International
Journal
Of Advanced Research in Engineering & Management (IJAREM)
ISSN: 2456-2033 || PP. 01-04

Uncertain Data Analysis Using Possibilisic Linkage Model

Madhavi Kolukuluri1, B.D.C.N. Prasad2
1Research Scholar in rayalaseema University, Kurnool, India
2PSCMR College of Engineering & Technology, Vijayawada, India

Abstract: Uncertainty is an unavoidable phenomenon in the real world data due to ambiguity and vagueness. Vagueness makes it difficult to make clear distinctions and Ambiguity is related to the problem of choosing the right one from few precise choices. Real world data lacks accuracy, completeness and efficiency and analyzing such data available in huge volumes is an essential task in real world applications. There are numerous applications that can handle uncertain data including Graphical models and linkage models, which is the most basic, common approach to process and deliver optimized outcomes from the Query processing. These models are generally represents either probabilistic based or plausibility based frameworks. In this paper we represent uncertain data using possibilistic linkage models.

1. Introduction

Information and uncertainty can be considered as two facets of the same coin. Increase in the volume of information reduces the uncertainty. Any information generally includes some element of uncertainty. The information may be regarding the structure and causal relationships of the system, the inputs of the systems, the goals and objectives, and the interpretation related to outcome of the analysis.

In this regard a novel proposal is essential to handle uncertainty. It requires additional information for validation of a specific proposition in a context dependent environment, apart from the required to arrive at absolute certainty.

The main reasons for modeling the large-scale systems in present day environment is with the assumption that all things are interdependent, so that a small variation in one sub system can affect many sub systems of the entire system. In such dynamic environment, accountability and risks attached to uncertainty become the critical aspects for making the policies and taking the decisions. This situation is observed in all domains including economic, political, engineering, science, and social issues and wider applicability in almost all domains where decision making is involved.

The amount of information sufficient to achieve entirely distinguishing specifications is required to measure the uncertainty. Marginal amount of information that satisfies the requirements of a particular task cannot meet the requirement for measuring uncertainty. In other context, uncertainty is in the view of analyst with a given level of uncertainty it may be sufficient for one specific problem but may be insufficient for another problem.

In the basic uncertain object model, assuming 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 generally has two sets of tuples “A” and “B” and a set of linkages $\mathfrak{T}$. Each linkage $\boldsymbol{\ell}$ in $\mathfrak{T}$ matches one tuple in “A” and one tuple in “B”. For a linkage $\boldsymbol{\ell}$ =(tA,tB), we say $\boldsymbol{\ell}$ is associated with tA and tB. We write $\boldsymbol{\ell}\varepsilon$ tA and $\boldsymbol{\ell}\varepsilon$tB . We consider each tuple tAεA as an uncertain object and tB ε B as an instance of tA if there is a linkage $\boldsymbol{\ell}$ =(tA,tB)ε$\mathfrak{T}$.

The membership possibility of instance tB with Object “tA” may contain multiple instances {tB1,tB2,…,tBk} where (tA, tBi)$\epsilon\mathfrak{I}(1\leq i\leq k)$. At the same time, an instance tB may belong to multiple objects {tA1,tA2,…,tAd} where (tAj,tB)$\epsilon\mathfrak{I}(1\leq j\leq d)$. A mutual exclusion rule RTB=(tA1,tB) ⊕ (tA2,tB) ⊕..... (tAd,tB) specifies that tB can only belong to one object in a possible world.

A record linkage is a technique that identifies the linkages among data entries that represent the same real world entities drawn from various data sources. In the real world applications, data is often incomplete or unclear. Hence, uncertainty arises even with the record linkages.

Possibility Record Linkages are generally applicable while modeling the uncertainty. For two records, possibility record linkage model can estimate the degree of possibility that the two records are related to the same real world entity. Let us consider two thresholds $\alpha 1$&$\alpha 2(0\leq\alpha 1<\alpha 2\leq 1)$. When the possibility linkage is less than α1 the records are not matched. When the possibility linkages are between α1&α2, 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 Table1.

