Generalised Fuzzy Bayesian Network with Adaptive Vectorial Centroid

Ku Muhammad Naim Ku Khalif[1] Alexander Gegov[2]

[1]School of Computing, University of Portsmouth, Buckingham Building, Portsmouth PO1 3HE, United Kingdom

Abstract

In this paper, the theoretical foundations of generalised fuzzy Bayesian Network based on Vectorial Centroid defuzzification is introduced. The extension of Bayesian Network takes a broad view by examples labelled by a fuzzy set of attributes, instead of a classical set. Combining fuzzy set theory and Bayesian Network's knowledge allows the use of fuzzy variables or attributes that widely used in various applications in science and engineering. It is so highlights the integration of both knowledge's considers the need of human intuition in data analysis. Through the experimental comparison and analysis on the BUPA-liver disorder dataset, the proposed methodology is then validated theoretically and empirically.

Keywords: Centroid defuzzification; Vectorial Centroid; Bayesian Network; Human Intuition.

1. Introduction

Over the last decade, a lot of techniques dealing with vagueness and imprecision have drawn the attentions of researchers and applied scientists. In real world phenomena, decisions are made based on information given which is known as data. However, information about decision is always uncertain. The uncertain information may include randomness, vagueness and fuzziness. Tang et al, (2002) claim, there are critical problems of the artificial intelligence research field that always arise which are: how to represent the uncertain information precisely; and how to reason using uncertain information. The theory of probability either in subjective or objective cases can deal with uncertainty due to randomness. Bayesian knowledge has been important tools that can represent the human knowledge under uncertain circumstances. On the other hand, the uncertainty due to vagueness and randomness can be dealt with the fuzzy knowledge.

Bayesian Network is developed method from classical mathematics theory where has a mathematical basis and stable structure development. Theoretically, the Bayesian Network model uncertain knowledge representation and reasoning where the fundamental bases are on probabilities and graph theory (Tang & Liu, 2007). It shows the factors and their interactions that relate to a response of interest (Jensen & Nelsen, 2007; Pearl 1998). Constructing Bayesian Network is a graphical and qualitative illustration of relationships among different nodes using directed arcs. In establishing the structure of constructing Bayesian Network is using data sets in machine learning either based on complete or incomplete data (Gasse et al, 2014). The second way in constructing the Bayesian Network is some researchers have proposed several approaches by using existing models of the system (Boujla et al, 2014). Aguilera et al. (2010) claim one of the most essential advantage of Bayesian Network is that the directed acyclic graphs determine the dependence and independence relationships among the variables, where it is possible to find out, with no need of carrying out any numerical calculations, which variables are relevant and irrelevant for some other variable of interest. The frequentist's model only take mean values into account where the Bayesian construct the model by means of probability distributions. Since nodes are modelled by means of probability distributions, risk and uncertainty can be estimated more accurately (Uusitalo, 2007). Bayesian Network capable to model complex systems with a large numbers of variable and manageable missing values in input data in proper prediction (Getoor et al, 2004; Uusitalo, 2007).

However, there are some limitations in handling Bayesian Network where in real world problems, the fact that most of the data available are continuous or hybrid and sometime in fuzzy nature. Even though Bayesian Network can manage them, the limitations are too restrictive (Nyberg et al, 2006; Uusitalo, 2007). In dealing with imprecision or fuzzy events, Bayesian Network is useful tool to deal with probabilistic theory, but unable to handle these fuzzy events properly since to express the ambiguity of data sets in a model. These models also are not yet incorporated to the typical commercial Bayesian Network software.

In dealing with imprecision or fuzzy events, Zadeh (1965) was introduced fuzzy set theory to represent vagueness or imprecision in a mathematical approach. The main motivation of using fuzzy set theory shows its ability in appropriately dealing with imprecise numerical quantities and subjective preferences of decision makers (Deng, 2013). It is typical needed that defuzzification plays a significant role in the performance of fuzzy system's modelling techniques (Yager, 1994). Defuzzification process is guided by the output fuzzy subset that one value would be selected as a single crisp value as the system output. Zimmerman (2000) claims that the fuzzy numbers are represented as possibility distribution where most of the real-world phenomena that exist in nature are fuzzy rather than probabilistic or deterministic.

Centroid defuzzification is most commonly technique that has been applied in various discipline areas where has been explored for the last decade. In ranking fuzzy numbers, centroid defuzzification is classified as one of the major classes where it can provide a very

useful to support the representation of fuzzy numbers possibility distribution. The involvement of centroid defuzzification concept in ranking fuzzy numbers only started in 1980 by Yager. Followed by Murakami et al. (1983), Cheng (1998), Chu and Tsao (2002), Chen and Chen (2003), Liang et al. (2006), Chen and Chen (2007) and Wang and Lee (2008). All researchers have their own definition of centroid concept in ranking index. Some of them contribute in ranking fuzzy numbers that are based on the value of $\tilde{x}$ alone whilst some are based on the contribution of both $\tilde{x}$ and $\tilde{y}$ values. Ramli and Muhamad (2009) claim that Wang et al. (2006) and Shieh (2007) methods produce correct centroid point formula for fuzzy numbers that very useful computational support in ranking fuzzy numbers based on the centroid concepts. Wang et al. (2006) and Shieh (2007) not focused on finding the best ranking fuzzy numbers, but on producing the correct formula for centroid defuzzification point $(\tilde{x}, \tilde{y})$ itself that can be used in many applications in real world problems other than ranking only. Shieh (2007) proposed the following centroid formulation for all fuzzy numbers

