevitable. Fuzzy theory is expected to have an impact on social sciences and humanities. Unfortunately, fuzzy theory is not known to many researchers in these fields (Smithson, 1987). There are no remarkable results yet. However, some researchers who are engineering and mathematics based are trying to apply fuzzy theory to non-engineering fields. Some researchers in the social sciences are beginning to use it as a new measure (Threadwell, 1995). Some of the applications of fuzzy sets theory in the social sciences will be discussed in this paper.

4.0 FUZZY SETS AND APPLICATION IN SOCIAL SCIENCES

Zimmerman (1991) summarised the development in fuzzy sets theory into two different areas. The first area categorised fuzzy set theory as a formal theory when it is matured it becomes more sophisticated, specified, and was enlarged by the original ideas or concepts, as well as embracing classical mathematical areas such as algebra, graph theory, topology, and so on generalising them.

The second area categorised fuzzy set theory as a very powerful modelling tool that can cope with a large fraction of uncertainties of real life situations. Armed with the nature of its generality, fuzzy sets theory can be well adapted to different circumstances and contexts.

Zimmerman (1991) outlined a wide range of different applications of the fuzzy set theory. He provided a classification of four different types of applications. Firstly, fuzzy sets were mainly applied in mathematics. These applications covered the
generalisation of traditional mathematics such as topology, graph theory, algebra, logic etc. The largest and most important of this type of application is undoubtedly fuzzy logic. Secondly, fuzzy sets are also applicable to the algorithms in clustering methods, control algorithms, and mathematical programming. Zimmerman (1991) also gave the third application of fuzzy sets in the standard models. Many models such as transportation model and inventory control model, used the fuzzy set theory. Its application to the different real-world problems is the fourth application of fuzzy sets. The most important of these would include fuzzy expert system and fuzzy control. Others would include applications to psychology (Kochen, 1975).

In the fourth classification, the application of fuzzy sets is likely to expand to broader fields including human sciences. These fields seem rather isolated from the world of mathematics but the fuzzy set theory tries to slowly place the applications in the context of the real world. The dialogue between the human sciences and fuzzy set theory has been scattered, unsystematic, and slow to develop. Smithson (1987) was cited as saying that fuzzy sets mathematics is couched in foreign and rather obtuse notation, which is forbidding even to the mathematically sophisticated behavioural scientist.

Apparently, most of the topics in fuzzy set assume either a mathematical computer, science or engineering orientations. Smithson (1987) forwarded his ideas in an attempt to bridge the gap and lighten mathematics to illustrate the basic elements of fuzzy set theory in the real-world research. The same spirit and goal in minimising the obtrusion of mathematics and make mathematics more palatable to a wider audience was shared by Treadwell (1995). He set out to provide some structure for handling fuzzy concepts and illustrated their use in budgeting, and decision making.

Fuzzy set theory was also reported to have been applied in library and information sciences by Hood & Wilson (2002). Within library and information system, fuzzy set theory has been applied to traditional librarianship, as well as to problems in information retrieval. Many librarians have to make decisions about when and whether to bind their periodicals. The decision may be based on a number of criteria including the number of missing issues, the future expected use of periodicals etc. Each of these criteria is vague and can be modelled with a fuzzy set theory. In addition, the decision may be based on the opinions of more than one decision maker. One very good example of fuzzy set theory application to library-decision making can be taken from Turner and O’Brien (1984).

Fuzzy set theory has been applied in many areas of human sciences such as production, management, and education. In Malaysia, many researchers have contributed to the development of fuzzy sets and various applications in these areas. Examples cited here have sparked the dissemination of the knowledge of fuzzy systems and encouraged applications of fuzzy systems by other researchers. Mahmod and Nazri (2002) reported the application of fuzzy sets in assessing the lecturers’ performances. The present methods of assessment had been compared with the
assessment using fuzzy evaluation approach. This research was based on the model proposed by Feng Chu (1990) and Biswas (1995).

The fuzzy sets approach also attracted Jasmani and Abu Osman (2002) to evaluate customers’ post-purchasing satisfaction level on three commodities i.e. onions, rice and fish. They were using the preference model proposed by Turksen and Willson (1994). Ilayni (2001) evaluated the levels of difficulties of learning faced by undergraduates. Most recently, Lazim (2004) explored the same model in measuring the effectiveness of computer algebra systems in the classroom. A brief explanation of the application of a fuzzy model is discussed in the following section.

