International Journal of Agricultural and Environmental Information Systems
Volume 13 • Issue 1

Decision Support Tool for the Agri-Food Sector Using Data Annotated by Ontology and Bayesian Network:
A Proof of Concept Applied to Milk Microfiltration

Cédric Baudrit, I2M, University of Bordeaux, INRAE, Bordeaux, France
Patrice Buche, IATE, University of Montpellier, INRAE, Institut Agro, Montpellier, France*
Nadine Leconte, STLO, INRAE, Institut Agro, Rennes, France
Christophe Fernandez, I2M, University of Bordeaux, INRAE, Bordeaux, France
Maëllis Belna, Boccard, Research and Development, F-35360 Montauban-de-Bretagne, France
Geneviève Gésan-Guiziou, STLO, INRAE, Institut Agro, Rennes, France

ABSTRACT

The scientific literature is a valuable source of information for developing predictive models to design decision support systems. However, scientific data are heterogeneously structured expressed using different vocabularies. This study developed a generic workflow that combines ontology, databases, and computer calculation tools based on the theory of belief functions and Bayesian networks. The ontology paradigm is used to help integrate data from heterogeneous sources. Bayesian network is estimated using the integrated data taking into account their reliability. The proposed method is unique in the sense that it proposes an annotation and reasoning tool dedicated to systematic analysis of the literature, which takes into account expert knowledge of the domain at several levels: ontology definition, reliability criteria, and dependence relations between variables in the BN. The workflow is assessed successfully by applying it to a complex food engineering process: skimmed milk microfiltration. It represents an original contribution to the state of the art in this application domain.

KEYWORDS

Bayesian Network, Data Integration, INRAE, Knowledge Base, Knowledge Integration, Milk Microfiltration, Ontology, Reliability, Uncertainty

1. INTRODUCTION

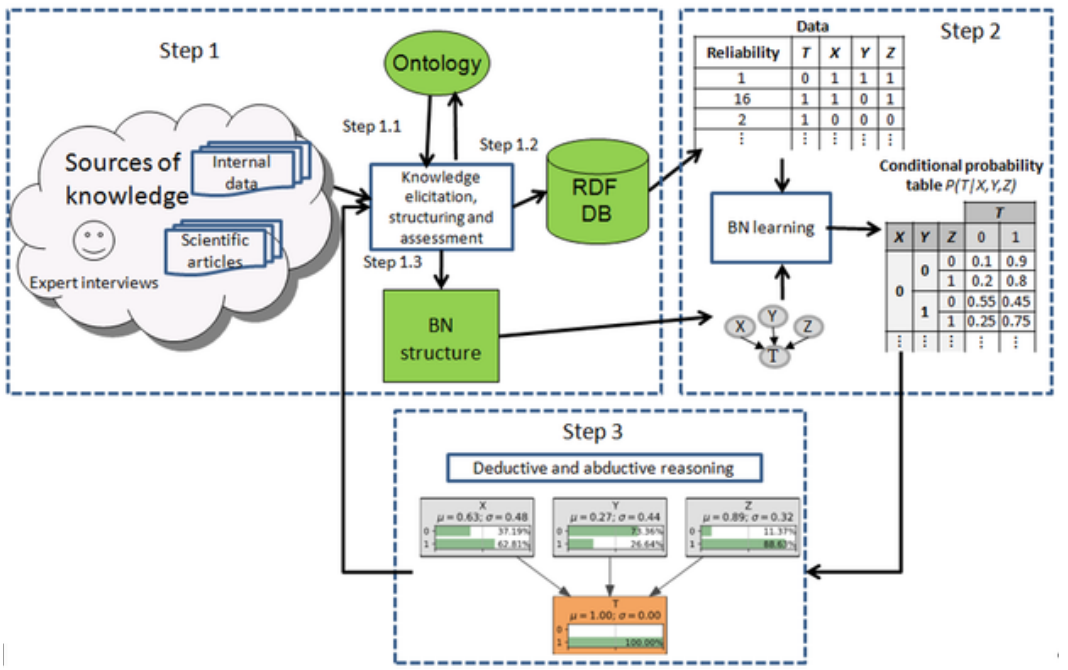
For decision tasks such as optimising food processes, an initial step is to predict variables of interest from process parameters. The scientific literature, including experimental data and knowledge expressed by domain experts, is a valuable source of information to reach this goal. However, the ever-increasing amount of scientific data is heterogeneously structured, found mainly in text format and expressed using different vocabularies. Addressing this difficulty requires innovative tools that

DOI: 10.4018/IJAEIS.309136 *Corresponding Author

can integrate and treat new information. In this context, using Semantic Web methods based upon ontologies seem relevant to structure experimental information (Lousteau-Cazalet et al. 2016; Yeumo et al. 2017; Aubin et al. 2019). As experiments use different methods and technologies, another difficulty is considering source (document) reliability when using the data in calculations. The theory of belief functions provides suitable solutions to address this issue (Destercke et al. 2013). Providing relevant conclusions and recommendations requires developing adequate modelling tools that can integrate, as much as possible, available knowledge which is heterogeneous in nature and quality. Such modelling tools must be able to manage heterogeneous sources of knowledge (experimental data and expert opinion), multiple manipulated scales and different forms of uncertainty (Perrot et al. 2016; Barnabe et al. 2018). With this goal in mind, Bayesian networks (BNs) (Jensen and Nielsen, 2007; Pearl, 1988) provide a practical mathematical structure that can describe complex systems which contain uncertainty. BNs are based on a coupling between graph and probability theory in which the graph provides an intuitively appealing interface with which model designers can represent strongly interacting sets of variables. Uncertainty in the system is considered by quantifying the dependence between variables in the form of conditional probabilities. The use of BNs has been investigated recently in agri-food domains (Baudrit et al. 2015; Drury et al. 2017; Chapman et al. 2018).

This article discusses a numerical workflow to treat data and knowledge that combines ontologies, databases and computer calculation tools based on the theory of belief functions and BNs. The workflow developed is based on a pluridisciplinary collective study involving experts in the domains of food processing and artificial intelligence, and comprises three sequential steps (see Fig. 1). The first step consists of elicitation, structuring and assessment of knowledge related to a food process of interest. More precisely, experimental data published in scientific articles are annotated using an ontology, and their reliability is assessed by experts in food processing. Data from scientific articles are annotated in a simple tabular format file that is semi-automatically generated using the ontology (see step 1.1 in Fig. 1). Then, in step 1.2, the file is uploaded and annotated data are stored in a Resource Description Framework (RDF) database. The complete annotation data set used in this

Figure 1. Workflow process developed in this study. RDF: Resource Description Framework, DB: database, BN: Bayesian network.



Trusted Integration of Wireless Sensor Networks into Herd Control Monitoring Application

Laurent Gomezand Michel Riveill (2013). *International Journal of Agricultural and Environmental Information Systems (pp. 1-24).*
www.irma-international.org/article/trusted-integration-of-wireless-sensor-networks-into-herd-control-monitoring-application/102942