$$\tilde{x}_0(A) = \frac{\int_{-\infty}^{\infty} x A(x) dx}{\int_{-\infty}^{\infty} A(x) dx} \quad (1)$$

$$\tilde{y}_0(A) = \frac{\int_{-\infty}^{\infty} \alpha |A(x)| d\alpha}{\int_{-\infty}^{\infty} |A(x)| d\alpha} \quad (2)$$

The centroids of fuzzy numbers normally are extracted from geometric aspects where is to construct various order relationship from the perspective of membership function to some extent. Fuzzy set theory has done for every single part of the official analysis when dealing with the vagueness and imprecision in human decision making. The power of human being in making logical decisions using imprecise and incomplete information has led to the uncertainty in terms of decision in formativeness.

This paper presents a novel approach that combines fuzzy set theory and Bayesian Network approach using Vectorial Centroid defuzzification which discusses the imprecision in probability of the fuzzy events. While classical Bayesian Network and established Shieh (2007) centroid method that used in Bayesian Network have been employed in this study, the proposed methodology is used there to work only if one can give a probability distribution for all attributes which doesn't seem meaningful in the case of fuzzy truth values. The proposed research covers all possible cases of fuzzy numbers that exist nowadays. This study employs alternative approach to extend the algorithm to cater all possible fuzzy cases. The extension of Bayesian Network into fuzzy states will increase robustness, allows impre-

cise or contradictory inputs, permits fuzzy thresholds of probabilistic dependence and reconciles conflicting that relates input to output.

The rest of this paper is organised as follows: In Section 2, this paper discuss the theoretical preliminaries of fuzzy set theory, generalised trapezoidal fuzzy number, Bayesian equation and Bayesian Network formulation. This is then proceeded to the proposed work of generalised fuzzy Bayesian Network using Vectorial Centroid method in Section 3. Section 4 discusses validation processes of proposed methodology theoretically and empirically.

2. Theoretical Preliminaries

In this section, some basic definitions and arithmetic operations concepts are presented in this paper

2.1 Fuzzy Set Theory

Fuzzy set is an extension of regular set of numbers in the sense that does not refer to one single crisp value but rather to a connected set of possible values with the membership function is between 0 and 1. The elementary function $\mu_{\tilde{A}}$ of a crisp set $A \subseteq X$ assign a value either 0 or 1 to each member in X. This function can be universal to a function $\mu_{\tilde{A}}$ such that the value is assigned to the element of the universal set X fall within a specified range. The assigned value indicates the membership grade of the element in the set A. The function $\mu_{\tilde{A}}(x)$ is known as a membership function where the set $\tilde{A} = \{(x, \mu_{\tilde{A}}(x)); x \in X\}$ defined by $\mu_{\tilde{A}}(x)$ for each $x \in X$ is called fuzzy set. Fuzzy set $\tilde{A}$ is defined on the universal set of real numbers $\Re$, where has the following characteristics (Kaufmann & Gupta, 1988)

a. $\mu_{\tilde{A}} : \Re \to [0,1]$ is continuous

b. $\mu_{\tilde{A}}(x) = 0$ for all $x \in (-\infty, a] \bigcup [d, \infty)$

c. $\mu_{\tilde{A}}(x)$ is strictly increase on $[a,b]$ and strictly decrease on $[c,d]$

d. $\mu_{\tilde{A}}(x) = 1$ for all $x \in [b,c]$, where $a < b < c < d$

2.2 Generalised Trapezoidal Fuzzy Number

A generalized trapezoidal fuzzy number can be represented by the following membership function is given by (Chen & Chen, 2009)

$$\mu_{\tilde{A}}(x) = (a,b,c,d;h) = \begin{cases} h \cdot \dfrac{(x-a)}{(b-a)} & if \quad a \le x \le b \\ h, & b \le x \le c \\ h \cdot \dfrac{(x-d)}{(c-d)} & if \quad c \le x \le d \\ 0 & otherwise \end{cases} \quad (3)$$

For a trapezoidal fuzzy number, if $b = c$, then the fuzzy number is in the form of the triangular fuzzy number.