Table1: Record linkages between the patients registered data and cause of death registered data

Link ID	Patient Registered Data			Reason of Death Data			Initial Possibility	New Possibility distribution	Min based Possibility	Product based possibility
	PID	Name of the Patient	Disease	DID	Name of the Patient	Age				
l1	x1	Sita M. Lakshmi	Heart attack	y1	Maha Lakshmi	42	0.3	0	0	0
l2	x1	Sita M. Lakshmi	Heart attack	y2	M. Lakshmi	45	0.3	0	0	0
l3	x1	Sita M. Lakshmi	Heart attack	y3	S. Lakshmi	32	0.4	0.4	0.4	0.5
l4	x2	S. MahaLakshmi	Blood Cancer	y3	S. Lakshmi	32	0.2	0.2	0.2	0.25
l5	x2	S. MahaLakshmi	Blood Cancer	y4	S. M. Lakshmi	55	0.8	0.8	1	1

Let E be the set of real world entities. Let us consider two tables A and B which describe subsets EA,EB⊆E of entities in E. Each entity is described by at most one tuple in each table. In general, EA&EB may not be identical, they may have different schemas as well.

1.1 Possibility Linkage: Consider two tables A and B each describing a subset of entities in E, a linkage function L:A×B→[0,1] gives a score L(tA,tB) for a pair of tuples tAϵA, tBϵB to measure the likelihood that tA & tB describes the same entity in E.

A pair of tuples $\boldsymbol{l}$=(tA,tB) is called a possibility record linkage, if L($\boldsymbol{l}$)>0, Poss($\boldsymbol{l}$)=L(tA,tB) is the possibility degree of ‘l’. Given a linkage $\boldsymbol{l}$=(tA,tB) , the larger the possibility degree Poss(l), the more likely the two tuples tA & tB describe the similarity entity.

A tuple tAϵA may participate in zero, one or multiple linkages. The number of linkages that tA participates in as called the Degree of tA denoted by d(tA). Similarly we can defined(tB).

For a tuple tA ϵA, let $\boldsymbol{l}$ 1=(tA,tB1),...., $\boldsymbol{l}$ d=(tA, tBd) be the linkages that tA participates in. For each tuple tA ϵA, we can write a Mutual Exclusive Rule (MER) RtA= $\boldsymbol{l}$ 1 $\boldsymbol{l}$ 2⊕…⊕ $\boldsymbol{l}$ d(tA), where d is the degree of tA ϵA , 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 $Poss(tA)$=Σ$Poss(li)$ $d(tA)i$=1that tA is matched by some tuples in B. Since the linkage function is normalized, Poss(tA)≤1. It is denoted by RA={RtA/ tA ϵA }, the set of mutual exclusion rules for tuples in A. Similarly RtB for tB ϵB, are symmetrically defined.

Therefore (£,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.

1.2 Connection with the Uncertain Object Model.

Given a set of Possibility linkages, L between tuple sets, A and B, we consider each tuple tA ϵA , as an uncertain object. For any tuple tB ϵB , if there is a linkage ***l*** =(tA,tB) such that Poss(***l***)>0. Then tB can be considered as an instance of object tA ϵA whose possibility degree is Poss(***l***).

In contrast to the basic uncertain object model where each instance only belongs to one object, in the Possibility Linkage model, a tuple tB ϵB may be the instance of multiple objects {tA1,tA2,…,tAd} where d is the degree and tAi is a tuple in A with linkage (tAi,tB) ϵ L(1≤i≤d).A mutual exclusion rule RtB=(tAi,tB) ⊕ .... ⊕ (tAd,tB) specifies that tB should only belong to one object in a possible world. Alternatively, we consider each tuple tB ϵB as an uncertain object and a tuple tA ϵA is an instance of tB if there is a linkage (tA,tB)ϵL.

