REPRESENTATION OF UNCERTAIN DATA USING POSSIBILISTIC NETWORK MODELS

P. E. S. N. Krishna Prasad[1], K. Madhavi[2] and B. D. C. N. Prasad[3]

[1 & 3]Prasad V. Potluri Siddhartha Institute of Technology, India
[1]surya125@gmail.com and [3]bdcnprasad@gmail.com
[2]Dadi Institute of Engineering and Technology, Anakapalli, India
[2]kolukulurimadhavi@yahoo.com

ABSTRACT

Uncertainty is a pervasive in real world environment due to vagueness, is associated with the difficulty of making sharp distinctions and ambiguity, is associated with situations in which the choices among several precise alternatives cannot be perfectly resolved. Analysis of large collections of uncertain data is a primary task in the real world applications, because data is incomplete, inaccurate and inefficient. Representation of uncertain data in various forms such as Data Stream models, Linkage models, Graphical models and so on, which is the most simple, natural way to process and produce the optimized results through Query processing. In this paper, we propose the Uncertain Data model can be represented as Possibilistic data model and vice versa for the process of uncertain data using various data models such as possibilistic linkage model, Data streams, Possibilistic Graphs. This paper presents representation and process of Possiblistic Linkage model through Possible Worlds with the use of product-based operator.

KEYWORDS

Uncertain Data; Uncertain Object model; Possibilistic Data model; Possibilistic Linkages; Possibilistic Linkage model; Possible worlds; Product-based Operator; min-based Operator;

1. INTRODUCTION

Artificial Intelligence aims in model human reasoning in order to help decision makers in the respective tasks. The development of expert systems is one of the popular and famous applications in this domain. But these systems are unable to manipulate correctly incontrollable variables, due to imprecise and uncertain data that characterizes in the real world.

Graphical models [1,2,3,8] are knowledge representation tools proposed for an efficient representation and analysis of uncertain data that can be used by researchers from different domains such as industry, space, medicine, wireless technology, and so on. Well known graphical models[2,3,5,14] are probabilistic graphical models like Bayesian networks, Bayesian belief networks, decision trees, influence diagrams and value based systems. Most of these models refer to probability theory and these models works out on uncertain data, but precise. Such models cannot be preferred for imprecise data, due to lack of ambiguity and incompleteness. In order to process such imprecise data, several non-classical models have been proposed such as evidence theory[2,22,32,33,34], utility theory, uncertain theory, Lehmann's ranked model, plausibility relations, Spohn's ordinal conditional functions and Possibility theory issued from the fuzzy set theory.

Sundarapandian et al. (Eds): CoNeCo,WiMo, NLP, CRYPSIS, ICAIT, ICDIP, ITCSE, CS & IT 07,
pp. 267–278, 2012. © CS & IT-CSCP 2012 DOI : 10.5121/csit.2012.2424

Among these models, Possibility theory [3,18,32] offers a natural and simple model to handle uncertain data. It is a framework for experts to express the uncertainty numerically in terms of possibility degrees in the universe of discourse.

The aim of this paper presents to represent the uncertain data model into Possibilistic data model and vice versa using graphical models with the support of possible worlds.

2. POSSIBILISTIC MODELS

Possibility theory [2, 14, 16, 20,32] is one the framework for the process of imprecise and uncertain data proposed by Zadah. This theory is based on the idea that we can evaluate the possibility of determinant variable 'x' belongings to a determinant set or event A. Here, fuzzy sets are called possibility distributions instead of membership degrees may often called possibility degrees.

[12,13,32]When the state of knowledge is expressed by a body of evidence it becomes clear that probability measures address precise but differentiated items of information, whereas possibility measures reflect imprecise but coherent items. So, possibility measures are useful for subjective uncertainty: one expects from information no very precise data; however, one expects the greatest possible coherence among his statements. On the other hand, precise, but variable data are usually the result of carefully observing physical phenomena. As a rule, the state of knowledge is neither precise nor totally coherent, in the general case, the of "A degree of credibility"(the degree of confidence).

Possibilistic Networks and Possibilistic logic [2,4,10,16] are two standard frameworks for representing uncertain pieces of knowledge. Possibilistic Networks exhibit relationship between the features where as Possibilistic logic ranks the logical formulas according to their level of certainty, it is well known that the inference process is a hard problem. These two types models representation are semantically equivalent when they lead to same possibility distribution. A possibility distribution can be decomposed using a chain of rule that may be based on two different kinds of conditioning that exist in possibility theory. These two types induce the possibilistic graphs.