Whereas, if $a = b = c = d$ for both triangular and trapezoidal fuzzy numbers, then both fuzzy numbers are said to be in the sort of singleton fuzzy number. The length between a and d are known as the core of the fuzzy number.
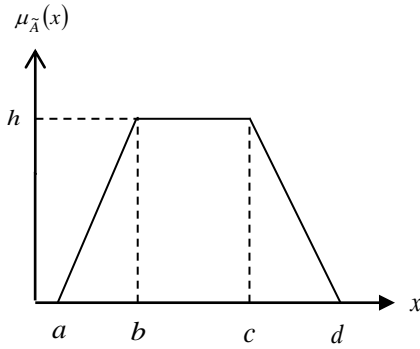

Fig. 1: Trapezoidal Fuzzy Number

2.3 Bayesian Equation

The principal of Bayesian knowledge applying simple Bayesian theory in complex causality events. If X and Y are events, where $P(Y) > 0$, then the conditional probability of A, given B is (Liu, 2000)

$$P(X|Y) = \frac{P(XY)}{P(Y)} \qquad (4)$$

The joint probability is

$$P(XY) = P(Y)P(X|Y) = P(X)P(Y|X) \qquad (5)$$

Assume that $\{B_i, i \in n\}$ is a number collection of events. Let X be another event and suppose that $P(Y_i)$ and $P(X|Y_i)$ for each $i \in n$.

The total probability formula is denoted as

$$P(X) = \sum_{i=1}^{n} P(Y_i)P(A|Y_i) \qquad (6)$$

Then the Bayesian equation is

$$P(Y_i|X_j) = \frac{P(X_j Y_i)}{P(X_j)}$$

$$= \frac{P(Y_i)P(X_J|Y_i)}{\sum_{i=1}^{n} P(Y_i)P(X_j|Y_i)} \qquad (7)$$

$$i = 1,2,...,n; \quad j = 1,2,...,m$$

Suppose that $V_1, V_2,..., V_N$ are events in a general case where intersection has positive probability. The multiplication rule of probability of Bayesian equation is

$$P(V_1, V_2,..., V_N)$$

$$= P(V_1)P(V_2|V_1)...P(V_N|V_1, V_2,..., V_{N-1}) \qquad (8)$$

2.4 Bayesian Network

Bayesian Network represents a directed acyclic graphs (DAG), which consist of nodes, directed arcs and the conditional probability (Pearl, 2000). According to Cheng et al. (2012), Bayesian Network is inference engine for the computation of beliefs of events given the observation of other events which is known as evidence, where the calculation of the probabilities of the occurrences of some events given the evidence. For every single node represents each event or variable. The directed arc between two events represent the direct causality and its degree can be expressed as by conditional probability formulation. In Bayesian Network, a node "(V_i)" given its parents is conditionally independent of its non-descendants.

$$P(V_i|A(V_i), F(V_i)) = P(V_i|F(V_i)) \qquad (9)$$

$A(V_i)$: The set of non-descendants of "V_i"

$F(V_i)$: The set of parent nodes of "V_i"

Referring to the conditional probability, the $V_1, V_2,..., V_N$ are nodes of a Bayesian Network and the joint probability can be expressed as

$$P(V_1, V_2,..., V_N) = \prod_{1}^{N} P(V_i|F(V_i)) \qquad (10)$$

where $F(V_i)$ is the set of parent nodes of "V_i" and the marginal probability of "V_i" is

$$P(V_i) = \sum_{except V_i} P(V_1, V_2,..., V_N) \qquad (11)$$

Let assume that the evidence e is given, and then the Bayesian Network formula can be denoted as

$$P(V_1, V_2,..., V_N|e) = \frac{P(e_i)P(e_i|V_1, V_2,..., V_N)}{\sum_{i=1}^{n} P(e_i)P(e_i|V_1, V_2,..., V_N)} \qquad (12)$$

3. Proposed Methodology

As noted in the introduction, the useful of fuzzy numbers nowadays a widely applied in many research problems in dealing with human intuition in data analysis. In machine learning systems, most of researchers attempt to eliminate the need of human intuition in their data analysis. Human intuition or human judgment can't be eliminated because it can lead us to uncertainty, vagueness and randomness.

In the fuzzy case, the authors generalise the concept of attributes to $\mu_{\tilde{A}} \in [0,1]$. The values of the varia-

bles correspond to generalised trapezoidal fuzzy numbers where capable to represent all type of fuzzy numbers in such a way that can cater human intuition in machine learning systems properly. In various applications in science and engineering, there will be a need to defuzzy the fuzzy values into crisp values. In this section, this study propose a new centroid defuzzification method that can be applied properly in Bayesian Network algorithm. The methodology consist of two stages here namely

A. Stage one

The development of Vectorial Centroid defuzzification method for fuzzy sets.

B. Stage two

The implementation of Vectorial Centroid in Bayesian Network algorithm.

