Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 11, November 2013, pg.70 – 78

RESEARCH ARTICLE

A HYBRID APPROACH FOR HANDLING UNCERTAINTY - PROBABILISTIC THEORY, CERTAINTY FACTOR AND FUZZY LOGIC

Sandhya Maitra[1], Nidhi Gupta[2]

[1]Associate Professor - Department of Computer Applications and Institute of Information Technology and Management, India

[2]Student - Department of Computer Applications and Institute of Information Technology and Management, India

[1] msan324@gmail.com; [2] nidhi_6690@yahoo.co.in

Abstract— Real world is actually revolving around data i.e. data plays a very important role in the current information era. Different types of uncertainty are addressed in different forms of data. Till date, probabilistic theory, fuzzy logic, certainty factor was developed to handle uncertainty. All these approaches were quite successful in handling uncertainty but there are some situations where when these methods taken individually, failed to handle uncertainty. So there was a need to develop a hybrid approach which will handle uncertainty to a high level.

In this paper, we present an approach wherein we integrate probabilistic theory, certainty factor and fuzzy logic concepts. Once we use all these approaches together, uncertainty model is developed which will address the various limitations inherent in these approaches when applied individually.

Keywords—Certainty Factor; Probability Theory; Fuzzy Logic; Uncertainty; Transitive Dependency; Baye's Theorem

I. INTRODUCTION

In the present scenario, each research deals with data in a different context. Data actually refers to some raw facts or figures. But these days a very important issue that has been inherited in data is the uncertainty. Data uncertainty is usually found in all applications.

Many researchers have been trying to find out the reasons of uncertainty, the type of uncertainty, sources of uncertainty and last but not the least how to deal with uncertainty. Many approaches have been developed for dealing with uncertainty like fuzzy sets theory, rough set theory, data mining algorithms, probability functions, statistical functions, certainty factor, fuzzy logic, etc but all these approach have some advantages and disadvantages. These approaches are somehow valid for some specific purpose only i.e. a single technique is only applicable to a particular problem. Therefore, it is essential to develop a hybrid approach that is composed of a combination of two or more approaches. When various methods are integrated, the disadvantages of these approaches when used individually are avoided.

The paper is organized into various sections:

Section II deals with uncertainty and Section III discuss some approaches to handle uncertainty proposed till now. Section IV explains about probability theory whereas Section V gives an overview about certainty factor approach. Section VI deal with transitive dependencies. Section VII describes the hybrid approach presented to handle uncertainty. Conclusions and future scope are drawn in Section VIII.

II. UNCERTAINTY

Uncertainty is usually associated with data i.e. uncertainty tends to revolve around real world data. It is said that the data in the real world is uncertain whereas real world is certain. Uncertainty is the property or features about the data including error, accuracy, validity, quality and noise.

There are some definitions of uncertainty given by different people in different context:

i. Klir and Wierman (1999) states that uncertainty itself has many forms and dimensions and may include concepts such as fuzziness or vagueness, disagreement and conflict, imprecision and non-specificity [1].
ii. Taylor and Kuyatt (1993) formulated guidelines for expressing the uncertainty of measurements [2].

Measurement uncertainty, although consisting of several components, can be broadly classified into categories according to the method used to estimate its numerical values:

- Evaluation by statistical methods such as standard deviation and least squares techniques and
- Evaluation by scientific judgment such as including previous measurements, manufacture's specification and experience.

1) Sources of Uncertainty

There are several sources of uncertainties i.e. uncertainty can arise from various sources such as:

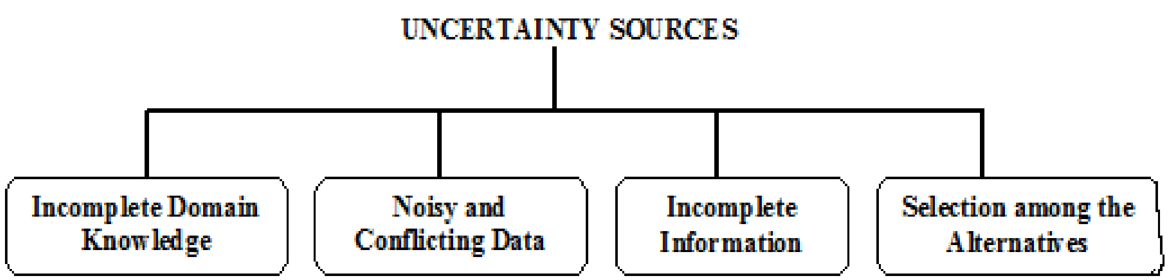

Fig 1: Sources of Uncertainty

1.1) Incomplete Domain Knowledge[3]

Incomplete Domain Knowledge refers to the knowledge which is not complete in a particular domain. The result of incomplete domain knowledge may or may not always results in accurate results.