The possibilistic graphical models are either a direct adaptation of probabilistic approach without knowledge representation or a way to perform learning from imprecise data. The notion of possibilistic graphs [17] for the representation of multidimensional possibility distributions are of two types 1) undirected graphs or hypergraphs 2) Directed graphs.

Possibilistic Networks[21] is an important tool for an efficient representation and analysis of Uncertain Data. The simplicity and capacity of representing and handling of independence relationships are important for an efficient management of uncertain information.

Possibilistic networks are Undirected / Directed graphs where each node (vertex) encodes a variable/feature and every edge represents a casual or an inference relationship between variables. Uncertainty is expressed as conditional possibility distribution for each node in the context.

Possibilistic logic [4,13,17] is an extension of classical logic. A weight is associated with each propositional formula. This weight represents the priority reading other formulas. The set such weighted formulas is called possibilistic knowledge base.

The basic element of possibility theory [32] is the possibility distribution 'π' which maps from Ω to [0, 1]. The degree $\pi(\omega)$ represents the compatibility of ω with the available information about the real world. A possibility distribution π is said to be normal if π(ω)=1, there exists at least one interpretation which is consistent with all the available beliefs..

The possibility distribution associated with the knowledge base $\Sigma = \{(p,\alpha)\} is\ \forall \omega \in \Omega$

$$\pi_{\{(p,\alpha)\}}(\omega) - \begin{cases} 1 \ \ if\ \omega \in [p_i] \forall (p_i, \alpha_i) \in \Sigma; , \\ 1 - \max\{\alpha_i : (p_i, \alpha_i) \in \Sigma;\ \omega \notin [p_i]\}\ otherwise \end{cases}$$

Thus using the minimum operator, $\pi(\omega) = \min\ \{\pi_{\{(p_i,\alpha_i)\}}(\omega) : (p_i, \alpha_i) \in \Sigma;$

2.1 Axioms of Possibilistic model

The list of axioms of the Possiblistic model[15,32] is given below:

$$\pi(X) = 1$$

$$\pi(\emptyset) = 1$$

$$\pi(E_1 \cup E_2) = \max(\pi(E_1), \pi(E_2))$$

$$\pi(E_1 \cap E_2) \le \min(\pi(E_1), \pi(E_2))$$

$$\pi(E_1 \cap E_2) = \min(\pi(E_1), \pi(E_2))$$ (for non-interactive events)

$$max(\pi(E\_1), \bar{\pi}(E\_2)) = 1$$

$$N(E) = 1 - \pi(E)$$

2.2 Conditional Independence

Let E_1,E_2 and E_3 be the three disjoint subsets of features, then E_1 is called conditionally independent[4,13,17] of E_2 given E_3 with respect to π, if $\forall \omega \in \Omega$

$$\pi(\omega_{E_1 \cup E_2} | \omega_{E_3}) = min\{\pi(\omega_{E_1} | \omega_{E_3}), \pi(\omega_{E_2} | \omega_{E_3})\}$$

Whenever $\pi(\omega_{E_3}) > 0$, where $\omega_{E_3} = proj_{W_3}(\omega)$ is the projection of a tuple ω to the features in E_1 and $\pi(.|.)$ is a non-normalized conditional possibility distribution

$$\pi(\omega_{E_1} | \omega_{E_3}) = \max\{\pi(\omega) | \omega \in \Omega \wedge proj_{E_1}(\omega) = \omega_{E_1} \wedge proj_{E_3}(\omega) = \omega_{E_3}$$

In possibility theory, the possible definition of conditioning is defined as

$$\pi(W \Lambda Q \mid \phi) = \begin{cases} \pi(\omega) & if \omega \in \phi \\ 0 & otherwise \end{cases}$$

and also two different ways to define the counterpart of casual Bayesian networks, due to existence of possibilistic conditioning[3,13,16,26,31,33]: 1) Min-based Conditioning 2) product-based conditioning.