Full description for both stages are as follow:

A. Stage one

Let consider $\tilde{A} = (a,b,c,d;h)$ as the generalised trapezoidal fuzzy number. The complete method process for Vectorial Centroid is signified as follow

Step 1: Find the centroids of the three parts of α, β and γ in trapezoid plane representation as shown in Fig. 2
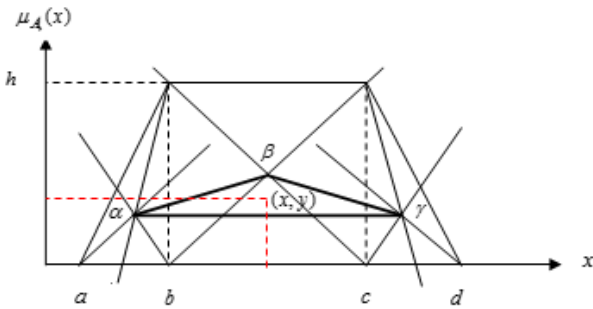

Fig. 2: Vectorial Centroid plane representation

$$\alpha(x,y) = \left(a + \left[\frac{2}{3} \left(\frac{a+b}{2} - a \right) \right], \frac{h}{3} \right) \quad (13)$$

$$\beta(x,y) = \left(\frac{b+c}{2}, \frac{h}{2} \right) \quad (14)$$

$$\gamma(x,y) = \left(d + \left[\frac{2}{3} \left(\frac{c+d}{2} - d \right) \right], \frac{h}{3} \right) \quad (15)$$

Step 2: Connect all vertices centroids points of α, β and γ each other, where it will create another triangular plane inside of trapezoid plane.

Step 3: The centroid index of Vectorial Centroid of $(\tilde{x}, \tilde{y})$ with vertices α, β and γ can be calculated as

$$VC_{\tilde{A}}(\tilde{x}, \tilde{y}) = \left(\frac{\alpha + \beta + \gamma}{3}, \beta + \left[\frac{2}{3} \cdot \left(\frac{\alpha + \gamma}{2} - \beta \right) \right] \right) \quad (16)$$

Vectorial Centroid can be summarised as

$$VC_{\tilde{A}}(\tilde{x}, \tilde{y}) = \left(\frac{4(a+d) + 5(b+c)}{18}, \frac{7h}{18} \right) \quad (17)$$

where

$\tilde{x}$: the centroid on the horizontal x-axis

$\tilde{y}$: the centroid on the vertical y-axis

$(\tilde{x}, \tilde{y})$: the centroid point of fuzzy number $\tilde{A}$

Centroid index of Vectorial Centroid can be generated using Euclidean Distance by Cheng (1998) as

$$R(\tilde{A}) = \sqrt{\tilde{x}^2 + \tilde{y}^2} \quad (18)$$

B. Stage two

Extend the Bayesian Network formulation in fuzzy states of nature, where if have fuzzy data set, defuzzification process is needed in converting into crisp values where at the same time the fuzzy nature is not lost. Re-interpretation of degree $\mu_{\tilde{A}} \in [0,1]$ using Vectorial Centroid to the $P(V_1, V_2, ..., V_N | e)$ is developed as follows:

Step 1: Lift the reintergration of the fuzzy values membership function using generalised trapezoidal fuzzy numbers. Vectorial Centroid formulation are applied for generalised trapezoidal fuzzy numbers rule formula. The $\mu_{\tilde{A}}$ represents as $(a,b,c,d;h)$ in calculation to avoid cluttering. Suppose that $\mu_{\tilde{V}}$ and $\mu_{\tilde{e}}$ are fuzzy averts for variable V and e.

Step 2: The centroid index of Vectorial Centroid, $R(\tilde{A}) = \sqrt{\tilde{x}^2 + \tilde{y}^2}$ is inserted into Bayesian Network formulation as

$$R(\tilde{A}) = \sqrt{\tilde{x}^2 + \tilde{y}^2} = \mu(\tilde{A}_i)$$

The computational process of fuzzy Bayesian Network using Vectorial Centroid is

$$P(\mu(\tilde{V}_1), \mu(\tilde{V}_2), ..., \mu(\tilde{V}_N) | \mu(\tilde{e}_i)) =$$

$$\frac{P(\mu(\tilde{e}_i)) P(\mu(\tilde{e}_i) | \mu(\tilde{V}_1), \mu(\tilde{V}_2), ..., \mu(\tilde{V}_N))}{\sum_{i=1}^{n} P(\mu(\tilde{e}_i)) P(\mu(\tilde{e}_i) | \mu(\tilde{V}_1), \mu(\tilde{V}_2), ..., \mu(\tilde{V}_N))} \quad (19)$$

4. Experimental Setting

In this section, we describe the required parameters to conduct the experiments. The experiment is conducted using 10-fold cross validation on BUPA liver-disorder data set from UCI machine learning repository is used where donated by BUPA Medical Research Ltd (Forsyth, 2015). This liver-disorder classification dataset has 345 examples, 7 attributes and binary classes for dependent attribute. The first 5 attributes measurements were taken by blood tests that are thought to be

sensitive to liver-disorders and might arise from excessive alcohol consumption. The sixth attribute is a sort of selector attribute where the subjects are single male individuals. The seventh attribute shows a selector on the dataset which being used to split into two categories that indicating the class identity. The attributes include:

    a.   Mean corpuscular volume,
    b.   Alkaline phosphatase,
    c.   Aspartate aminotransferase,
    d.   Gamma-glutamyl transpeptidase,
    e.   Alamine aminotransferase,
    f.   Number if half-pint equivalents of alcoholic beverage drunk per day, and
    g.   Output attributes either liver disorder or liver normal

