RESEARCH PAPER | Computer Science | Volume : 2 | Issue : 2 | November 2012 | ISSN - 2249-555X

Uncertain Data in Application- A case Study

KEYWORDS

Mr. Vishal K. Pandya	Dr. Dhaval R. Kathiriya
HOD, Shri V J Modha College Of IT	Director, Information Technology Anand Agricultural University, Anand

ABSTRACT *The appearing or found everywhere of uncertain data in modern-day application has outputted in a growing require for techniques or methods to work with such type of data. This work addresses challenges in managing uncertain data in principled, usable and scalable way. The arrival and growing importance of several new applications fields, information extraction on the web, integration, scientific databases, sensor, data management and de-duplication. Available DBMSs do not support uncertain data. This work develops techniques or methods to incorporate uncertainty in data.*

Uncertain Data in Application
Here are some illustrations of uncertain data arising in our day-to-day applications, motivating the need for some tools to manage some kind of information.

- Information extraction when extracting structured data from unstructured information such as on the Web, extractors typically provide a surety or probability of the extracted information being correct.
- Deduplication, the process of deciding automatically whether two data records represent the same real world entity is often approximate. Typically associates similarity scores with pairs of records.
- Data Integration, automatically integrating multiple sources of structured data may involve uncertainty at various levels.

Fundamentals
Here fundamentals of uncertain databases, an uncertain relation which is define R a set of possible instances. A relation instance is a multi-set of tuples and set semantic is obtained through explicit duplicate removal operations.

Data Management
Consider the data management for the Animals count, each year, volunteers and professionals worldwide observe or supervised animals for a fixed period of time and recording their observation the data from year to year is used to understand trends in animal populations and to correlate animal life with short term and long term environmental conditions. Individual animal sightings may not always be precise in terms of class, location or time and the observations of professionals may provide more reliable data than those of amateur.

Thus, there is inherent uncertainty in the data we use the following scheme for the actual animal count scheme. Like AnimalData: AnimalName, Color, Size and Other is Sightings: Observer, When, Where, Time, AnimalName. Let us begin with an observer Bhakti who definitely saw a Deer may or may not have seen another animal that was either an Elephant or Zibra. These observations can be represented in the sightings relations that are"Uncertain Relation", relations as follows: Observer: Bhakti, AnimalName: Deer, When: 25-09-11, Where: Chhambal, Time: 11:25AM (approx) Observer: Bhakti, AnimalName: Elephant / Zibra (*), When: 25/26-09-11, Where: Chhambal, Time: 14:27PM (approx) Here Elephant, Zibra is an attribute or defining uncertainty between two values and *denotes a may be tuple, that is uncertainty whether the tuple is in the relation. Intuitively this uncertain relation represents the three set of possible relation instances the first containing only a single tuple like Observer Bhakti saw Deer (Animal) on date 25-09-11 and time is 11:25 AM at Chhambal and the other two containing tuples and differing on AnimalName like Observer Bhakti saw Elephant (Animal) on date 25-09-11 and time is 14:27 AM and also next day Bhakti saw Zibra on same time, that is uncertainty for observer to storing record same animal on same time or day-date.

Data Model
To represent uncertainty has many ways, ranging from alternative values for attributes to rich constraint languages.

A data model M is complete if any finite set of relation instances to given schema can be represented by an uncertain relation in M. Consider the following three instances, in which Bhakti saw either Zibra, an Elephant or both. Note that we would need two separate tuples for Zibra and Elephant; otherwise we would not get the third instance, which has two tuples. Both of these tuples would have to be marked * - to get the two instances with one tuple each, However, then the empty relation (No animal Sighted) would also be a possible instance, which we did not mean to include. Now at last after that Closure is the condition that formalizes the existence of relation in a model M.A Model M is said to be closed under an operation if performing operation on any set of uncertain relations in M results in an uncertain relation that can be represented in data model M. In fact, when data model is closed under operation a reasonable implementation would compute operation directly on relation and not through the set of possible instances as represent by the implementation certain instance relation to uncertain instance relation. The next example show that models consisting of attributes and maybe tuples are not closed under certain operations. For instance, we have the following sighting of either tiger or lion:

Observer: Hetvi, AnimalName: tiger, lion, When: 28-09-11, Where: Gir, Time: 10:36AM (approx)

And the following relevant tuples in AnimalData relation:

AnimalName: Tiger, Color: White, Size: Medium

AnimalName: Lion, Color: Brown, Size: Large

If we perform a natural join of these two relations there are two possible instances in result:

First Possible Instance:
Observer: Hetvi, When: 28-09-11, Where: Gir, Time: 10:36AM (approx), AnimalName: Tiger, Color: White, Size: Medium.

Second Possible Instance:

Observer: Hetvi, When: 28-09-11, Where: Gir, Time: 10:36AM (approx), AnimalName: Lion, Color: Brown, Size: Large.

Using types of uncertainty we have looked at so far attributes or and maybe tuples there is no way to represent that exactly

Volume : 2 | Issue : 2 | November 2012 | ISSN - 2249-555X

and perfect one of these two tuples exists.

Consider the same sighing tuple from the previous example but now as a contrived example for illustrative purpose, suppose the AnimalData relation contains:

AnimalName: Tiger, Color: White, Size: Medium

AnimalName: Tiger, Color: Yellow, Size: Small

Now the natural join produces the following two instances – the empty instance is shown by displaying just the schema and no data:

First Possible Instance:

Observer: Hetvi, When: 28-09-11, Where: Gir, Time: 10:36AM (approx), AnimalName: Tiger, Color: White, Size: Medium.

Second Possible Instance:

Observer: Hetvi, When: 28-09-11, Where: Gir, Time: 10:36AM (approx), AnimalName: Tiger, Color: Yellow, Size: Small.

Again, using the types of uncertainty we have looked at so far, there is no way to represent that either both of the tuples exist or neither do.

REFERENCE (1) J. Widom. "Trio: A System for Integrated Management of Data, Accuracy, and Lineage". In: Proc. of Conf. on Innovative Data Systems Research (CIDR), 2005. | (2) T. J. Green and V. Tannen. "Models for Incomplete and Probabilistic Information". In: Proc. of IIDB Workshop, 2006. | (3) T. Imielinski and W. Lipski. "Incomplete Information in Relational Databases". Journal of the ACM, Vol. 31, No. 4, 1984. | (4) L. Libkin and L. Wong. "Semantic representations and query languages for or-sets". In: Proc. of ACM Symp. on Principles of Database Systems (PODS),1993. |