1. Min-based Conditioning: In min-based conditioning (ordinal elements), the maximal possibility degree is

$$\pi(\omega |_m, \phi) = \begin{cases} 1 & if \pi(\omega) = \Pi(\phi) and \omega \in \phi \\ \pi(\omega) & if \pi(\omega) < \Pi(\phi) and \omega \in \phi \\ 0 & otherwise \end{cases}$$

2. Product-based Conditioning: In product-based conditioning (numerical elements), the possibility degree is

$$\pi(\omega |_p, \phi) = \begin{cases} \dfrac{\pi(\omega)}{\Pi(\phi)} & if \omega \in \phi \\ 0 & otherwise \end{cases}$$

According o the Bayesian rule, these two conditions satisfy the a unique equation

$\forall \omega, \pi(\omega) = \pi(\omega \mid \phi) \otimes \Pi(\phi)$

For instance, consider an uncertain knowledge base with the features weight(H,M,L) and quality(G,B) of XYZ. product with the initial distribution, new distribution, min-based and product –based conditioning, $\phi = (L \Lambda G, M \Lambda B)$, $\Pi(\phi) = \max(0.8, 0.5) = 0.8$, which is presented in ***Table1***.

Table1: Simple Uncertain Knowledge with Possibilistic Conditioning.

Weight	Quality	Initial Distribution $\pi(W \Lambda Q)$	New Distribution $\pi(W \Lambda Q \mid \phi)$	Min-based Condition $\pi(W \Lambda Q \mid_m \phi)$	Product-based Condition $\pi(W \Lambda Q \mid_p \phi)$
H	G	0.3	0	0	0
M	B	0.5	0.5	0.5	0.63
L	G	0.4	0.4	0.4	0.5
H	G	0.5	0	0	0
L	G	0.8	0.8	1	1

2.3 Independence Relations in Possibilistic Framework

From [3,31,33], two types of independence relations have been considered such as 1) Possibilistic Casual Independence 2) Possibilistic Decompositional Independence: Non-Interactivity. For these two independence relations, possibility theory has several kinds of conditioning relations such as 1) Plausibility independent 2) min-based independent 3) Product-based independent 4) Pareto Independent 5) Leximin independent 6) Leximax independent relations. These relations have been discussed in [3,31,33]. From the observation of [3,31], min-based independence decomposition is considered in the possibilistic decompositional independence relation with the help of Graphical models. Basically a graphical model supports 5 properties such as P1: Symmetry, P2: Decomposition, P3: Weak Union, P4: Contraction, P5: Intersection.

Table2: Graphical Properties on Independence Relations

Independence Relation	**Symmetry**	**Decomposition**	**Weak Union**	**Contraction**	**Intersection**
Non-Interactivity	yes	yes	yes	yes	no
Min-based	no	yes	yes	yes	yes
Min-based Symmetry	yes	yes	yes	yes	yes
Product-based	yes	yes	yes	yes	yes if $\pi>0$
Pareto	yes	yes	yes	yes	yes
Plausibility	no	yes	yes	yes	yes
Leximax	yes	yes	no	no	no
Leximin	yes	yes	no	no	no

The independence relations supports only few of the properties[3] which are mentioned above, listed in Table2.

From these observations, the casual and decompositional independence relations based on *product* and *min* operators are reasonable relations with good properties, since they are semi-graphiods. Indeed, the min-based with symmetry property independence relation is too strong than the rest of the relations,[3,14], this relation has been considered in the Possibilistic Graphical model. The product-based relation is good but it cannot satisfy the intersection property, it is good and most used operator in the Probabilistic network models. It is good and most used operator, the min-based independence relation is considered to represent uncertain data model using Possibilistic graphical models.

2.4 Product-based Possibilistic Networks

A product-based possibilistic graph[3,16,19,31] over a set of variables, denoted by, ΠG_p, *is a possibilistic graph* where conditionals are defined using product-based conditioning:

$$\pi(\omega|_p,\phi)=\begin{cases}\dfrac{\pi(\omega)}{\Pi(\phi)} & if \omega\in\phi\\ 0 & otherwise\end{cases}$$

The joint distribution relative to product-based possibilistic networks can be computed using the product-based chain rule.

Product-based Chain Rule: Given a Product-*based possibilistic network* ΠG_p, the global joint possibility distribution over a set of variables $V=\{A_1,A_2,.....,A_n\}$ can be expressed as the product of the N initial a priori and conditional possibilities via product-based chain rule: $\pi_p(A_1,A_2,.....,A_n)=\prod_{i=1...N}\Pi(A_i\,|\,U_i)$, which is derived from the product independence relations induced by the local product based conditional degrees. From the Product-based relation, the local independence is defined by $\Pi(x\,|\,y\wedge z)=\Pi(x\,|\,z),\forall x,y,z$ to express that the variables sets X and Y are non-interactivity independent in the context of Z.