Among all the people, there are 145 belonging to the liver-disorder group and 200 belonging to the liver-normal group. These attributes are selected with the aid of experts. In operating centroid methods, the original dataset are fuzzified randomly with range ±3 in generalised trapezoidal fuzzy number fuzzy set form. Below depicts the example of fuzzy sets are used in this research study

Example:

If the generalised trapezoidal fuzzy set $\tilde{\tilde{A}}_i = (90.09, 91.45, 93.32, 94.14; 0.9)$, then the centre points are computed using proposed (Vectorial Centroid) and established Shieh (2007) formulation respectively as follows:

Vectorial Centroid:

$VC(\tilde{x}) = 92.2650$ and $VC(\tilde{y}) = 0.35$

Centroid index Vertical Centroid, $VC(\tilde{R}) = 92.2657$

Shieh centroid:

$\tilde{x} = 92.2334$ and $\tilde{y} = 0.3948$

Crisp index Shieh centroid, $Shieh(\tilde{R}) = 92.2343$

5. Simulation Results

This section illustrates the validation process that are divided into two parts which are theoretically and empirically. Therefore, the theoretical of Vectorial Centroid validation process are as follow

A. Stage one

The relevant properties considered for justifying the applicability of centroid for fuzzy numbers, where they depend on the practically within the area of research however, they are not considered as complete. Therefore, with no loss of generality, the relevant properties of the centroid are as follow:

Let $\tilde{A}$ and $\tilde{B}$ are be trapezoidal and triangular fuzzy numbers respectively, while $VC_{\tilde{A}}(\tilde{x}, \tilde{y})$ and $VC_{\tilde{B}}(\tilde{x}, \tilde{y})$ be centroid for $\tilde{A}$ and $\tilde{B}$ respectively. Centroid index of Vectorial Centroid represents the crisp value of centroid point that is denoted as $R(\tilde{A}) = \sqrt{\tilde{x}^2 + \tilde{y}^2}$

Property 1: *If $\tilde{A}$ and $\tilde{B}$ are embedded and symmetry, then $R(\tilde{A}) > R(\tilde{B})$.*

Proof:
Since $\tilde{A}$ and $\tilde{B}$ are embedded and symmetry, hence we know that $\tilde{x}_{\tilde{A}} = \tilde{x}_{\tilde{B}}$ and $\tilde{y}_{\tilde{A}} > \tilde{y}_{\tilde{B}}$.
Then, from equation (18) we have $\sqrt{\tilde{x}_{\tilde{A}}^2 + \tilde{y}_{\tilde{A}}^2} > \sqrt{\tilde{x}_{\tilde{B}}^2 + \tilde{y}_{\tilde{B}}^2}$. Therefore, $R(\tilde{A}) > R(\tilde{B})$.

Property 2: *If $\tilde{A}$ and $\tilde{B}$ are embedded with $h_{\tilde{A}} > h_{\tilde{B}}$, then $R(\tilde{A}) > R(\tilde{B})$.*

Proof:
Since $\tilde{A}$ and $\tilde{B}$ are embedded and *with $h_{\tilde{A}} > h_{\tilde{B}}$*, hence we know that $\tilde{x}_{\tilde{A}} = \tilde{x}_{\tilde{B}}$ and $\tilde{y}_{\tilde{A}} > \tilde{y}_{\tilde{B}}$.
Then, from equation (18) we have $\sqrt{\tilde{x}_{\tilde{A}}^2 + \tilde{y}_{\tilde{A}}^2} > \sqrt{\tilde{x}_{\tilde{B}}^2 + \tilde{y}_{\tilde{B}}^2}$. Therefore, $R(\tilde{A}) > R(\tilde{B})$.

Property 3: *If $\tilde{A}$ is singleton fuzzy number, then $R(\tilde{A}) = \sqrt{\tilde{x}_{\tilde{A}}^2 + \tilde{y}_{\tilde{A}}^2}$.*

Proof:
For any crisp (real) numbers, we know that $a = b = c = d = \tilde{x}_{\tilde{A}}$ and $\tilde{y}_{\tilde{A}} < 1$ which are equivalent to equation (17). Therefore, $R(\tilde{A}) = \sqrt{\tilde{x}_{\tilde{A}}^2 + \tilde{y}_{\tilde{A}}^2}$.

Property 4: *If $\tilde{A}$ and $\tilde{B}$ are any symmetrical or asymmetrical fuzzy number, then $a < R(\tilde{A}) < d$ and $a < R(\tilde{B}) < d$.*

Proof:
Since $\tilde{A}$ and $\tilde{B}$ are any symmetrical or asymmetrical fuzzy numbers, hence $a < VC_{\tilde{A}}(\tilde{x}, \tilde{y}) < d$ and $a < VC_{\tilde{B}}(\tilde{x}, \tilde{y}) < d$. Therefore, $a < R(\tilde{A}) < d$ and $a < R(\tilde{B}) < d$ respectively.