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 LA,B be the set of linkages between tuples A and B. A Possible world of LA,B denoted by W⊆LA,B is a set of pairs ***l*** =(tA,tB) such that for any mutual exclusion rule, RtA, if Poss(tA)=1, then there exists one pair (tA,tB)ϵL . Symmetrically, for any mutual exclusion rule, RtB, if Poss(tB)=1, then there exists one pair (tA,tB)ϵW.

Each tuple t_A ϵA participates in at most one pair in W, so does each tuple t_B ϵB . $W_{LA,B}$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 using Possibilistic Networks that can be discussed in future presentations.

Conclusions

The object of this paper is to represent and analyze the uncertain/vague data using Possibilistic object linkage models for the process and evaluation of Query and also provide the ranking to the evaluated query. Here, an uncertain object model is represented as Possibilistic Database Model using Possibilistic Networks through record linkage and tuple analysis 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 system using Possibility theory.

References

[1]. Amihai Motro (1995) Imprecision and Uncertainty in Database Systems. In: Bosc P., Kacprzyk J. (eds) Fuzziness in Database Management Systems. Studies in Fuzziness, vol 5. Physica, Heidelberg
[2]. Chickering D.M., Geiger D., and Heckerman D. (1994). Learning Bayesian Networks is NP-Hard (Technical Report MSR-TR-94-17). Microsoft Research, Advanced Technology Division, Redmond, WA,USA
[3]. Dubois D., Fargier H. and Prade H. (1996a). Possibility theory in constraint satisfaction problems: Handling priority, preference and uncertainty, Applied Intelligence, 6, 287-309.
[4]. Dubois D., Foulloy L., Galichet S. and Prade H. (1997). Two different views of approximate reasoning, Proc. of the 7th Inter. Fuzzy Systems Assoc. World Congress (IFSA'97), Academia, Prague, 238-242.
[5]. Dubois D., Lang J. and Prade H. (1987) Theorem proving under uncertainty — A possibility theory-based approach", Proc. of the 10th Inter. Joint Conf. on Artificial Intelligence, Milano, 984-986
[6]. Dubois D., Lang J. and Prade H. (1991a). Fuzzy sets in approximate reasoning — Part 2: Logical approaches, Fuzzy Sets and Systems, 40, 203-244.
[7]. Dubois D., Lang J. and Prade H. (1991b). Timed possibilistic logic, FundamentaInformaticae, XV, 211-234.
[8]. Gibbs W. (1902). Elementary Principles of Statistical Mechanics. Yale University Press, New Haven, Connecticut, USA.
[9]. Heckerman D, Geiger D, and Chickering D.M(1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. Machine Learning 20:197–243. Kluwer, Dordrecht, Netherlands.
[10]. Prade H. (1984). Lipski's approach to incomplete information databases restated and generalized in the setting of Zadeh's possibility theory. Information Systems, 9, 27-42.
[11]. Yager R. R. (1987b). Possibilistic qualification and default rules, Uncertainty in Knowledge-Based Systems (Bouchon B. and Yager R. R., eds.), Lecture Notes in Computer Science, Vol. 286, Springer Verlag, Berlin, 41-57

[12]. Zadeh, L. A. (1971). Quantitative fuzzy semantics, Information Sciences. 3 (2): 159–176. doi:10.1016/S0020-0255(71)80004-X.
[13]. Zadeh L. A. (1972). A fuzzy set-theoretic interpretation of linguistic hedges, J. of Cybernet., 2, 4-34.
[14]. Zadeh L.A. (1975). Fuzzy Logic and approximate reasoning (In memory of GrigoreMoisil), Synthese, 30, 407-428.
[15]. Zadeh L.A (1978). Fuzzy Sets as a Basis for a Theory of Possibility. Fuzzy Sets and Systems 1:3–28. North-Holland, Amsterdam, Netherlands.
[16]. Zadeh L. A. (1984). Fuzzy probabilities, Information Proc. and Manag., 20, 363-372.
[17]. Zadeh L.A and Kacprzyk J. (1992). Fuzzy Logic for the Management of Uncertainty. J. Wiley & Sons, New York, NY, USA.