It may be noted that t*he product-based chain rule allows the recovering of initial data through the local distributions. Probabilistic networks use the product-based chain rule, so that the initial data can be recovered from local distributions, the same logic can be extended in possibilistic networks.*

2.5 Min-based Possibilistic Networks

A min-based possibilistic graph[3,19,31] over a set of variables, denoted by ΠG_m, is a possibilistic graph, where conditionals are defined using min-based conditioning :

$$\pi(\omega|_m,\phi)=\begin{cases}1 & if \pi(\omega)=\Pi(\phi) and \omega\in\phi\\ \pi(\omega) & if \pi(\omega)<\Pi(\phi) and \omega\in\phi\\ 0 & otherwise\end{cases}$$

The joint distribution relative to min-based possibilistic networks can be computed using the min-based chain rule.

Min-based chain rule: Given a *min-based possibilistic network ΠG_m*, the global joint possibility distribution over a set of variables $V = \{A_1, A_2, \ldots\ldots, A_n\}$ can be expressed as the minimum of the N initial a Priori and conditional possibilities via min-based chain rule: $\pi_m(A_1, A_2, \ldots\ldots, A_n) = \min_{i=1\ldots N} \prod(A_i \mid U_i)$, which is derived from the minimum independence relations. From the non-interactivity of the min-based is defined by $\Pi(x \wedge y \mid z) = \min(\Pi(X \mid z), \Pi(y \mid z)), \forall x, y, z$ to express that the variables sets X and Y are non-interactivity independent in the context of Z.

It may be noted that, i*n the min-based Possibilistic networks do not recover the initial data provided by the experts, since the unrecovered data correspond to the redundant data that can be ignored and have no influence on independence relations.*

In this paper, we considered product-based operator to represent uncertain data using Possibilistic networks

3. UNCERTAIN DATA MODEL

Modeling and Querying uncertain data [6,14,20,22,24] has been a fast growing research direction and receives an increasing attention. Various models of uncertain and fuzzy data have been developed. We proposed a novel model for modeling uncertain data in the fuzzy environment using *Possibilistic data model***.** The working model for uncertain data describes the existence possibility of a tuple in an uncertain data set and the constraints on the uncertain tuples.

A fuzzy database [10,11,14,18] comprises of multiple fuzzy tables. A fuzzy table contains a set of tuples, where each tuple is associated with a fuzzy membership value, which is treated as Degree of Possibility in the Possibilistic Model. A fuzzy table may also come with some generation rules to capture the dependencies among tuples, where a generation rule specifies a set of exclusive tuples, and each tuple is involved in at most one generation rule.

Another useful model is the *Uncertain Object Model* [6,9,18]; an uncertain object is conceptually described by a fuzzy membership function, i.e. Possibility Distribution in the data space. In this scenario a possibility degree of an uncertain object is unknown, a set of samples (instances) are collected to approximate the fuzzy distribution, which is a possibility distribution.

Definition: An *Uncertain Object* is a set of instances $X = \{x_1, x_2, \ldots\ldots, x_m\}$ such that each instance $x_i \ (1 \leq i \leq m)$ takes a a Possibility degree $\pi(x_i) and \ \sum_{i=1}^{m} \pi(x_i) \leq 1$.

The cardinality of an uncertain object $X - \{x_1, x_2, \ldots\ldots, x_m\}$ denoted by |X| is the number of instances contained in *X*. The set of all uncertain objects denoted by *U*, $U - \{X_1, X_2, \ldots\ldots, X_n\}$.

Possible worlds of Uncertain Objects:

Let $U = \{X_1, X_2, \ldots\ldots, X_n\}$ be a set of Uncertain Objects. A possible world $W = \{x_1, x_2, \ldots\ldots, x_m\}$ is a set of instances such that one instance is taken from each uncertain Object. The existence membership of ω is $\mu(\omega) = \pi(\omega) = \prod_{i=1}^{n} \pi(x_i)$, where W denotes the set of all possible worlds. Let $|X_i|$ be the cardinality of Object X_i ($1 \le i \le m$), the number of possible worlds is $|W| = \prod_{i=1}^{n} |X_{i}|$. Thus, $\pi(W) = \sum_{w \in W} \pi(w) = 1$.

A Possibilistic database model is used to represent uncertain data, which is a finite set of Possibilistic tables; A Possibilistic table contains a set of *uncertain tuples* T and a set of *generation rules* $\mathfrak{R}$. Each uncertain tuple $t \in T$ is associated with a *possibility degree* $\pi(t) > 0$. Each generation rule $R \in \mathfrak{R}$ specifies a set of exclusive tuples in the form $R: t_{r_1} \oplus t_{r_2} \oplus \ldots \oplus t_{r_m}$, where $t_{r_i} \in T (1 \le i \le m)$, $\pi\left(t_{r_i} \wedge t_{r_j}\right) = 0 \ (1 \le i, j \le m, i \ne j)$ and $\sum_{i=1}^{m} \pi\left(t_{r_i}\right) = 1$.