5.0 FUZZY QUESTIONNAIRE

It is a normal practice for researchers, especially for social scientists to analyse data using a statistical analysis package software such as SPSS. This software facilitates all numerical calculations, graphs, and tables in accordance with research objectives. Firstly, data must be collected using a specific instrument. There are five major techniques for gathering quantitative data: tests, questionnaires, interviews, observation, and unobtrusive measures (McMillan, 2001). One of the most popular techniques among researchers in social sciences study is using questionnaires. There are several kinds of question and response modes in questionnaires, for example, dichotomous questions, multiple choice questions, rating scales, and open-ended questions (Cohen, 2001).

A questionnaire normally comes with statements or questions, but in all cases the subject is responding to some specific purposes. There are many ways in which a question or statement can be worded, and several ways in which the response can be made. One of the most popularly used response is a combination of statements and gradation of the statement in the form of scales such as the Likert scale. Every item is usually stated in the form of a statement in which the respondents have to choose a response from a scale of 1 to 5. For instance, the statement ‘I am not very interested in joining the armed forces’ is followed by a 5 point rating scale of strongly agree (1), agree (2), neither agree nor disagree (3), disagree (4), and strongly disagree (5). Respondents will make their choice from this 5 point rating scale. After collecting all the data, statistical analysis such as descriptive statistics and graphical analysis are used to analyse the data. Finally, the researchers draw a conclusion based on the analysed data.

These are the basic steps that researchers follow in analysing data. Looking thoroughly at the prescribed ways of collecting and analysing data, several issues should be look into. When a respondent fill out a questionnaire, he or she may notice that there are some uncertain choices other than agree or disagree. For example, the choices could be partially agree, neither, probably or don’t know. Also, the statement in the questionnaire could be fuzzy and vague. The word ‘not very interested’ is very hard to grasp objectively. Furthermore, the degree of ‘strongly
agree’ and ‘agree’ are bounded with nominal integers of 1 and 2. Is there any other value between these two numbers?

Also, the value of the mean is merely a measurement of central tendency with little emphasis on extreme values. This central tendency takes into account all integers (rating scales), and the value of the mean will emerge after the process of cancellations of the two-end extreme values. Similarly, the percentage measure just portrays the pattern of the frequencies. Despite the fuzziness and ambiguity of questionnaires, researchers conclude with convincing and comprehensive findings. Hence, it is important to look into an alternative approach using highly reliable fuzzy questionnaires. This is where fuzzy theory can be applied successfully. Since the questionnaires are considered as fuzzy, then the answers to the questionnaire (from the respondents) can also be considered as fuzzy data.

Fuzzy sets theory recognises the properties of agree, disagree, slow, tall, adequate, and competent by giving variations for each category. Cohen (2001) stressed specifically the significance of fuzzy logic in social science research. He stated that fuzzy logic enables us to gain a more precise measurement of the variance within and between these semantic categories (e.g. on rating scale); it recognises that imprecision, rather than bivalence (either something is or is not the case), is a characteristic of many phenomena. The following example of a social science research is to measure the fuzzy variables in a computer algebra system learning environment. The fuzzy questionnaires which had linguistics modifiers and fuzzy statements could be treated in a distinct manner. The alternative approach was purely based on the Fuzzy Set Theory pioneered by Zadeh (1965).

6.0 FUZZY MODEL: AN EXAMPLE OF RESEARCH

As explained in section 5.0 of this article, the source of data was entirely obtained from a fuzzy questionnaire. The questionnaire was constructed based on the preference models widely used in new product design, marketing management, and market segmentation (Green & Srinivasan, 1990; Wittink & Cattin, 1989). Preference models can also be extended in many areas which are vague, uncertain, and very subjective. This study is a modification of the analogous application of Turksen and Willson (1995) preference model, whereby the linguistic variables are prevalent in both psychological attributes (for example ‘difficult’), and subject preferences (for example ‘strongly agree’). Vagueness of rating ‘strongly agree’ is inherent rather than due to a lack of knowledge about the available rating. The fuzzy set for ‘strongly agree’ would consists of element pairs, each a domain variable and a degree of membership. A membership function maps each value of the domain variable to a degree of membership or belongingness in the set that range from 0 to 1.

