International Journal of Approximate Reasoning 55 (2014) 1147–1163

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

A label semantics approach to linguistic hedges

Martha Lewis *, Jonathan Lawry

Department of Engineering Mathematics, University of Bristol, BS8 1TR, United Kingdom

ARTICLE INFO

Article history:
Received 28 August 2013
Received in revised form 28 January 2014
Accepted 31 January 2014
Available online 4 February 2014

Keywords:
Linguistic hedges
Label semantics
Conceptual spaces
Random sets

ABSTRACT

We introduce a model for the linguistic hedges 'very' and 'quite' within the label semantics framework, and combined with the prototype and conceptual spaces theories of concepts. The proposed model emerges naturally from the representational framework we use and as such, has a clear semantic grounding. We give generalisations of these hedge models and show that they can be composed with themselves and with other functions, going on to examine their behaviour in the limit of composition.

1. Introduction

The modelling of natural language relies on the idea that languages are compositional, i.e. that the meaning of a sentence is a function of the meanings of the words in the sentence, as proposed by [13]. Whether or not this principle tells the whole story, it is certainly important as we undoubtedly manage to create and understand novel combinations of words. Fuzzy set theory has long been considered a useful framework for the modelling of natural language expressions, as it provides a functional calculus for concept combination [30,32].

A simple example of compositionality is hedged concepts. Hedges are words such as 'very', 'quite', 'more or less', 'extremely'. They are usually modelled as transforming the membership function of a base concept to either narrow or broaden the extent of application of that concept. So, given a concept 'short', the term 'very short' applies to fewer objects than 'short', and 'quite short' to more. Modelling a hedge as a transformation of a concept allows us to determine membership of an object in the hedged concept as a function of its membership in the base concept, rather than building the hedged concept from scratch [31].

Linguistic hedges have been widely applied, including in fuzzy classifiers [6,7,20,22] and database queries [1,3]. Using linguistic hedges in these applications allows increased accuracy in rules or queries whilst maintaining human interpretability of results [4,23]. This motivates the need for a semantically grounded account of linguistic hedges: if hedged results are more interpretable then the hedges used must themselves be meaningful.

In the following we provide an account of linguistic hedges that is both functional, and semantically grounded. In its most basic formulation, the operation requires no additional parameters, although we also show that the formulae can be generalised if necessary. Our account of linguistic hedges uses the label semantics framework to model concepts [17]. This is a random set approach which quantifies an agent's subjective uncertainty about the extent of application of a concept. We refer to this uncertainty as *semantic uncertainty* [19] to emphasise that it concerns the definition of concepts and categories, in contrast to stochastic uncertainty which concerns the state of the world. In [19] the label semantics approach is combined

* Corresponding author.
E-mail addresses: martha.lewis@bristol.ac.uk (M. Lewis), j.lawry@bristol.ac.uk (J. Lawry).

http://dx.doi.org/10.1016/j.ijar.2014.01.006

with conceptual spaces [14] and prototype theory [25], to give a formalisation of concepts as based on a prototype and a threshold, located in a conceptual space. This approach is discussed in detail in Section 2. An outline of the paper is then as follows: Section 3 discusses different approaches to linguistic hedges from the literature, and compares these with our model. Subsequently, in Section 4, we give formulations of the hedges 'very' and 'quite'. These are formed by considering the dependence of the threshold of a hedged concept on the threshold of the original concept. We give a basic model and two generalisations, show that the models can be composed and investigate the behaviour in the limit of composition. Section 5 compares our results to those in the literature and proposes further lines of research.

