International Journal of Approximate Reasoning 52 (2011) 501–518

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

Interpretability assessment of fuzzy knowledge bases: A cointension based approach

C. Mencar *, C. Castiello, R. Cannone, A.M. Fanelli

Department of Informatics, University of Bari, Bari 70125, Italy

ARTICLE INFO

Article history:
Received 26 April 2010
Revised 18 November 2010
Accepted 19 November 2010
Available online 30 November 2010

Keywords:
Computing with words
Cointension
Interpretability
Fuzzy rule-based classifier
Boolean logic
Logical view

ABSTRACT

Computing with words (CWW) relies on linguistic representation of knowledge that is processed by operating at the semantical level defined through fuzzy sets. Linguistic representation of knowledge is a major issue when fuzzy rule based models are acquired from data by some form of empirical learning. Indeed, these models are often requested to exhibit interpretability, which is normally evaluated in terms of structural features, such as rule complexity, properties on fuzzy sets, partitions and so on. In this paper we propose a different approach for evaluating interpretability that is based on the notion of cointension. The interpretability of a fuzzy rule-based model is measured in terms of cointension degree between the explicit semantics, defined by the formal parameter settings of the model, and the implicit semantics conveyed to the reader by the linguistic representation of knowledge. Implicit semantics calls for a representation of user's knowledge which is difficult to externalise. Nevertheless, we identify a set of properties – which we call "logical view" – that is expected to hold in the implicit semantics and is used in our approach to evaluate the cointension between explicit and implicit semantics. In practice, a new fuzzy rule base is obtained by minimising the fuzzy rule base through logical properties. Semantic comparison is made by evaluating the performances of the two rule bases, which are supposed to be similar when the two semantics are almost equivalent. If this is the case, we deduce that the logical view is applicable to the model, which can be tagged as interpretable from the cointension viewpoint. These ideas are then used to define a strategy for assessing interpretability of fuzzy rule-based classifiers (FRBCs). The strategy has been evaluated on a set of pre-existent FRBCs, acquired by different learning processes from a well-known benchmark dataset. Our analysis highlighted that some of them are not cointensive with user's knowledge, hence their linguistic representation is not appropriate, even though they can be tagged as interpretable from a structural point of view.

1. Introduction

Computing with words (CWW) is a recent paradigm for processing information through words instead of numbers [1–3]. Its inception is justified by the need of providing for a tool to represent and manipulate knowledge in linguistic form rather than numerical. The ability of manipulating words and inferring knowledge that can be expressed in linguistic forms gives added value to intelligent systems, because of their wider applicability in several real-world contexts in which accuracy is not the main (or the only) concern. Hence, CWW may be beneficial in human-centric fields such as medicine, psychology, economics and linguistics, where a main concern is the need of communicating knowledge to users.

* Corresponding author.
E-mail addresses: mencar@di.uniba.it (C. Mencar), castiello@di.uniba.it (C. Castiello), raffaelecannone@di.uniba.it (R. Cannone), fanelli@di.uniba.it (A.M. Fanelli).

0888-613X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2010.11.007

Historically, the approach used for designing expert systems is mainly based on symbolic representation of knowledge derived by a syntactic inference process. Indeed an appropriate choice of symbols leads to a linguistic – usually structured – representation that can be read by users. On the other hand, CWW operates at the semantic level, i.e. inference is carried out by taking into account the semantic definition of words (and their connectives). These two approaches are equivalent in the case of Boolean semantics, i.e. when words refer to sets of objects and connectors correspond to set operations. Unfortunately, natural language is very far from being expressible as Boolean expressions: fuzziness is an intrinsic feature of most linguistic terms that do not belong to mathematics. As a consequence, the utility of expert systems is strictly limited in contexts where knowledge can be expressed by extremely sharp concepts. CWW comes into play when there is the need of representing and manipulating knowledge where fuzziness is an important feature. Fuzziness, indeed, can be captured by using fuzzy sets, and fuzzy knowledge can be manipulated by fuzzy set theory (FST), which is the formal underpinning of CWW.

The greatest enhancement of CWW over FST is mainly methodological. In CWW emphasis is put on the representation of knowledge, while FST is used as a tool for knowledge manipulation and inference. Hence, interpretability of knowledge is a necessary requirement in CWW. CWW does not directly address interpretability, being focused on the structures and rules for representing and manipulating knowledge. However, granting interpretability is a main issue when designing CWW-based systems, since the lack of interpretability damages all the benefits of CWW. Without interpretability of knowledge, purely numerical methods can be an effective alternative.

Interpretability is an ill-posed property, as there are several main issues to be addressed concerning its definition, its achievement and its assessment. Whilst several works in literature try to capture the property of interpretability with a collection of constraints (both crisp and fuzzy), here we adopt a more general approach by considering the notion of "cointension", firstly defined by Zadeh in [4]. Roughly speaking, cointension can be viewed as a relation between concepts such that two concepts are cointensive if they refer to almost the same object. Thus, in our view, a knowledge base is interpretable if its semantics is cointensive with the knowledge acquired by the user.

The point of departure of our approach relies on a close involvement of user understanding in the definition of interpretability. This could allow for a more effective assessment of interpretability with respect to the structural approach based on interpretability constraints, whose fulfilment cannot assure that the knowledge base is actually cointensive with user's knowledge. Indeed, interpretability constraints are mainly based on common-sense and there is no agreement on the minimal set of constraints to be used. Finally, there is no guarantee that any set of constraints could be an exhaustive characterisation of interpretability.

By proposing an approach based on interpretability as cointension, we try to cast interpretability, intended as "ability to read and understand", in the realm of semantics. Thus, assessing interpretability mainly concerns a comparison between the semantics of a knowledge base and the semantics of the knowledge acquired by a user after reading and understanding the knowledge base. The semantics of a knowledge base (we call it *explicit semantics*) is completely specified by the involved fuzzy sets and operators used for inference. On the other hand, the semantics acquired by users when reading the knowledge base (we call it *implicit semantics*) is much more difficult to externalise. Nevertheless, common features can be identified for both implicit and explicit semantics. By exploiting such common features it is possible to analyse cointension and hence interpretability.

In assessing interpretability of knowledge bases that are represented in a linguistic form, we identify a common feature shared by implicit and explicit semantics derived from a fuzzy knowledge base which we call "logical view". The logical view is here intended as the set of properties of propositional calculus which are assumed to hold both in implicit and explicit semantics. The validity of the logical view is tested through the application of a minimisation algorithm on the fuzzy knowledge base. By testing the validity of the logical view in the explicit semantics, we are able to assess the interpretability of a fuzzy knowledge base in the sense of cointension. We limit our argumentation to fuzzy rule-based classifiers (FRBCs).

The paper is organised as follows: in the next section, the problem of defining and assessing interpretability is analysed. In Sections 3 and 4 the proposed approach for interpretability assessment is described in detail. Successively, in Section 5 we report an experimentation concerning ten knowledge bases for the same classification problem (Wine data) obtained through a system that preserves a number of interpretability constraints. In the conclusive part of the paper we address points of strength and weakness of the proposed approach, along with some remarks on future developments. Additionally, for a self-contained discussion the paper is completed with an appendix containing the description of the minimisation algorithm.

2. The problem of interpretability assessment

Interpretability assessment should be regarded as a major issue in the field of fuzzy knowledge-based system modelling. However, a proper evaluation of interpretability appears to be a controversial problem, since the definition of interpretability eludes any formal characterisation.

2.1. Interpretability definition

In [5], an attempt to provide a definition of interpretability is made, in the form of the following "Comprehensibility Postulate" (CP):