The cardinality of a generation rule R, denoted by |R|, is the number of tuples involved in R. The generation rule R is the set of all tuples $t_{r_1}, t_{r_2}, \ldots \ldots t_{r_m}$ involved in the rule, at most one tuple can appear in a possible world. R is a singleton rule if there is only one tuple involved in the rule, otherwise R is a multiple rule, and thus the Possibilistic database follows a possible world.

Given a Possibilistic Table, $\tilde{T}$ a possible world W is a subset of $\tilde{T}$ such that for each generation rule $\in \mathfrak{R}_{\tilde{T}}, |R \cap W| = 1, if\ \pi(R) \le 1 and\ |R \cap W| \le 1, if\ \pi(R) < 1$. Thus, the existing membership of W is

$$\pi(W) = \prod_{R \in \mathfrak{R}_{\tilde{T}}, R \cap W = 1} \pi(R \cap W). \prod_{R \in \mathfrak{R}_{\tilde{T}}, R \cap W = \emptyset} \pi(1 - \pi(R)).$$

For an uncertain table $\tilde{T}$ with a set of generation rules $\mathfrak{R}_{\tilde{T}}$, the number of all possible worlds is

$$|W| = \prod_{R \in \mathfrak{R}_{\tilde{T}}, \pi(R) = 1} |R| . \prod_{R \in \mathfrak{R}_{\tilde{T}}, \pi(R) < 1} |R| + 1.$$

We can convert the *Uncertain Object model* into *Possibilistic database model*[9,24,27,39,] and vice versa. When, it is concerned that both are equivalent. The conversion process will be presented below.

1. *Conversion between Uncertain Object Model to Possibilistic Database Model*: A set of uncertain objects can be represented by a fuzzy table. For each instance '*x*', of an uncertain Object 'X', to create a tuple t_x, whose membership or possibility degree value is $\mu(x) = \pi(x)$. for each uncertain object $X = \{x_1, x_2, \ldots\ldots, x_m\}$, to create one generation rule $R_X: t_{x_1} \oplus t_{x_2} \oplus \ldots \oplus t_{x_m}$.
2. *Conversion between Possibilistic Database models to Uncertain Object Model*: A fuzzy table can be represented by a set of uncertain objects with discrete instances. For each tuple *t* in a fuzzy table, to create an instance x_t, whose membership or possibility degree is $f(x_t) = \pi(t)$. For a generation rule $R: t_{r_1} \oplus t_{r_2} \oplus \ldots \oplus t_{r_m}$, to create an uncertain object X_R, which includes instances $x_{t_{x_1}}, \ldots\ldots x_{t_{xm}}$ corresponding to $t_{r_1}, \ldots\ldots\ldots t_{r_m}$,

respectively. Moreever, $\sum_{i=1}^{m} \pi(t_{r_i}) < 1$. We create another instance x_Q whose fuzzy membership function is $f(x_Q) = 1 - \sum_{i=1}^{m} \pi(t_{r_i})$ and u_Q to the uncertain object X_R.

4. POSSIBILISTIC LINKAGE MODEL

In the basic uncertain object model, we assume that each instance belongs to a unique object, though the object may have multiple instances, if an instance may belong to different objects in different possible worlds. Such a model is useful in Possibility Linkage analysis.

A Possibilistic linkage model[9,14,15,18] contains two sets of tuples A and B and a set of linkages $\mathfrak{T}$. Each linkage ℓ in matches one tuple in A and one tuple in B. For a linkage ℓ =(t_A,t_B), we say ℓ is associated with t_A and t_B. We write $\ell \in t_A$ and $\ell \in t_B$. We consider each tuple $t_A \in A$ as an uncertain object and $t_A \in B$ as an instance of t_A if there is a linkage $\ell = (t_A, t_B) \in \mathfrak{T}$. The membership possibility of instance t_B with respect to object t_A is $\pi(\ell)$, which is $\mu(\ell)$.