2. Theoretical approach to concepts

2.1. Prototype theory and fuzzy set theory

Prototype theory views concepts as being defined in terms of prototypes, rather than by a set of necessary and sufficient conditions. Elements from an underlying metric space then have graded membership in a concept depending on their similarity to a prototype for the concept. There is some evidence that humans use natural categories in this way, as shown in experiments reported in [25]. Fuzzy set theory [30] was proposed as a calculus for combining and modifying concepts with graded membership, and extended these ideas in [32] to linguistic variables as variables taking words as values, rather than numbers. For example, 'height' can be viewed as a linguistic variable taking values 'short,' 'tall', 'very tall', etc. The variable relates to an underlying universe of discourse Ω, which for the concept 'tall' could be $\mathbb{R}^+$. Then each value L of the variable is associated with a fuzzy subset of Ω, and a function $\mu_L : \Omega \to [0, 1]$ associates with each $x \in \Omega$ the value of its membership in L. Prototype theory gives a semantic basis to fuzzy sets through the notion of similarity to a prototype, as described in [10]. In this context, concepts are represented by fuzzy sets and membership of an element in a concept is quantified by its similarity to the prototype. In this situation the fuzziness of the concept is seen as inherent to the concept. An alternative interpretation for fuzzy sets is random set theory, see [10] for an exposition. Here, the fuzziness of a set comes from uncertainty about a crisp set, i.e. semantic uncertainty, rather than fuzziness inherent in the world. This second approach is the stance taken by [19], and which we now adopt in this paper.

2.2. Conceptual spaces

Conceptual spaces are proposed by Gärdenfors in [14] as a framework for representing information at the conceptual level. Gärdenfors contrasts his theory with both a symbolic, logical approach to concepts, and an associationist approach where concepts are represented as associations between different kinds of basic information elements. Rather, conceptual spaces are geometrical structures based on quality dimensions such as weight, height, hue, brightness, etc. It is assumed that conceptual spaces are metric spaces, with an associated distance measure. This might be Euclidean distance, or any other appropriate metric. The distance measure can be used to formulate a measure of similarity, as needed for prototype theory – similar objects are close together in the conceptual space, very different objects are far apart.

To develop the conceptual space framework, Gärdenfors also introduces the notion of integral and separable dimensions. Dimensions are integral if assignment of a value in one dimension implies assignment of a value in another, such as depth and breadth. Conversely, separable dimensions are those where there is no such implication, such as height and sweetness. A *domain* is then defined as a set of quality dimensions that are separable from all other dimensions, and a *conceptual space* is defined as a collection of one or more domains.

Gärdenfors goes on to define a *property* as a convex region of a domain in a conceptual space. A *concept* is defined as a set of such regions that are related via a set of salience weights. This casting of (at least) properties as convex regions of a domain sits very well with prototype theory, as Gärdenfors points out. If properties are convex regions of a space, then it is possible to say that an object is more or less central to that region. Because the region is convex, its centroid will lie within the region, and this centroid can be seen as the prototype of the property.

2.3. Label semantics

The label semantics framework was proposed by [17] and related to prototype theory and conceptual spaces in [19]. In this framework, agents use a set of labels $LA = \{L_1, L_2, \ldots, L_n\}$ to describe an underlying conceptual space Ω which has a distance metric $d(x, y)$ between points. In fact, it is sufficient that $d(x, y)$ be a pseudo-distance. When x or y is a set, say Y, we take $d(x, Y) = \min\{d(x, y)\colon\ y \in Y\}$. In this case, the set Y is seen as an ontic set, i.e., a set where all elements are jointly prototypes, as opposed to an epistemic set describing a precise but unknown prototype, as described in [11]. Each label L_i is associated with firstly a set of prototype values $P_i \subseteq \Omega$, and secondly a threshold ε_i, about which the agents are uncertain. The thresholds ε_i are drawn from probability distributions δ_{ε_i}. Labels L_i are associated with neighbourhoods $\mathcal{N}_{L_i}^{\varepsilon_i} = \{x \in \Omega\colon\ d(x, P_i) \leqslant \varepsilon_i\}$. The neighbourhood can be seen as the extension of the concept L_i. The intuition here is that ε_i captures the idea of being sufficiently close to prototypes P_i. In other words, $x \in \Omega$ is sufficiently close to P_i to be appropriately labelled as L_i providing that $d(x, P_i) \leqslant \varepsilon_i$.

Given an element $x \in \Omega$, we can ask how appropriate a given label is to describe it. This is quantified by an appropriateness measure, denoted $\mu_{L_i}(x)$. We are intentionally using the same notation as for the membership function of a fuzzy set. This quantity is the probability that the distance from x to P_i, the prototype of L_i, is less than the threshold ε_i, as given by: