Modelling and Representation of Uncertain Temporal Information

P. Chountas and I. Petrounias
UMIST, Manchester, UK

The main task of an information system is the representation and management of large amounts of indicative information from multiple sources describing the state of some enterprise. Most conceptual and database models represent enterprises with no imprecise data. Very few approaches in the literature are dealing with imprecise data. In temporal data models and databases, approaches are dealing mainly with precise absolute times. Little consideration has been given to imprecise absolute times or infinite absolute times. No consideration has been given to imprecise infinite absolute times either. Many algebraic models are dealing with temporal or value imperfection, with no description of semantics of uncertain information. There is a need for conceptual models that capture the essential semantics of data imperfection, belief, and the temporal nature of imperfect information. This paper is proposing a conceptual framework that describes the semantics of temporal and incomplete information, and shows how this can be translated to database representation.

Keywords: Conceptual modelling; Factual and temporal uncertainty; Multiple information sources; Nested relational databases

1. Introduction

Imperfect information is the partial knowledge of the true value of the real world. It is essentially an epistemic property caused by lack of information. A major source of uncertainty is imprecision, since an imprecise bit of information entails a low level of certainty and vice versa.

The elements of the enterprise ontology involved in information imperfection, in conjunction with their semantic meaning and nature, are:

- the might happen ability of things, or the tendency of things to occur;
- the concept of time;
- the information source or provider.

Most conceptual and database models represent enterprises that are crisp, with no imprecise data. A crisp enterprise is defined as one that is highly quantifiable; all relationships are fixed, and all attributes are atomic valued. The premises on which this paper is based are precise enterprises, where data is imperfect. In such cases information can be certain, imprecise or uncertain, temporal, or any possible combination of each two of them, depending on the application domain.

A database is a description model of the enterprise world. Like any other description model it attempts to represent an abstracted version of the enterprise reality, and the level of the abstraction is usually determined by the expected applications. It is recognised that multiple information sources or providers, either internal or external in relation to the enterprise environment, will be present, propelling the information flow and obviously intervening in the final output of enterprise decision-making mechanisms. More than one source might be describing the same fragment of the enterprise world, expressing logical views about the same enterprise facts that are defined over the dimensional space of belief and time (multidimensionality). In addition they may be doing so independently or in conjunction with each other. Under such a dynamic environment it is practically inevitable that information would be inharmonious and unstable. Reducing the multidimensionality of a fact may result in loss of
information or uncertainty. Furthermore, a source itself is carrying a degree of credibility, which determines the conclusive belief of an independent observer, for the particular multidimensional facts embed by this particular source.

Many real-world phenomena require the explicit representation of both uncertain facts and uncertain time. Taking a measurement of some physical quantity, for example, can be both imprecise with respect to the time of the measurement and to the measurement itself. Another example is in the medical domain, where a patient may give details about symptoms of a disease previously suffered which are both imprecisely dated as well as imprecisely observed. In the above examples two different kinds of imperfection are derived: factual uncertainty and temporal uncertainty. Whereas factual uncertainty is about evidence available for a proposition, temporal uncertainty is about its occurrence in time. Temporal and factual uncertainty are independent. For example, someone may have symptoms typical of a certain disease, while the time of occurrence of those symptoms does not support the presupposed disease.

In the literature [1–4] most approaches that address uncertainty in temporal reasoning, represent crisp propositions, which may occur at an uncertain time. Others [5–9] consider propositions, which may be uncertain, while the time related to those propositions is not. Time serves as an index to the probability or possibility of propositions. No approach considers certain or uncertain propositions that may be repeated over a fixed but arbitrary calendar, with either an absolute or inconstant characteristic frequency, as a part of a multi-source environment [10,11]. From a modeling point of view, metamodels that contain all the necessary knowledge about the concepts, semantics and rules of uncertain information are absent from the literature.

This paper is proposing a novel approach to temporal representation and reasoning dealing with both uncertain facts and uncertain temporal information at the metamodel level. A point of particular concern is that temporal uncertainty should not influence factual imperfection. Decomposition does not prevent time and facts having a mutual influence. Temporal reasoning is exactly about the relation between time and facts.

The rest of the paper is organised as follows. Section 2 discusses relevant work in this area. Section 3 provides details on the semantics of temporal information and Section 4 discusses the interval-based finite time representation. Section 5 presents a conceptual modeling formalism to capture temporal and factual uncertainty and defines the basic elements for a temporal representation. Section 6 defines the metamodel for the proposed formalism and Section 7 describes the instance level. Section 8 presents a recursive relational algebra and transformations as a part of the standard select operator for fast searching, through the time hierarchy. Section 9 summarises the main points of the proposed approach and points out further research.

2. Classification of Relevant Work

In the literature, most approaches that address uncertainty start from either the perspective of formal temporal theories [1,2] or probabilistic logic [3,5,6]. Very few are trying to express and interpret uncertainty as part of a conceptual formalism.

Many applications require the explicit representation of both uncertain facts and uncertain time. Relevant work found in the uncertainty reasoning literature is judged by the following criteria:

- independent representation of factual uncertainty;
- independent representation of temporal uncertainty;
- representation of definite, indefinite, infinite information;
- representation of the factual and temporal uncertainty at the metamodel level.

Definite temporal information is defined when all times associated with facts are absolute time intervals. Indefinite temporal information is defined when the duration of a fact is indeterminate but it is constrained between a lower and upper bound of time points. Periodic or recurring (infinite) information is expressed when an infinite number of times are associated with a fact with either a constant or inconstant characteristic frequency of occurrence.

As part of a formal temporal theory, global and local inequality constraints on the occurrence time of a fact have been added to a temporal data model in Koubarakis [1,2]. The resulting model supports indefinite instants. An indefinite instant is a general instant that includes indeterminate instants, instants with disjoint sets of possible chronons, and instants with incompletely specified upper and lower bounds. The model proposed can conclude that $\alpha$ happened between $\beta$ and $\gamma$, $\beta < \alpha < \gamma$, where $\beta$, $\gamma$ are not explicitly defined. In this way indefinite temporal information is expressed encoded in the tuples. However, the model assumes that facts are certain and, therefore, there is no factual uncertainty. Furthermore, the model cannot represent infinite temporal uncertain or certain information.

In probabilistic logic uncertainty is addressed from two different perspectives. The first perspective considers propositions, which may be uncertain, while the time related to those propositions is not. That is, time is