All validation are related with computation for single crisp value $R(\tilde{A})$, where $\tilde{A}$ is any fuzzy number.

B. Stage two

In this stage, the empirical validation is implemented where the BUPA liver-disorder dataset is used. These attributes are selected with the aid of experts. The original dataset are fuzzified randomly with range ±3 as

mentioned before. The fuzzy data bases consist of all possible cases of fuzzy sets for each variable or attribute randomly which are:

a. 4 different cases of trapezoidal fuzzy number
b. 4 different cases of triangular fuzzy number
c. 2 different cases of singleton fuzzy number

Details on all these cases can be found in the Fig. 3 – Fig. 12.

Table I presents a comparative results between classical Bayesian Network (BN-Classic), fuzzy Bayesian Network using established Shieh centroid method (FBN-Shieh), and fuzzy Bayesian Network using Vectorial Centroid (FBN-VC). The comparison results are based on accuracy, precision, Kappa statistic and some error terms.

TABLE I.   ACCURACY, PRECISION, KAPPA STATISTIC AND ERRORS RESULTS

BN-Classic	FBN-Shieh	FBN-VC
Accuracy 56.2319%	*Accuracy* 56.8116%	*Accuracy* 57.3913%
Precision 0.535	*Precision* 0.54	*Precision* 0.552
Kappa Statistic 0.035	*Kappa Statistic* 0.0419	*Kappa Statistic* 0.0644
Errors MAE: 0.482 RMSE: 0.4995 RAE: 98.8992% RRSE: 101.187%	*Errors* MAE: 0.4804 RMSE: 0.4963 RAE: 98.575% RRSE: 100.539%	*Errors* MAE: 0.4759 RMSE: 0.4956 RAE: 97.6398% RRSE: 100.369%

The accuracy and precision of a measurement system play important role in quantifying the actual measure value. It is commonly used as metric for evaluation of machine learning systems. Accuracy refers to the closeness of agreement between a measured value and the true value. The precision is dependent of accuracy where the model can be very precise but inaccurate. The higher the value of accuracy and precision, the better classification prediction is made. In this research study, Table I shows the accuracy results that show the correctness of a model classifies the dataset in each class. The accuracy results of BN-Classic, FBN-Shieh and FBN-VC are 56.2319%, 56.8116% and 57.3913% respectively. It shows that the proposed methodology is significantly more accurate compared to others. The highest precision in this case study is FBN-VC with 0.552, followed by FBN-Shieh with 0.54 and BN-Classic with 0.535.

Kappa statistic technique is a chance-corrected tool that used to measure the agreement of two classifiers and estimate the probability of two classifiers agree simply by chance (Jeong, 2010). The higher the value of kappa statistic, the stronger the strength of agreement between two classifiers by chance. Referring Table 1, FBN-VC shows the highest value of kappa statistic with 0.0644 followed by FBN-Shieh and BN-Classic with 0.0419 and 0.035 respectively.

The last part in Table I depicts the errors for the experiment carried out. The errors are computed by using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE) and Root Relative Square Error (RRSE). The proposed method, FBN-VC performs better results in error terms where all of these errors are less than FBN-Shieh and BN-Classic.

6.   Conclusion

This study has addressed a generalised of the fuzzy Bayesian Network that takes into account the need of fuzzy events in variables or attributes in classification case study in which vagueness and ordinality are available. This work recommends new extension of fuzzy Bayesian Network methodology which consist of two stages which are: the development of Vectorial Centroid defuzzification method for fuzzy sets: and the implementation of Vectorial Centroid in Bayesian Network algorithm. For the first stage, the development of new centroid method can cater all the possible cases of fuzzy numbers precisely that matching for human intuition or human judgment. The generalised fuzzy Bayesian Network using proposed method on stage two is easily capable constructed and handled in data analysis when dealing with fuzzy data sets.

Shieh (2007) method shows some shortcomings where it can't compute for singleton cases. The proposed Vectorial Centroid method capable in handling all possible cases of fuzzy numbers and its looks more balance compare to Shieh (2007) method in discovering centre point where Vectorial Centroid has three multiple centre points, α, β, and γ that used to support more the shape in Cartesian plane before we find the core centre point. This proposed method can be used in many applications either in fuzzy problems or machine learning systems where it can cater all possible cases of fuzzy events.

Moreover, this study can be valuable alternatively in the set of existing Bayesian Network for many problems in machine learning problems. For the theoretical validation, there are four relevant properties for centroid development are constructed and well proved, where corresponding with all possible fuzzy numbers representation. Several tests for validation have been done and the results have been studied in-depth using BUPA liver-disorder classification dataset from UCI machine learning repository dataset. The validation results show the proposed research study more efficient and consistent in dealing with fuzzy events empirically. Finally, it can be concluded that the main focus of this research study can be proceeded in order to make some contributions by considering real case study drawn for diverse fields crossing ecology, health, genetics, finance and so forth. In the future, we are planning to focus on generalisation of the proposed Vectorial Centroid method for type-2 fuzzy sets and Z-numbers under the uncertainty and reliability conditions.