Example:
A doctor must have the complete knowledge about a patient's disease, its history, period of suffering or about symptoms of disease etc. if he has incomplete domain knowledge, than he might not be able to treat him accurately.

1.2) Noisy and Conflicting Data[3]

Real world data is dirty. When data is collected from different sources, then there are many chances that the data is noisy and conflicting. Noisy data usually refers to the meaningless data whereas conflicting data refers to the contradictory data.

Example:
If a coin is flipped, then output cannot be a number. It has to be either head or tail (Noisy Data) and the salary of a person is 10000 and it lies between 7000-9000 (Conflicting Data).

1.3) Incomplete Information[3]

Incomplete Information is partial or insufficient information i.e. the information that does not have all the necessary information. Information may become incomplete due to the changes in factors with time.

Example:
Software cannot be build properly if the user will not provide complete information about the functionality that the software is supposed to do.

1.4) Selection among the Alternatives[3]
Alternatives mean choices or options that are considered in a situation. There are many alternatives to a problem and there may be a disagreement in choosing the best alternative among all. It is always not necessary that the best solutions comes out for a solving a problem.

Example:
There are various methods to handle uncertainty. So choosing the best way to handle it requires a complete understanding about the problem as well as the methods and then applying the appropriate one.

2) Types of Uncertainty
There are several types of uncertainties that can arise in data. Basically there are four types of uncertainty:

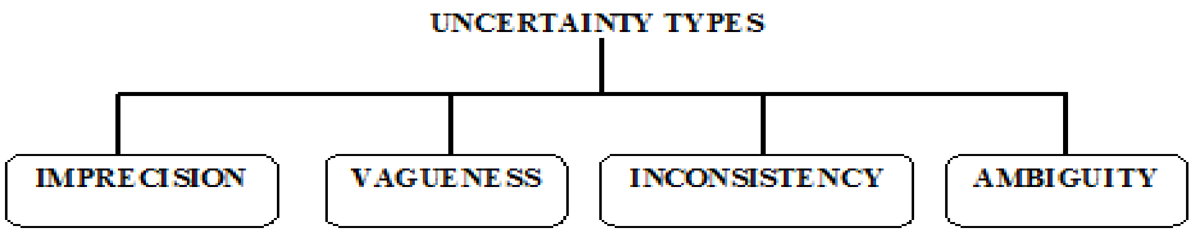

Fig 2: Types of Uncertainty

2.1) Imprecision
The information available in the model is not as specific as it should be [4].

Example:
The salary of a person lies in between 10000-15000 (uncertainty in range) or the salary of a person is not 5000 (negative uncertainty)

2.2) Vagueness
The model includes that are inherently vague [4].

Example:
Raman is in early middle age.

2.3) Inconsistence
The model contains two or more assertions that cannot be true at the same time [4].

Example:
The salary of a person lies in between 10000-15000 and the salary of that person is 20000.

2.4) Ambiguity
Some elements of model lack complete semantics, leading to several possible interpretations [4].

Example:
It is not clear whether the 10:15 means 10:15 AM or 10:15 PM.

III. SOME APPROACHES TO HANDLE UNCERTAINTY

This sections deal with some approaches developed till today to handle uncertainty in data:

1) Fuzzy set theory

The concept of fuzzy sets was given in 1965 by Lotfi A. Zadeh and Dieter Klaua. Initially classical set theory was introduced which tells that the membership of elements in a set is accessed in terms of binary i.e. an element either belongs to the set or not. But fuzzy set theory extended this concept. Fuzzy sets introduced the membership function and the membership of elements in a set lies in the interval [0, 1]. Therefore fuzzy sets can be defined as the sets that contain elements that have degree of membership. Fuzzy sets can be used in different areas for handling uncertainty. Fuzzy set theory permits the assessment of the membership of the element in a set.

