Handling of Uncertainty – A Survey

Lakshmi.S.Dutt*, Mathew Kurian**

* Department of Computer Science and Engineering, Karunya University, India
** Department of Computer Science and Engineering, Karunya University, India

Abstract- Uncertainty is a state of lack of certainty, where having incomplete knowledge can make it impossible to describe the outcome. There are many challenges that occur due to uncertainty. Uncertainty in events can cause losses of data, providing inaccurate data, noisy output. Thus the need to handle uncertainty is essential. This paper represents an outline of various methods through which the uncertainty can be analyzed and handled.

Index Terms- Bayesian Network, fuzzy logic, probability, uncertainty

I. INTRODUCTION

Uncertainty the term covers a lot of concepts. Uncertainty can be due to lack of knowledge or insufficient information. This can be due to vagueness, no specificity and conflict in the information. It can be defined as a situation n where the information available to the decision makers is imprecise to be summarized by a probabilistic measure. In truth scientific literature have lots of definitions for uncertainty. According to [2], [6] uncertainty can be classified into four classes namely: epistemic, linguistic, ambiguity and variability. Epistemic uncertainty is uncertainty due to lack of knowledge[14].It is also referred to as state of subjective uncertainty, reducible uncertainty, Type B or knowledge uncertainty means that where uncertainty can be reduced through more relevant data. This epistemic uncertainty can be categorized as: the basis of the model that is the description of the system, completeness, systematic error, subjective uncertainty and measurement uncertainty. The epistemic uncertainty can be represented in many ways including probabilistic theory, fuzzy set, possibility theory, etc. But the major challenge in it is the selection of appropriate mathematical structure for the representation. Linguistic uncertainty produced by statements in natural language [3]. Linguistic uncertainty is important because it pervades in workshop, other face-to-face language based methods on which risk assessment relies on to measure and communicate risk. Linguistic uncertainty can be classified into five distinct types: vagueness, ambiguity, context dependence, under specificity and indeterminacy of theoretical terms. This may be accidental or deliberate which create problems for risk assessment. Ambiguity can be defined as uncertainty about the probability which is created by lack of information that is relevant and could be known. Ambiguity uncertainty is applied to the situation where the probabilities of the outcome are unknown. Ambiguity uncertainty arises from the concept that a word can have more than one meaning. The ambiguous events are having more uncertainty associated with it and are confused with vagueness. Variability as name specifies is due to variations or differences in a process or quantity. Assigning an exact value for a quantity is difficult because it depends on many parameters. The variability can be categorized as: natural variation which occurs naturally and inherent randomness are repeated processes and are irregular by nature [2]. Sometimes it can be reduced but cannot be eliminated completely.

Uncertainty analysis is a process that measures, recognizes, identifies and minimizes the all types of uncertainty in a risk estimates and separates this uncertainty among the risk factors that contributes to risk estimates. The uncertainty analysis includes many statistical problems [14] such as: uncertainty factor, decision making with uncertain information, estimation of uncertainty in complex models of risk, structural uncertainty and model specification and monitoring methods to reduce uncertainty. Quantitative approaches to measure uncertainty vary with complexity of the problem and the method to reduce risk.

II. UNCERTAINTY ANALYSIS TECHNIQUES

Various techniques are available for analyzing uncertainty. This section will discuss some of the uncertainty analysis techniques.

A. Probabilistic Analysis

Probability is the study of uncertain events or outcomes. Probability analysis is a method of uncertainty propagation for making qualitative and quantitative calculations in the face of uncertainties of various kinds. A probability distribution for the individual outcomes are defined, the probability of any event is defined on the sample space as the sum of the probabilities assigned to the outcomes in that event. Here the uncertainties are characterized by probabilities associated with the events. In probabilistic analysis the probability of the event is interrupted by the frequency of the occurrences of the event.

[13] Proposses a probabilistic approach to high-level event extraction from RFID data. Exploitation of RFID data presents significant challenges. RFID readers produce streams of low-level events; these must be transformed into high-level events. But these transformation is challenging due to two main issues i) reliability ii) ambiguity. These issues make deterministic event detection

unworkable and lead event recall near zero. A probabilistic model is enabled here complex event extraction to handle the uncertainty involved. Also includes techniques to extract events and to handle reliability and ambiguity.

[12] Proposes a frame work for knowledge representation and reasoning enabling the event interface i.e., the events to which the system must respond must be derived from other events that are based on complex temporal predicates, but are not generated by monitoring tools. Based on probability theory, the representation of the associated uncertainty is defined. In addition, the probability space also defined, and shows how the relevant probabilities can be calculated. Here defines the representation and semantics of event composition for probabilistic settings, and shows how to apply the extensions to the quantification of the occurrence probability of events. Thus these results enable any kind of active system to handle the uncertainty. Also defines a class of specification languages that make it possible to represent both types of uncertainty, using probability as the uncertainty representation mechanism. In addition, the general principles underlying a reasoning mechanism for inferring the occurrence probabilities of events also detailed. To illustrate these principles, a specific inference algorithm for a specific language is detailed.