References

[1] Y. Tang, W. Pan, H. Li, and Y. Xu, Fuzzy Naive Bayes Classifier Based on Fuzzy Clustering. *Pro-*

ceeding if the Systems, Man and Cybernatics 2002 IEEE International Conference on (Volume: 5).

[2] H. Tang and S. Liu, Basic Theory of Fuzzy Bayesian Networks and Its Application in Machinery Fault Diagnosis, proceedings of the 4th *International Conference on Fuzzy Systems and Knowledge Discovery* (FSKD 2007), IEEE Computer Society, 2007.

[3] F. V. Jensen and T. D. Nielsen, Bayesian Networks and Decision Graphs, Springer, New York, 2007.

[4] J. Pearl, Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kauffman, San Mateo, CA, 1998.

[5] M. Gasse, A. Aussem and H. Elghazel, A hybrid algorithm for Bayesian network structure learning with application to multi-label learning, *Expert Systems with Applications*, 41, pages 6755-6772, 2014.

[6] A. Bouejla, X. Chaze and F Guarnieri and A. Napoli, A Bayesian network to manage risk of maritime piracy against offshore oil fields, *Safety Science*, 68, pages 222-230, 2014.

[7] P. A. Aguilera, A. Fernandez, F. Reche and R. Rumi, Hybrid Bayesian network classifiers: applications to species distribution models, *Environmental modelling & Software* 25, pages 1630-1639, 2010.

[8] L. Uusitalo, Advantages and challenges of Bayesian networks in environmental modelling, *Ecological Modelling* 203, pages 312-318, 2007.

[9] J. B. Nyberg, B. G. Marcot, R. Sulyma, Using Bayesian belief networks in adaptive management, *Canadian Journal of Forest Research* 36, pages 3104-3116, 2006.

[10] L. Getoor, J. T. Rhee, D. Koller and P. Small, Understanding tuberculosis epidemiology using structured statistical models, *Artificial Intelligent in Medicine* 30, pages 233-256, 2004.

[11] L. A. Zadeh, Fuzzy sets, Inf. Control 8, pp. 338-353, 1965.

[12] H. Deng, Comparing and ranking fuzzy numbers using ideal solutions. *Applied Mathematical Modelling*, in press, 2013.

[13] R. R. Yager, and D.P. Filev, Essential of fuzzy Modelling and Control. Wiley, New York, 1994.

[14] H-J. Zimmermann, An application – oriented view of modelling uncertainty. *European Journal of Operational Research 122,* pp. 190 – 198, 2000.

[15] S. Murakami, S. Maeda and S. Imamura, Fuzzy decision analysis on the development of the centralized regional energy control; system, paper presented at the IFAC Symp on Fuzzy Inform. Knowledge Representation and Decision Analysis, 1983.

[16] C. H. Cheng, A new approach for ranking fuzzy numbers by distance method. *Fuzzy Sets and System*, 95, pages 305-317, 1998.

[17] C. T. Chu and C. T. Tsao, Ranking fuzzy numbers with an area between the centroids point and original point, *Computer and Mathematics with Applications* 43, pages 111-117, 2002.

[18] S. J. Chen and S. M. Chen, A new method for handling multicriteria fuzzy decision making problems using FN-IOWA operators. *Cybernatics and Systems*, 34, pages 109-137, 2003.

[19] C. Liang, J. Wu and J. Zhang, Ranking indices and rules for fuzzy numbers based on gravity center point, paper presented at the 6th *World Congress on Intelligent Control and Automation*, Dalian, China, 2006.

[20] S. J. Chen, and S. M. Chen, Fuzzy risk analysis based on the ranking of generalized trapezoidal fuzzy numbers. *Applied Intelligence*, 26(1), pages 1-11, 2007.

[21] W. J. Wang and H. S. Lee, The revised method of ranking fuzzy numbers with an area between the centroid and original points. *Computers and Mathematics with Applications*, 55, pages 2033-2042, 2008.

[22] N. Ramli and D. Muhamad, A comparative analysis of centroid methods in ranking fuzzy numbers. *European Journal of Scientific Research*. ISSN 1450-216X Vol.28 No.3, pages 492-501, 2009.

[23] B. S. Shieh, An approach to centroids of fuzzy numbers .*International Journal of Fuzzy Systems 9,* pages 51-54, 2007.

[24] A. Kaufmann and M. M. Gupta, Fuzzy mathematical models in engineering and management science, *Elsevier Science Publishers*, Amsterdam, Netherlands, 1988.

[25] S. M. Chen and J. H. Chen, Fuzzy Risk Analysis Based on Ranking Generalized Fuzzy Numbers with Different Heights and Different Spreads, *Expert Systems with Applications,* 36, pages 6833--6842, 2009.

