Quaternary International xxx (2015) 1–22

Contents lists available at ScienceDirect

Quaternary International

journal homepage: www.elsevier.com/locate/quaint

Assessing the effects of temporal ambivalence on defining palaeosystem interrelations, and applicability to the analysis of archaeological survey data

Fernando Sánchez-Trigueros [a, b, *], Alfonso Benito-Calvo [c], Marta Navazo [c, d], Antoni Canals [b, e]

[a] The School of Geography, Garstang North, University of Leeds, Leeds, LS2 9JT, UK
[b] Departament d'Història i Història de l'Art, Universitat Rovira i Virgili (URV), Avinguda de Catalunya 35, 43002 Tarragona, Spain
[c] Centro National de Investigación en Evolución Humana (CENIEH), Paseo Sierra de Atapuerca 3, 09002 Burgos, Spain
[d] Departamento de Ciencias Históricas y Geografía, Universidad de Burgos (UBU), C/Villadiego s/n, 09001 Burgos, Spain
[e] IPHES, Institut Català de Paleoecologia Humana i Evolució Social, Campus Sescelades URV, C/Marcel.lí Domingo s/n, 43007 Tarragona, Spain

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Uncertainty
Epistemic ambivalence
Archaeological survey data
Data mining
Middle Palaeolithic
Sierra de Atapuerca

ABSTRACT

We present a possibility-based method with applications to the analysis of archaeological survey data under constraints of temporal ambivalence. The method is a generalizable heuristic for automated learning and data mining that estimates the set of possible responses associated with ambivalent numerical data, by (i) assessing the degree of epistemic ambivalence that is carried in a data set with ambivalent information, and by (ii) monitoring the effects of this degree on data analysis, enabling the adaptation of classical data analysis to the treatment of both crisp and fuzzy information. Owing to the strong methodological foundations of the application, a detailed introduction of the ambivalent paradigm is presented first so its adaptability to archaeological survey data can be followed later. The approach is demonstrated in selecting informative features of uncertain palaeosystems by ambivalent Principal Component Analysis (PCA), and exemplified with archaeological survey data collected in the Sierra de Atapuerca and the Arlanzón valley (Burgos, Spain), focusing on the effects of temporal ambivalence on the analysis of associations between lithological palaeoenvironments and Middle Palaeolithic technologies in an open-air context. Although survey data in this region is affected by varying degrees of temporal ambivalence that, on occasion, span several isotopic stages of the late Pleistocene sequence, its effects on feature selection, regarding PCA, are not significant enough to produce contradictory reprojections of palaeolithological data onto the principal space, thanks to a gradual stabilization of the relief in the Middle Arlanzón valley, on a regional scale, during the late Pleistocene. By applying correlation-based dissimilarity distances and cluster analysis to the grouping of ambivalent score patterns, ambivalent features that describe the palaeolithological context of the Middle Palaeolithic scatters can be grouped on the $PC_1 \times PC_2$ principal plane into four contexts with a similar palaeolithological evolution among group members: the Mesozoic crest of the Sierra de Atapuerca (Group 1), the Neogene levels of the Villalval-Rubena platform (Group 2), lithologically mixed areas between fluvial valleys and Neogene hills (Group 3), and the valleys of the Arlanzón, Pico and Vena rivers (Group 4). In none of these groups the evolution of the local palaeolandscape is notable enough, once again on a regional scale, to produce conflicting descriptions of the palaeolithological signatures where the archaeological scatters occur, although Group 3 (a geographical interface between different lithological palaeoenvironments) has been more dynamic than the rest of environments over the time range the lithic scatters would have formed.

* Corresponding author. The School of Geography, Garstang North, University of Leeds, Leeds, LS2 9JT, UK
E-mail address: f.sancheztrigueros@leeds.ac.uk (F. Sánchez-Trigueros).

http://dx.doi.org/10.1016/j.quaint.2015.09.089
1040-6182/© 2015 Elsevier Ltd and INQUA. All rights reserved.

1. Introduction

Uncertainty stems from an imperfect knowledge of the world. Tacking account of uncertain information entails the awareness that missing information, inaccuracies, imprecisions and contradictions may be contained in observations about real-world phenomena that are brought by factors such as the complex and even unpredictable nature of many real-world systems, limitations and defects of the data collection process, transmission of spurious data and loss of information content. Whatever the cause, the consequence of uncertainty is always a loss of assurance with respect to the characteristics of the entities under study.

This paper revolves around this problem and considers general aspects of uncertainty theory and the analysis of uncertain data in the study of dynamic paleosystems, that is, evolving systems that ceased to exist in the past, can just be indirectly observed by their material remains, and may be modeled based on deduction, experimentation and uniformitarian premises. Quite naturally, palaeosystems modeling has become a prime line of research among sciences addressing historical reconstruction, particularly in archaeology, palaeontology, palaeoecology, geological history and forensics, provided they are typically faced with the problem of finding interrelations between elements of extinct systems, but primarily supported by ambiguous knowledge on the exact occurrence of these elements in space-time. The theoretical framework we consider, in other words, addresses the problem of describing associations between events and the environmental context they have occurred, where the date of the event is not known with certainty or is not infinitesimal (i.e., it is a time span) and the context it interacted with is assumed to have changed in time (e.g., an archaeological deposit or the traces of a crime).