1.1) Definition of Fuzzy Set

Let U denote a universal set. A fuzzy set is defined as:

$m:U\rightarrow[0,1]$

where [0,1] denotes the interval of real no from 0 to 1 and m is the membership function.

For each $x \in U$, the value m(x) is called the grade of the membership of x in (U,m).

There are two types of fuzzy sets – type-1 fuzzy set and Interval type-2 fuzzy set. In type-1 fuzzy sets, the degree of membership is specified by a crisp number belonging to the interval [0, 1], in type-2 fuzzy sets, the degree of membership is in itself fuzzy and is represented by what is usually referred to as a secondary membership function [5].

2) Probabilistic theory

Probability theory is a branch of mathematics concerned with the analysis of random phenomena. The outcome of a random event cannot be determined before it occurs, but it may be any one of several possible outcomes. The actual outcome is considered to be determined by chance. There are many probability theory formulas like conditional probability, conditional independence, Baye's theorem, etc that can be used for analysing uncertainty in data.

3) Statistical theory

Statistics is the branch of mathematics that was firstly used in 1719-1772 by Gottfried Achenwall. It is a process of collecting the data, analysing it and then summarizing it. There are various terms used in statistics like arithmetic mean, median, standard deviation, co-variance, correlation, regression, skewness, kurtosis, etc that can be used to handle the uncertainty. All these terms have specific meaning in the statistics literature. The techniques or the methods for statistics provide the basis for statistics theory. This theory helps to find the best and accurate approach for handling uncertainty.

4) Rough set theory

The rough set theory was proposed by Zdzisław I. Pawlak. It is is a formal approximation of a crisp set in terms of a pair of sets which give the lower and the upper approximation of the original set. This approach is used to handle various types of uncertainty like vagueness. This method is less commonly used as compared to other methods like probability, statistics, etc but the only advantage of this method is that it does not require any additional information.

5) Fuzzy Logic

The concept of fuzzy logic was given by Lotfi Zadeh, is based on fuzzy sets. FL is a problem-solving control system methodology that lends itself to implementation in systems ranging from simple, small, embedded micro-controllers to large, networked, multi-channel PC or workstation-based data acquisition and control systems [6]. It actually focuses more on the degree to which an event is true. Fuzzy logic is often used to handle uncertainty arising from vagueness and imprecision.

IV. PROBABILITY THEORY

The term probability is most frequently used keyword. It has different meaning in different context. One interpretation of probability is that it tells the occurrence of an event. Probability is often used to handle uncertainty arising from incomplete information. Probability lies in between 0 and 1. The formula for probability is:

$$P(A) = \frac{\text{total number of favorable outcomes}}{\text{total number of possible outcomes}}$$

where ‘A’ refers to an event and ‘P’ refers to the probability of that event.

Example:
When a die is tossed, the probability denotes the occurrence of 1, 2,3,4,5 and 6.
P (getting an even number) = 3/6 = 0.5

1) Conditional Probability

Conditional probability means that an event will occur when another event is known to occur or has occurred. Conditional probability is expressed as:

$$P(A|B) = P(A \cap B) / P(B)$$

where ‘A’ and ‘B’ refers to the event and ‘P’ refers to the probability of an event.

2) Bayes' Theorem or Bayes' Rule

Baye’s Theorem is based on conditional probability. It is used to recognize the dealing of sequential events. It is expressed as:

$$P(A|B) = P(B|A).\ P(A) / P(B)$$

where ‘A’ and ‘B’ refers to the event and ‘P’ refers to the probability of an event.

3) Analysis of probability theory

i. It is a formal foundation and well defined semantics for decision making [3].
ii. Probability can be assigned even if no information is possible.
iii. It describes the frequency of occurrence of reproducible events [3].
iv. It is used where statistical data is available.
v. It may be inaccurate as data may or may not be reliable.

V. CERTAINTY FACTOR

Certainty factor is a number that reflects the beliefs in a hypothesis given available information [3]. This number ranges from -1 to +1.Certainty is often used to handle uncertainty arising from judgments.