[26] G. Liu, Probability Theory and Applied Mathematical, Higher Education Press, Beijing, 2000.

[27] R. S. Forsyth,"Liver disorder data set," UCI Machine Learning Repository. Internet: https://archive.ics.uci.edu/ml/datasets/Liver+Disorders, 2015.

[28] D. Jeong, D. Kang and S. Won, "Feature selection for steel defects classification," *in proceedings of International Conference on Control, Automation and Systems,* 2010, Gyeonggi, Korea.
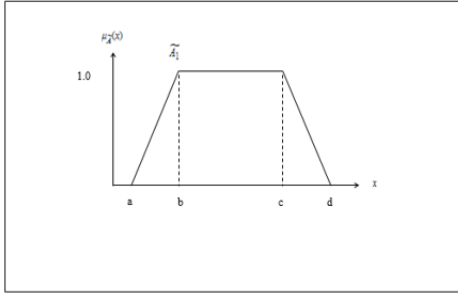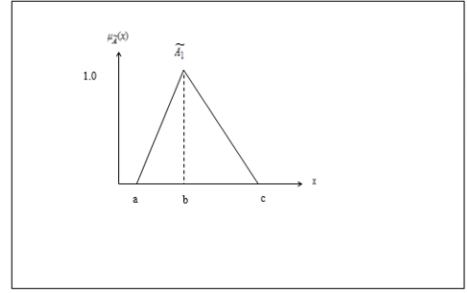
Fig. 3: Trapezoidal Symmetry Normal

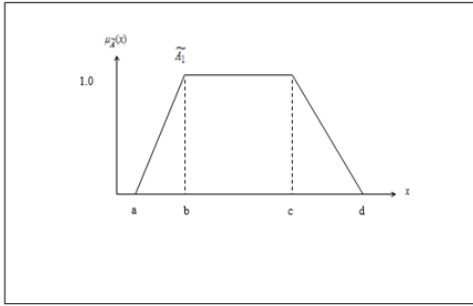Fig. 4: Trapezoidal Non-Symmetry Normal

Fig. 5: Trapezoidal Symmetry Non-Normal

Fig. 6: Trapezoidal Non-Symmetry Non-Normal
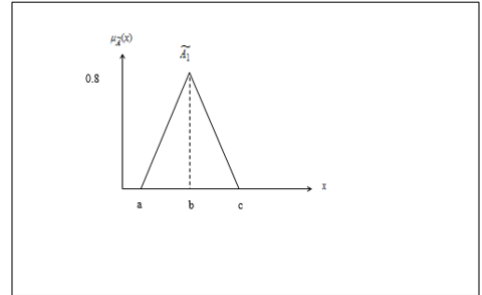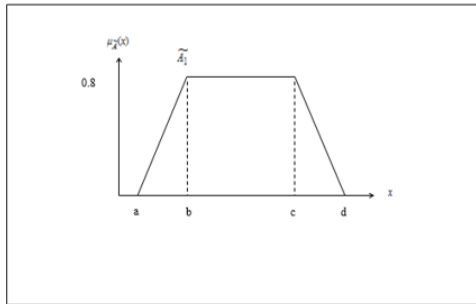

Fig. 7: Triangular Symmetry Normal
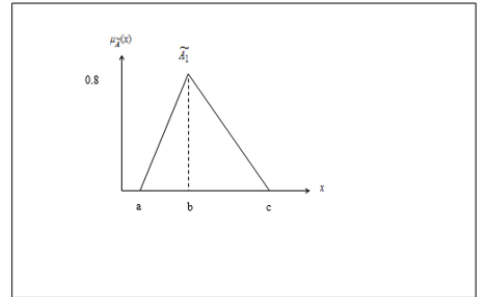

Fig. 8: Triangular Non-Symmetry Normal

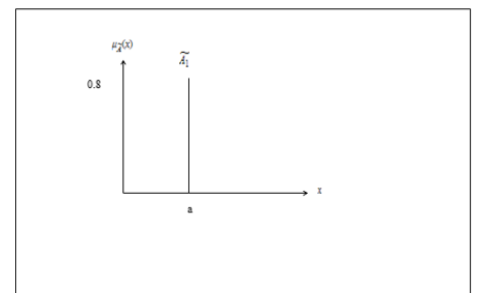Fig. 9: Triangular Symmetry Non-Normal

Fig. 10: Triangular Non-Symmetry Non-Normal
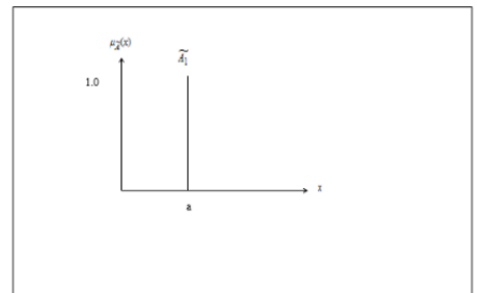
Fig. 11: Singleton Normal

Fig. 12: Singleton Non-Normal