B. Fuzzy Analysis

Fuzzy logic theory provides a very useful solution to understanding, quantifying and handling vague, ambiguous and uncertain data. Fuzzy sets are mainly used to handle and analyze uncertainty of data, environmental data and fuzzy logic to handle inaccurate reasoning in knowledge-based models of ecological processes. The fuzzy decision support systems aims at making imprecise reasoning to have rational decisions in an environment of uncertainty and imprecision. A decision support system consists of analytical tools and information management components for decision makers to solve relatively large and unstructured problems. This system can be termed as a multi-criteria analysis system which implies satisfying a set of criteria or constraints. A Fuzzy Logic System can deal with vagueness and uncertainty residing in the human knowledge base, and allows us to represent linguistic terms. Fuzzy rules with fuzzy sets have been used to express fuzzy knowledge. Moreover, the relative importance of criteria that may not equally influence a decision can also be considered by the Fuzzy Logic System.

[9] Proposes that the modeling of the real world complex systems has been always a challenging problem. Complexity and uncertainty are both features of the real world systems which are treated differently in different modeling approaches. A deterministic model of the system presents a deterministic mapping between inputs and outputs while probabilistic models are used for modeling of stochastic systems and can be characterized by the statistical properties of some random processes in the system. A probabilistic model in which probability is used as a way of representing statistical uncertainties in complex systems and Fuzzy logic is a way of representing non-statistical uncertainty. Fuzzy logic and probability theory are both powerful tools for handling uncertainty. However, they are complementary rather than competitive. The former represent non-statistical uncertainty while the later represent statistical uncertainty. This is while both of the above uncertainties can exist at the same time in real world systems. Consequently, probabilistic fuzzy logic [5] has been introduced as a general framework for combination of these theories and hence better handling such cases. A probabilistic fuzzy system is hence a generalized fuzzy system which utilizes the probabilistic fuzzy logic in presenting a new concept of probabilistic fuzzy rule. A general mathematical framework named DSU has been developed here for representing probabilistic fuzzy systems.

[11] Proposes a reliable human activity recognition methodology which handles the structural variations of an activity. When a new observation containing an execution of an activity is provided, the system measures how semantically similar a given observation is to the optimal structure of the activity. This similarity measure is not deterministic but is designed to consider uncertainties of the activities structures. A fuzzy interval describes the possible range of its starting time and that of its ending time as well as the confidence value associated with time frames within the ranges. Once fuzzy intervals are calculated, a dynamic programming algorithm that we have designed and presented here is applied to measure the similarity between the detected fuzzy intervals and the structure of the activity specified in the representation. The algorithm searches for the time points in ranges that satisfy the temporal structure specified in the activity representation while maximizing the fuzzy membership values.

C. Bayesian Analysis

In short Bayesian Networks, also known as belief networks or Bayes nets and it belong to the family of probabilistic Graphical Models. These models are used to represent knowledge about an uncertain domain. Bayesian Networks cross the divide between qualitative models, mechanistic models and statistical models, and are one of the few methods that can perform forward uncertainty propagation with little or no data, and statistical inference when data is available. Each node in the graph represents a variable and the edges represent the dependencies among the variables. These dependencies are evaluated using statistical and computational methods. Bayesian networks always forms acyclic graph. Bayesian Networks provide a transparent, mathematically a logic way to express one's belief in a conceptual model, and the conditional probability of events, in a manner that can be updated as data are gathered during the monitoring and validation stages of an evaluation. The technique can be used in case where data is missing to understand

ISSN 2250-3153

various problem domains and to predict the future. Bayesian Networks have much to over as a risk assessment tool and have been identified as a practical and scientific approach to modeling complex systems in the presence of “high uncertainty”.

[4] Proposes an industrial application of Bayesian Network for constructing an expert system for gas-liquid contacting device selection. Bayesian Networks support representation and reasoning under uncertainty. Bayesian Networks are useful in the study problem domain since they can accommodate the large number of contacting devices available and the uncertain nature of the attributes involved in the selection task. The directed acyclic graph component of Bayesian Networks is also useful since it shows the causal relationships between the attributes in the domain. Given data or a method to obtain the conditional probabilities of the attributes, Bayesian Networks are an effective uncertainty modeling technique. However one of the disadvantages of this technique is that, they are not always practical when the domain is large. To support more easy interaction with the user, an expert system using Bayesian functions is currently being implemented.

D. Soft Computing Technique

Soft computing techniques have become increasingly visible in finance applications. Soft computing is a pool of techniques that includes fuzzy logic, neural networks, and probabilistic reasoning. The basic aim of soft computing is to trade precision for a reduction in solution cost and tractability by exploiting the tolerance for imprecision and uncertainty.

[7] Proposes a system to focus on the issue of learning representations and adaptive learning algorithms through the combination of fuzzy, neural, and evolutionary algorithm techniques. To solve the structural problems, here propose the use of hybrid adaptive learning. Their flexible properties increase the range of their applicability and efficiency relative to simple learning algorithms. The system is based on neuro-fuzzy and genetic algorithm techniques. The structure of a fuzzy system lends itself to being able to capture expert knowledge. The fuzzy system representation allows re-mapping the knowledge contained in the system back into linguistic terms. Hence, the possibility of extracting knowledge from fuzzy systems remains high. Thus is more efficient in extracting knowledge. Here the representation parameters are rolled into the genetic coding and automatically tuning them along with the structure and other model parameters.