Object t_A may contain multiple instances $\{t_{B_1}, \ldots\ldots, t_{B_k}\}$ where $(t_A, t_B) \in \mathfrak{T}(1 \leq i \leq k)$. At the same time, an instance t_B may belong to multiple objects $\{t_{A_1}, \ldots\ldots, t_{A_d}\}$ where $(t_{A_j}, t_B) \in \mathfrak{T}\,(1 \leq j \leq d)$. A mutual exclusion rule $R_{t_B} = (t_{A_1}, t_B) \oplus \ldots\oplus (t_{A_d}, t_B)$ specifies that t_B can only belong to one object in a possible world.

A ***record linkage***[5,9,24] is a technique that finds the linkages among data entries referring to the same real world entities from different data sources. In the real world applications, data is often incomplete or ambiguous. Thus, record linkages are often uncertain.

Possibility Record Linkages [5,9,24,27]are often used to model the uncertainty. For teo records, a state- of-the art, possibility record linkage model can estimate the possibility degree that the two records refer to the same real world entity. Let us consider teo thresholds $\alpha_1 \& \alpha_2 (0 \leq \alpha_1 < \alpha_2 \leq 1$. When the possibility linkage is less than α_1, the records are not matched. When the possibility linkages is between$\alpha_1 \& \alpha_2$, then records considered possibly matched.

To build a possibility record linkage effectively and efficiently with the some real world scenarios. Each linked pair of records as an uncertain instance and each record as an uncertain object. Two uncertain objects from different data sets may share zero or one instance. Thus the uncertain objects may not be independent. For instance, let us consider the patient data from hospitalized registered and cause of death data, which is presented in ***Table3***.

Table3: Record linkages between the patients registered data and cause of death registered data.

Linkage ID	Patient Registered Data			Cause of Death Data			Possibility Degree
	PID	Name of the Patient	Disease	DID	Name of the Patient	Age	
l1	*x1*	Sita M. Lakshmi	Flu	*y1*	Maha Lakshmi	42	0.4
l2	*x1*	Sita M. Lakshmi	Flu	*y2*	M. Lakshmi	45	0.4
l3	*x1*	Sita M. Lakshmi	Flu	*y3*	S. Lakshmi	32	0.5
l4	*x2*	S. MahaLakshmi	Cancer	*y3*	S. Lakshmi	32	0.2
l5	*x2*	S. MahaLakshmi	Cancer	*y4*	S. M. Lakshmi	55	0.8

Let E be the set of real world entities. Let us consider two tables A and B which describe subsets $E_A, E_B \subseteq E$ of entities in *E*. Each entity is described by atmost one tuple in each table. In general, E_A *and* E_B may not be identical, they may have different schemas as well.

Possibility Linkage: Consider two tables A and B each describing a subset of entities in E, a linkage function $L: A \times B \rightarrow [0,1]$ gives a score $L(t_A, t_B)$ for a pair of tuples $t_A \in A$ *and* $t_B \in B$ to measure the likelihood that t_A *and* t_B describes the same entity in *E*.

A pair of tuples $l = (t_A, t_B)$ is called a possibility record linkage, if $L(l) > 0, \pi(l) = L(t_A, t_B)$ is the possibility degree of '*l*'. Given a linkage $l = (t_A, t_B)$, the larger the possibility degree π(*l*), the more likely the two tuples t_A *and* t_B describe the similarity entity.

A tuple $t_A \in A$ may participate in zero, one or multiple linkages. The number of linkages that t_A participates in as called the ***Degree*** of t_A denoted by $d(t_A)$. Similarly we can define $d(t_B)$.

For a tuple $t_A \in A$, let $l = (t_A, t_{B_1}), \ldots \ldots l_d = (t_A, t_{B_d})$ be the linkages that t_A participates in. For each tuple $t_A \in A$, we can write a ***Mutual Exclusive Rule (MER)*** $R_{t_A} = l_1 \oplus l_2 \oplus \ldots \ldots \oplus l_d$, where *d* is the degree of $t_A \in A$ that indicates atmost one linkage can hold based on the assumption that each entity can be described by atmost one tuple in each table. The possibility degree is computed as $\pi(t_A) = \sum_{i=1}^{d(t_A)} \pi(l_i)$. that t_A is matched by some tuples in B. Since the linkage function is normalized, $\pi(t_A) \leq 1$. It is denoted by $R_A = \{R_{t_A} | t_A \in A\}$, the set of mutual exclusion rules for tuples in A. Similarly R_{t_B} for $t_B \in B$, are symmetrically defined.

Therefore *(L,A,B)* specifies a bipartite Graph, where tuples in A and those in B are two independent sets of nodes respectively and the edges are the linkages between the tuples in the two data tables.