Certainty factor denotes the beliefs in a hypothesis H given that some pieces of evidence E are observed [3]. There are three cases:

i. When CF = +1, it means hypotheses H is fully believed
ii. When CF = 0, it means hypotheses H is not known
iii. When CF = -1, it means hypotheses H is not believed

1) Analysis of probability theory

i. CF represents the amount of support for a particular fact based on some other facts [3].
ii. It is used where judgments are used.
iii. CF can be used where there are complex relationships between data.
iv. When probability cannot be used, CF is one of the most important methods used.
v. It is based on availability of prior information.

VI. THE ADOPTION OF CONCEPT OF TRANSITIVE DEPENDENCY TO UNCERTAINTY IN FACTS

A transitive dependency is a functional dependency which holds by virtue of transitivity. A transitive dependency can occur in a relation that has three or more attributes.

Let A, B, and C designates three distinct attributes (or distinct collections of attributes) in the relation. Suppose all three of the following conditions hold:

i. $A \rightarrow B$
ii. It is not the case that $B \rightarrow A$
iii. $B \rightarrow C$

Then the functional dependency $A \rightarrow C$ is known as transitive dependency.

In database normalization, one of the important features of third normal form is that it excludes certain types of transitive dependencies. E.F. Codd, the inventor of the relational model, introduced the concepts of transitive dependence and third normal form in 1971[7].

We shall adopt the concept to handle uncertainty where several transitive relationships exist on evidences supporting a fact [3]. When decisions are based on evidences and interpretation of evidences is subjective. The main idea here is to minimize amount of computation required to update the probability, as evidences are accumulated incrementally [3]. We propose to suggest a procedure where dependent evidences supporting a fact are investigated for transitivity which may result in incremental accumulation of evidences. Such cases suggest usage of certainty factor approach rather than Bayesian or other approaches which may prove to be computationally prohibitive.

VII. COMBINED APPROACH OF PROBABILITY THEORY, CERTAINTY FACTOR AND FUZZY LOGIC

When data is collected, first and most important step is to collect the evidences that support that data. Once evidences are accumulated, the dependency between facts is determined. The dependency will help in judging whether to apply Baye's Theorem or Certainty Factor to determine occurrence of event.

1) Definition:

Say a□ is data and f□ is the fact supporting data. Say a□ has n facts supporting it and there exists transitive dependencies of facts i.e. f□ →f□ →f□ for some u, v and x,
then use Certainty Factor
else
use Baye's Theorem.

Fuzzy logic is used to determine the degree of truth of event i.e. the degree to which event is true. Fuzzy logic theory is based on the computation with fuzzy sets which is an extension of classical sets [5]. Even though fuzzy logic was successful in modeling uncertainty in many situations but the most important issue raised was that fuzzy logic does not focuses on the occurrence of event. The occurrence of event can be very well determined by the probabilistic theory.

As mentioned above, various methods have been developed to model uncertainty. Each method has its applications in some or the other real world data. Therefore this paper presented a hybrid approach wherein the uncertainty is determined by first checking the input data for transitivity. As per the result got from this, probability formulas are used to determine the occurrence of event and once the occurrence of event exists, uncertainty can be handled using fuzzy logic system [8].

The basic structure for handling uncertainty proposed in this paper is:

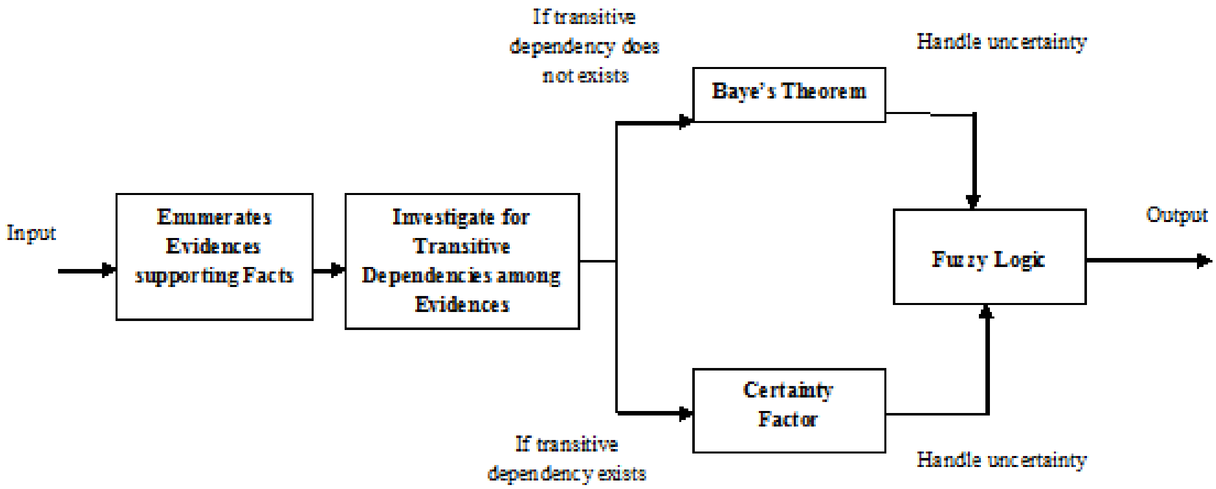

Fig 3: Basic Structure for Handling Uncertainty

Basic process of fuzzy logic includes that input data is taken and then input data is then fuzzified using membership functions. The output of this step is the input fuzzy sets. Then some inferences are made using some sets of rules. The output of this step is the output fuzzy sets. Then these output sets are de-fuzzified using membership function to produce the output data.

The flowchart for handling uncertainty is depicted below. The basic steps for handling uncertainty using probability theory, certainty factor and fuzzy logic [8] are:

i. Inputs are taken first i.e. the data or the facts.
ii. Enumerate the evidences that support these facts.
iii. Check for transitive dependencies among the evidences.
iv. If transitive dependency exists, determine occurrence of an event by using Certainty Factor, else the occurrence of an event is determined by using Baye's Theorem.
v. Once the occurrences of events are determined according to the given condition in the flowchart, the uncertainty is handled using fuzzy logic [8].
vi. The final outputs are the produced.