E. Rule Based Classification Technique

[10] Proposes a rule-based classification and prediction algorithm called uRule for classifying uncertain data. The algorithm introduces new measures for pruning, generating, and optimizing rules. These new measures are computed considering uncertain data interval and probability distribution function. The uRule algorithm uses the sequential covering approach to extract rules from the data. The algorithm extracts the rules one class at a time for a data set. A rule is desirable if it covers most of the positive examples and none of the negative examples. The algorithm is based on the RIPPER algorithm [1]. This algorithm follows the new paradigm of directly mining uncertain datasets and shows an excellent performance even when data is highly uncertain.

III. CONCLUSION

The need for handling uncertainty increases as uncertainty leads to incomplete information and unpredictability. There are many techniques available to analyze the uncertainty from the past decades. The main techniques discussed are Probabilistic analysis, fuzzy analysis, Bayesian Network analysis, Soft computing technique and Rule based classification technique. Among these the probability analysis, fuzzy analysis and Bayesian Network analysis are the most technically challenging techniques. The basic concepts behind these techniques are not so complex. The probabilistic analysis and fuzzy analysis helps to reduce linguistic uncertainty. The researches are going on this area to find an efficient mechanism to handle uncertainty thus will provide accurate result.

ACKNOWLEDGMENT

I express my thanks to my guide who have helped me to do this survey and ,all staffs and my friends who helped me with their knowledge and experiences to work on this paper “ Handling of Uncertainty – A Survey”.

REFERENCES

[1] W. W. Cohen (1995), “*Fast effective rule induction*,” in Proc. of the 12th Intl.Conf. Machine Learning, pp. 115–123.

[2] Daniel. P. Thunnissen. (2003) “*Uncertainty Classification for the Design and Development of Complex Systems*,” Proc. Third Annual Predictive Methods Conf. Veros Software.

[3] Helen. M. Regan., Mark. Colyvan. and Mark. A. Burgman (2002) “*A Taxonomy And Treatment Of Uncertainty For Ecology And Conservation Biology*,” Proc. Ecological Applications.

[4] E. Hui., Amornvadee Veawab. and Christine. W. Chan. (2004) “Uncertainty *Modeling in Selection of Gas/Liquid Contactors Using Bayesian Networks*,”Proc. 3rd IEEE Int’l Conf.Cognitive Informatics.

[5] F. Jensen. , S. Lauritzen. and K. Olsen. (1990), "*Bayesian updating in recursive graphical models by local computations*," Computational Statisticals Quarterly, pp.269–282.
[6] Keith. R. Hayes. (2011) "*Uncertainty and Uncertainty Analysis Methods*".
[7] M. A. Lee. and M. H. Smith. (1995), "Handling *Uncertainty in Finance Applications Using Soft Computing*," Proc. 3rd Int'l Symp Uncertainty Modelling and Analysis.
[8] Martin. Basil., Christos. Papadopoulos., Donald. Sutherland. and Hoi. Yeung. (2001) " *Application Of Probabilistic Uncertainty Methods(Monte Carlo Simulation) In Flow Measurement Uncertainty Estimation*," Int'l Conf.Flow Measurement.
[9] Meghdadi. H. Amir., R. Mohammad. and T. Akbarzadeh. (2003) "*Uncertainty Modeling through Probabilistic Fuzzy Systems*," Proc. 4th Int'l Conf.Uncertainty Modeling and Analysis, pp. 56 – 61.
[10] B. Qin., Yuni Xia., S. Prabhakar and Yicheng Tu. (2009), "A *Rule-Based Classification Algorithm for Uncertain Data*," Proc.IEEE Int'l.Conf on Data Engineering (ICDE), pp. 1633-1640.
[11] M. S. Ryoo. and J. K. Aggarwal. (2008) "Human *Activities: Handling Uncertainties Using Fuzzy Time Intervals*," Proc.19th Int'l.Conf.Pattern Recognition(ICPR 2008)., pp.1-5.
[12] S. Wasserkrug.,A. Gal. and O. Etzion. (2005) "*A Model for Reasoning with Uncertain Rules in Event Composition Systems*," Proc. 21st Conf. Uncertainty in Artificial Intelligence (UAI '05), pp. 599-606.
[13] S. Wasserkrug.,A. Gal., O. Etzion and Y. Turchin. (2008), "*Complex Event Processing over Uncertain Data*," Proc. Second Int'l Conf. Distributed Event-Based Systems (DEBS '08), pp. 253-264.
[14] Laura P. Swiler., Thomas L. Paez. and Randall L. Mayes (2009) " *Uncertainty Analysis*," Proc. IMAC.

AUTHORS

First Author – Lakshmi.S.Dutt, PG Scholar, Karunya University (Coimbatore, India) and lsdutt@gmail.com.
Second Author – Mathew Kurian, Assistant Professor, Karunya University (Coimbatore, India) and mathewk80@karunya.edu