4.1 Connection with the Uncertain Object Model.

Given a set of Possibility linkages, *L* between tuple sets, *A and B* , we consider each tuple $t_A \subset A$,as an uncertain object. For any tuple $t_B \subset B$, if there is a linkage $l = (t_A, t_B) \subset L$ such that $\pi(l) > 0$. Then t_B can be considered as an instance of object $t_A \in A$ whose possibility degree is $\pi(l)$.

In contrast to the basic uncertain object model where each instance only belongs to one object, in the Possibility Linkage model, a tuple $t_B \in B$ may be the instance of multiple objects$\{t_{A_1}, t_{A_2}, \dots, t_{A_d}\}$ where d is the degree and t_{A_i} is a tuple in A with linkage $(t_{A_i}, B) \in L (1 \le i \le d)$. A *mutual exclusion rule* $R_{t_B} = (t_{A_1}, t_B) \oplus \dots \oplus (t_{A_d}, t_B)$ specifies that t_B should only belong to one object in a possible world.

Alternatively, we consider each tuple $t_B \in B$as as uncertain object and a tuple $t_A \in A$ is an instance of $t_B$if there is a linkage $(t_A, t_B) \in L$. Thus, a *linkage function* can be regarded as the summarization of a set of possible worlds.

For a *linkage function L* and *tables A and B*, let $L_{A,B}$ be the set of linkages between tuples A and B. A Possible world of $L_{A,B}$ denoted by $W \subseteq L_{A,B}$ is a set of pairs $l = (t_A, t_B)$ such that

1. For any *mutual exclusion rule* $R_{t_A}, if \pi(t_A) = 1$, then there exists one pair $(t_A, t_B) \in W$. Symmetrically, for any *mutual exclusion rule*, $R_{t_B}, if \pi(t_B) = 1$ then there exists one pair $(t_A, t_B) \in W$.
2. Each tuple $t_A \in A$participates in at most one pair in W, so does each tuple $t_B \in B$. $W_{L_{A,B}}$ denotes the set of all possible worlds of $L_{A,B}$.

Similarly we can represent the uncertain data models in the form of Data Streams as well as Possibilistic Graphical models in the form of Possibilistic Networks that can be discussed in future presentations.

5. Conclusion

The object of this paper is to represent the uncertain data in various forms of Object models for processing and evaluating Query and also give the ranking to the evaluated query. Here, an uncertain object model is represented as Possibilistic Database Model using Possibilistic Networks through Product-based operator and vice versa so that the uncertain data model can be evaluated through the query evolution mechanism using Possibilistic Database model. Further, the uncertain data may be represented as Data streams and Possibilistic Graphical Models that process the data objects to evaluate through query evaluation mechanism using Possibility theory.

References

[1] Christian Borgelt, Learning Possibilistic Graphical Models from Data, IEEE Transactions on Fuzzy Systems, 2003.
[2] Christian Borgelt, J. Gebhardt, and Rudolf Kruse, Learning from Imprecise data: Possibilstic Graphical Models, Journal Computational Statistics & Data Analysis - Nonlinear methods and data mining, Vol 38, No 4 2002, pp. 449 – 463.
[3] Nahla Ben Amor, Qualitative Possiblistic Graphical Models: From Independence to Propogation algorithms. Ph. D. Thesis, 2012
[4] Salem Benferhat, Didier Dubois Laurent Garcia and Henri Prade, Possibilistic logic bases and Possibilistic Graphs.
[5] Thomos N. Herzog, Fritz J. Scheuren and William E. Winkler, Data Quality and Record Linkages, Springer, 2007,pp.81-92.
[6] Amihai Motro, Imprecision and Uncertainty in Database Systems,