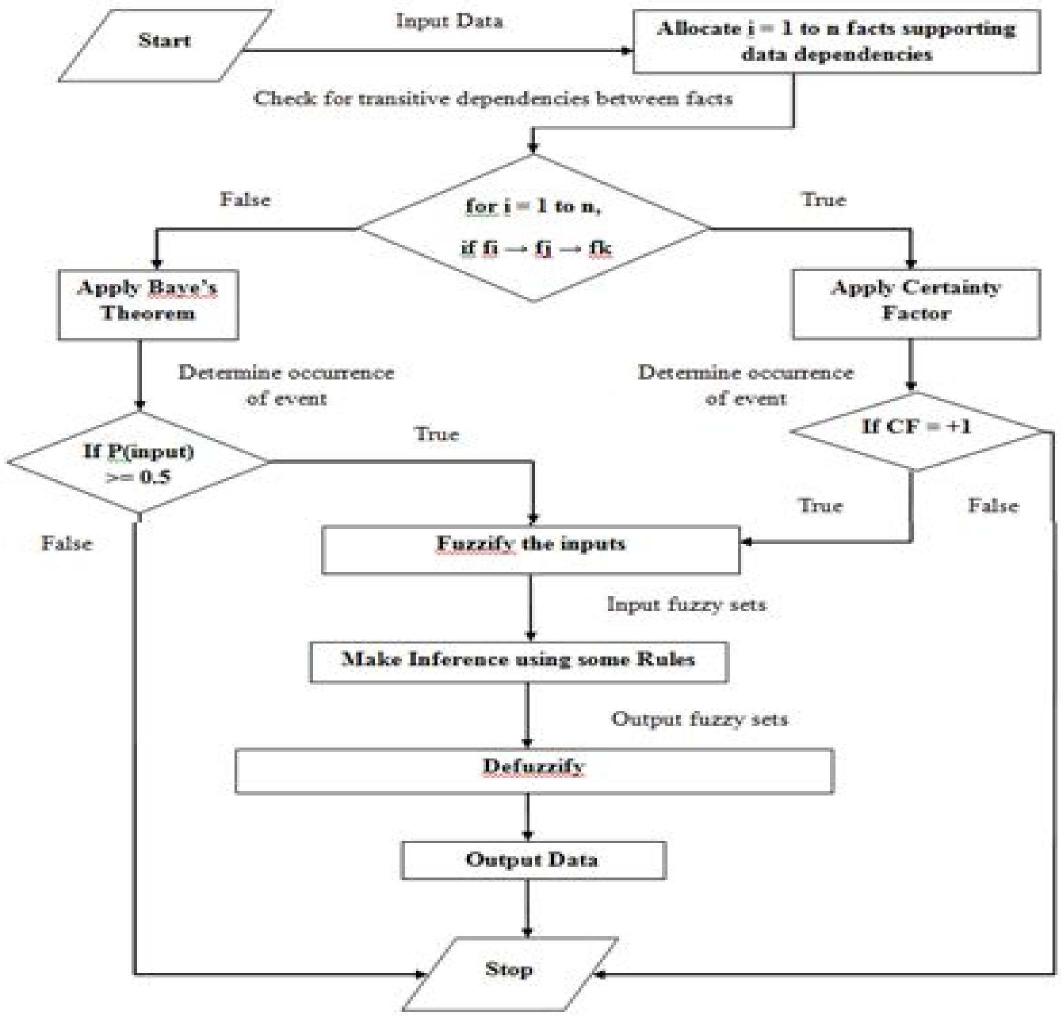

Fig 4: Flowchart for Handling Uncertainty

VIII. CONCLUSION AND FUTURE SCOPE

There are various types and sources of uncertainties and it can arise in any form in any type of data. This paper combines the probability theory, certainty factor and fuzzy logic to handle it.

Through this paper, a hybrid approach has been presented to handle uncertainty in the data.
The certainty factor approach eliminates the disadvantages associated with Bayesian approach where evidences supporting a fact can get accumulated incrementally. But same approach on the other hand cannot be used if there are complex relationships between the evidences supporting facts. As far as possible when relationships between evidences limit to simple dependencies, it is effective to use Bayesian Theorem as:

i. Combination of non-independent evidences is unsatisfactory [3].
ii. New knowledge may require changes in the certainty factor of existing knowledge [3].

But if the above statement does not hold i.e. if complex relationships exist between evidences supporting data, then it is efficient to use Certainty Factor.

The idea may be extended to include other approaches such as fuzzy sets, rough set theory, statistical measures, etc. There are numerous ways to deal with uncertainty these days. The approach proposed in this paper can be taken for further study and research. Each approach has its limitations. So the presented approach can be refined if possible by integrating with other approaches used for handling uncertainty.

REFERENCES

[1] Klir, G. and M. Wierman. Uncertainty-Based Information: Elements of Generalized Information Theory, 2nd edition. Physica-Verlag. 168pp, 1999.

[2] Taylor, B. N. and C. E. Kuyatt. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. Technical report, National Institute of Standards and Technology Technical Note 1297, Gaithersburg, MD, 1993.

[3] Ashutosh Dwivedi, D Mishra and Prem k Kalra. Handling Uncertainties – using Probability Theory to Probabilistic Theory, 2006.

[4] AMIHAI MOTRO. 22 Management of Uncertainty in Database Systems., 1994.

[5] Hani Hagras and Christian Wagner. Introduction to Interval Type-2 Fuzzy Logic Controllers -Towards Better Uncertainty Handling in Real World Applications, June 2009.

[6] Steven D. Kaehler. Fuzzy Logic Tutorial - **Fuzzy Logic - An Introduction Part 1,** 1998

[7] ^ Codd, E.F. "Further Normalization of the Data Base Relational Model." (Presented at Courant Computer Science Symposia Series 6, "Data Base Systems," New York City, May 24th-25th, 1971.) IBM Research Report RJ909 (August 31st, 1971). Republished in Randall J. Rustin (ed.), Data Base Systems: Courant Computer Science Symposia Series 6. Prentice-Hall, 1972. See pages 45-51, which cover third normal form and transitive dependence.

[8] J. Mendel. Fuzzy logic systems for engineering: a tutorial. Proceedings of the IEEE, 83(3):345{377, (Mar 1995).