A fuzzy set preference model requires a fuzzy set definition for each of linguistic ratings on the measurement scale. In this research, an approach similar to the Likert
scale is used for all preference ratings that provided subject with a balanced selection of 7 linguistic terms to pick from the attributes of the agreement. This scale has a central neutral evaluation with three positive and three negative evaluations. The underlying theory of fuzzy sets in the preference modelling can be retrieved further from Turkson and Willson (1995). The appeal of using fuzzy sets in preference models comes from representing linguistic variables in a mathematical structure that closely correspond to the actual subject preferences. An overall preference for a statement can be decomposed into a combination of preferences for its constituent parts (attributes), which are combined using a combination function. A combination of preferences becomes the main underlying philosophy in the fuzzy set conjoint model.

Turksen and Willson (1994) had used the application of fuzzy set into ordinary conjoint model based on preference model. This model is known as the Fuzzy Set Conjoint Model. By using this model, a fuzzy set $R$ is formed to represent the hierarchy of all respondent against the specific attributes. This approach gives a degree of agreement for each selected attribute that were used in this study.

Approximate degree of membership for each element, $y_j (j = 1,2,3,\ldots,l)$ in fuzzy set $R$ is defined as:

$$
\mu_R(y_j, M) = \sum_{i=1}^{n} W_i \cdot \mu_{R_i}(x_j, M)
$$

in which,

$\mu_{R_i}(x_j, M)$: degree of membership for each element $x_j$ in linguistic evaluation $R_i$ by $i$-th respondent against attribute $M$ (for each element fuzzy set $R_i$, $x_j = 1,2,3,\ldots,l$

$R_i \in \{\text{very strongly agree (L1), strongly agree (L2), agree (L3), undecided (L4), disagree (L5), strongly disagree (L6), very strongly disagree (L7)}\}$ by $i$-th respondent, $i = 1, 2, 3, \ldots, 23$ against attribute $M$

$W_i$: weight for $i$-th respondent and for $W_i = \frac{w_i}{\sum_{k=1}^{n} W_k}$, as $w_i$ is a score of linguistic values given by $i$-th respondent.

$L$: number of linguistic values used (in this study, $l = 7$)

$\mu_R(y_j, M)$: approximate overall degree of membership of the linguistic value $R$ for all factor $M$ attributes based on domain value $y_j = 1, 2, \ldots, 23$. 
6.1 Definition of Fuzzy Linguistic Variable

According to Bojadziev (1999), variables whose values are words or sentences in natural or artificial languages are called linguistics variables. In this study, ‘agreement’ is a linguistic variable whose values are words like very strongly agree, strongly agree, agree, undecided, disagree, strongly disagree, and very strongly disagree. They are called terms or labels of linguistic variable ‘agreement’. As mentioned earlier, a fuzzy set is completely characterised by its membership function. Therefore, each term or label is defined by an appropriate membership function. The seven terms or labels of linguistic variable of ‘agreement’ that is Lk = {very strongly agree, strongly agree, agree, undecided, disagree, strongly disagree, very strongly disagree} are defined by membership functions as:

- **Very strongly agree**, L1 = \{1/1, 0.8/2, 0.5/3, 0.2/4, 0/5, 0/6, 0/7\}
- **Strongly agree**, L2 = \{0.7/1, 1/2, 0.6/3, 0.4/4, 0/5, 0/6, 0/7\}
- **Agree**, L3 = \{0.4/1, 0.6/2, 1/3, 0.6/4, 0.4/5, 0/6, 0/7\}
- **Undecided**, L4 = \{0/1, 0.3/2, 0.7/3, 1/4, 0.7/5, 0.3/6, 0/7\}
- **Disagree**, L5 = \{0/1, 0.2/2, 0.4/3, 0.6/4, 1/5, 0.6/6, 0.4/7\}
- **Very disagree**, L6 = \{0/1, 0/2, 0/3, 0.4/4, 0.6/5, 1/6, 0.7/7\}
- **Very strongly disagree**, L7 = \{0/1, 0/2, 0/3, 0.2/4, 0.5/5, 0.8/6, 1/7\}

The choice of fuzzy sets of linguistics variables of ‘agreement’ and the fuzzy data from respondents are then inserted into the equation (3). Turksen and Willson(1995) added that in order to get the overall measurement of factor attribute for the linguistic variable, the procedures of getting the degree of similarity must be implemented.