[7] Angelos Vasilakopoulos andVerena Kantere, Efficient Query Computing for Uncertain Possibilistic Databases with Provenance.
[8] Jorg Gebhardt and Rodulf Kruse , Learning Possibilistic Graphical Networks from Data, IEEE Transaction on Fuzzy Systems
[9] Ming Hua, Jian Pei, Ranking Queries on Uncertain data, Advances in Database Systems.
[10] M. S. Sunitha and A. Vijay Kumar, Complement of a Fuzzy Graph, Indian J. Pure Applied Mathematics, 33(9), 1451-1464, September 2002.
[11] Damiani. E, Tanca L.,Building Queries to XML data, Database and Expert Systems Applications, 2000, pp 266-279.
[12] Christian Borgelt, J¨org Gebhardt, and Rudolf Kruse, Possibilistic Graphical Models,
[13] Pascale FONCK, Conditional Independence in Possibility Theory, UAI'94 Pp.224 – 226.
[14] Christian Borgelt, and Rudolf Kruse, Data Mining with Possibilistic Graphical Models
[15] Salem Benferhat and Salma Smaoui, A Hybrid Possibilistic Networks, International Journal of Approximate Reasoning, Vol 44, No.3, 2007,224-243.
[16] Salem Benferhat and Salma Smaoui, on the use of Possibistic bases for local computations in product based possibilistic Networks
[17] Christian Borgelt and J¨org Gebhardt, A Naïve Bayes style Possibilistc Classifier, Proc. 7th European Congress on Intelligent Techniques and Soft Computing.
[18] Anish Das Sarma, Omar Benjelloun, Alon Halevy,Jennifer Widom, Working Models for Uncertain data, Data Engineering ICDE-2006
[19] Christian Borgelt and Rudolf Kruse, Operations and Evoluation measures for Learning Possibilistic Graphical Models, Artificial Intelligence, Vol. 148, No.1-2, 2003, pp. 385-418.
[20] Jiang Chen and Ke Yi, Dynamic structures for Top-kqueries on Uncertain Data, Proceedings of the 18th international conference on Algorithms and computation, 2007, pp. 427-438
[21] AbdelKader Heni, M Nazih Omri and Adel M Alimi, Knowledge Representation with Possibilstic and Certain Bayesian Networks, Proceedings of the 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain, 2005, pp472-477
[22] C.C. Agarwal, Managing and mining Uncertain Data
[23] Brian Babcock Shivnath Babu Mayur Datar Rajeev Motwani Jennifer Widom, Models and Issues in Data Stream Systesms, Springer, 2002 pp1-16.
[24] Anish Das Sarma, Omar Benjelloun, Alon Halevy, Shubha Nabar, and JenniferWidom, Representing Uncertain Data: Models,Properties and Algorithms, VLDB Journal, 2009.
[25] Salem Benferhat and Salma Smaoui, Possibilistic Casual Networks for handling Interventions: A New Propogation Algorithm
[26] Salem Benferhat, Didier Dubois Laurent Garcia and Henri Prade, On the Transformation between Possibilistic Logic bases and Possibilistic Casual Networks, International Journal of Approximate Reasoning 29 (2002) 135–173.
[27] Jiang Chen and Ke Yi , Ranking Distributed Probabilistic Data, ACM SIGMOD-2009, pp:361-374.
[28] Goetz Graefe, Query Evaluation Techniques for Large Databases, ACM Computing Surveys, Vol25,No. 2, 1993, pp. 73-169
[29] Graham Cormode, Divesh Srivastava, Entong Shen and Ting Yu, Aggregating Query Answering on Possibilistic Data with Cardinality ConstraintsICDE,2012
[30] D. Dubois and H. Prade, Possibility Theory: an approach to Computerized Processing of Uncertainty, Plenum Press, 1988.
[31] Lawrence J. Mazlack, Using Imprecise Complexes to Computationally Recognize Casual Relations in Very Large Data Sets, Proceedings of IPMU, 2008,pp. 698-704
[32] Irina Georgescu, Possibility Theory and the Risk, Studies in Fuzziness and Soft Computing, Springer,2012
[33] Jirina Vejnarova, Conditional Independence in Evidence Theory, Proceedings of IPMU, 2008, pp.938-945. G. Shafer, Mathematical Theory of evidence, Princston Press,1976.

AUTHOR'S PROFILES

P. E. S. N. Krishna Prasad, currently a Researcher in the area of Machine Intelligence and Neural Networks. He is working in Computer Applications at Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India..He is a member of ACM, IEEE, ISTE. He has presented and published papers in several International Conferences and Journals. His areas of interest are Artificial Intelligence, Neural Networks and Machine Intelligence, Computational Intelligence and Information security Applications.

Dr. B.D.C.N.Prasad, currently is a Professor & Head, Departments of Computer Applications at Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh, India. He received Ph.D in Applied Mathematics from Andhra University, Visakhapatnam, A.P, India in 1984. His research interest includes Data Mining, Rough Sets, Fuzzy Sets, Machine Intelligence, Soft Computing and Information Security . He has several publications in Mathematics and Computer Science in reputed national and International journals, and Conferences. He is a member of ACM, IEEE, ISTAM and ISTE. He is a national executive member of Indian Society for Rough Sets.