Determining such a domain of feasible associations soon becomes a *big data* problem as the size of the domain grows exponentially with the degree of uncertainty in the associations, so its full description is intractable in most cases. To tackle this tractability problem we present a data mining protocol that goes through data quality assessment, feature selection, feature extraction and pattern recognition in estimating the uncertainty carried by temporally ambivalent data, and demonstrate its suitability for the analysis of archaeological field survey data and the palaeolandscape contexts it may have related to, giving an in-depth definition of their methodological composition and assembly. Notably, an automated learning mechanism is detailed for exploring the universe of possible scenarios that derive from an ambivalent data set, based on iteratively retrieving samples of such scenarios until the amount of information gained with additional samples is no longer significant. Section 2 formalizes the concept of temporal ambivalence and describes a protocol for ambivalent analysis that includes this mechanism, while Section 3 demonstrates the ambivalence problem in the context of archaeological survey data analysis, showing its analytical behavior in studying land-use patterns associated with Middle Palaeolithic technologies in the Sierra de Atapuerca and the Arlanzón valley (Burgos, Spain).

2. Formal principles for the modeling of epistemic ambivalence and conceptual adaptations to the study of palaeosystems

2.1. An informal introduction to the notion of uncertainty

In defining uncertainty, we follow an ontology based on two properties of any entity or event, namely, (i) the direct observation of its actual state and (ii) the ambivalence or variability of this state, from which we define five sources of knowledge uncertainty: missing data, error, polysemy, ambiguity and imprecision (Fig. 1). Of

	True data is known	**True data is not known**
Single-valued data	Certainty	Missing data Error
Ambivalent data	Polysemy Ambiguity	Imprecision

Fig. 1. Factors that affect knowledge certainty (in bold) and classes of uncertainty.

these, sources with effects on the knowledge of single-valued observations are missing data and errors. *Missing data* affect the completeness of a data set as they encompass information about the state of the attributes in the objects that is lost, so it is a major source of concern in preparing the data view for exploratory data analysis, data mining and statistical modeling. *Error*, in a statistical sense, is a measure of the difference between observed states and the actual ones, whether systematic (i.e., "biased") or stochastic and so unable to be specified beforehand, in which case they can be modeled in probabilistic terms.

The key feature of an ambivalent observation is a range of states it can or may take. Depending on whether such states can be observed directly or not, we define the properties of polysemy, ambiguity and imprecision. In this taxonomy the concept of *polysemy* applies to either (a) multiple states happening along the lifespan of an occurrence or to (b) an occurrence that can be associated with alternative states (Fig. 2). Data *ambiguity* refers to the problem of considering conflicting systems for the classification of states (Foody, 2003). For example, on quantizing scalar distributions, given that setting the breaks between classes may be based on different rules, but also in the quantification of qualitative labels, as alternate scaling systems could apply. When faced with *imprecision* alternate states are observed, whether on the terms' or on the response's side or on both, although their actual existence is mutually exclusive so possible states need not map to actual existence.

2.2. Formalizing epistemic ambivalence from a possibilistic perspective

In the remainder of the paper we treat the concept of ambivalence in greater depth, and present a formal data model for ambivalent analysis. The first premise of this theoretical framework is that possible relations between phenomena (that is, relations that may happen or have happened) can be better modeled as fuzzy associations (as a range of possible scenarios, each associated with

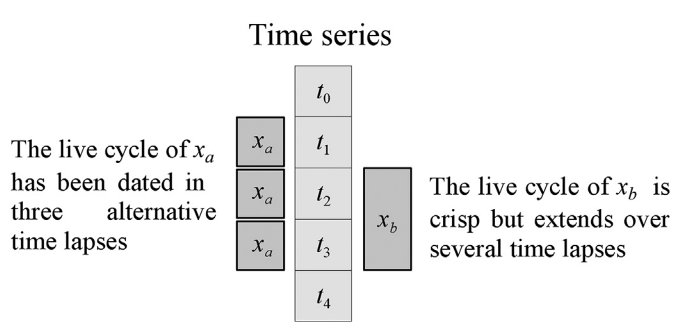

Fig. 2. Temporal ambivalence. Left: temporal imprecision. x_a could be dated at three mutually-exclusive points in time (each with a possibility grade). Right: temporal polysemy. x_b occurs in the course of a time range that spans lapses t_2 and t_3 of the time series.

Please cite this article in press as: Sánchez-Trigueros, F., et al., Assessing the effects of temporal ambivalence on defining palaeosystem interrelations, and applicability to the analysis of archaeological survey data, Quaternary International (2015), http://dx.doi.org/10.1016/j.quaint.2015.09.089