6.2 Degree of Similarity

The attribute’s actual overall measurement is compared to the model’s calculated overall measurement for each of the term or label in the order of ranking. The comparison continues until the correspondence in ranking between attribute measurement and estimated model of measurement is broken. For the fuzzy set model, a similarity measure is used to calculate the sum of the Euclidean distance between corresponding elements in the calculated fuzzy set and the attribute’s actual fuzzy rating (linguistic term). The formula for the similarity of two sets is given as:
\[
Sim(R, L_k) = \frac{1}{1 + \sqrt{\sum_{j=1}^{7} (\mu_{R(j,M)} - \mu_{L_k(j)})^2}}
\]

where \( \mu_{L_k(j)} \) is the fuzzy set defined for linguistic rating \( l \) and \( \mu_{R(j,M)} \) is the calculated set for attribute \( M \) from equation (3). The squared difference of the degree membership of the \( j \)-th element of each set is summed for the seven elements in the two sets. The similarity measure is the reciprocal of the square root of the sum added to 1. The similarity is computed between attribute \( M \) and each of the seven possible linguistic terms \( l \). The similarity score ranges from 0 to 1 and provides ordinal information, which is sufficient to determine the measurement.

### 6.3 Result

Let’s consider one statement or in this study called as an attribute in the questionnaire. ‘Procedural knowledge methods’ are difficult to master with 7 scales of preferences. In this study, the word labels or terms are used as scales of preferences. For example, if there are six respondents’ responses to the degree of agreement about the above attribute, then the set of respondents is denoted as \( S = \{S_1, S_2, S_3, S_4, S_5, S_6\} \). The set of terms or labels of the linguistic variable ‘agreement’ given by six respondents is the ‘degree of agreement’ = \{1, 2, 3, 1, 2, 2\}. The set of respondents are mapped to the set of degree of agreement, i.e, for respondent 1, \( S_1 \) is mapped to very strongly agree (1), for respondent 2, \( S_2 \) is mapped to strongly agree and so on.

Normally, by using the mean formula, the value of mean is easily calculated. In this example, the mean of that statement is 1.83. This value can be laterally interpreted that the six respondents almost strongly agreed that procedural knowledge methods are difficult to master. This is a very simple statistical interpretation even though the exact value of 2 is representing ‘strongly agree’. In order to make a conclusion, the approximation is forced to be used and eventually neglecting the actual value. It seems impossible to get the exact value (in this case, integer 2), since the calculation is based on the mean formula.

By considering the same method of data collection, the numerical inputs were inserted into the fuzzy model as expressed in equations (3) and (4). The calculated degrees of similarity are:

\[
\begin{align*}
Sim(R, L_1) &= 0.6878 \\
Sim(R, L_2) &= 0.8293 \\
Sim(R, L_3) &= 0.6275 \\
Sim(R, L_4) &= 0.4464
\end{align*}
\]
Sim (R, L5) = 0.3967
Sim (R, L6) = 0.3519
Sim (R, L7) = 0.3483

These ordinal results, then were brought to the conclusive measurement at Sim (R, strongly agree)_{max} = 0.8293. This value showed that the six respondents strongly agreed that procedural knowledge were difficult to master at 0.8293 degree of agreement. The fuzzy statement does not stop at the conclusion ‘strongly agree’. Fuzzy set model provides a very significant value to represent the strength of ‘strongly agree’. This degree of membership makes fuzzy approach more reliable, distinct, and precise. Since the process of getting these values was quite complex, this paper does not intend to give the details of such algorithms and procedures involved.

7.0 CONCLUSION

The study presented an example of research work using mathematical theory of fuzziness which has been used to model a variety of situations in social sciences. The subjective vagueness and uncertainties in social sciences can be overcome to a certain degree. Studies highlighted here is just a few examples of the broad applications of fuzzy sets. Indeed, the research of fuzzy systems and its application in the social sciences and the use of fuzzy theory itself in Malaysia is still new. There is still room for further applications using the fuzzy systems and probably more promising models will be developed in the future to give alternative ways of managing social science related researches. The benefits to be gained by using fuzzy theory need to be tested and explored.

REFERENCES


