Expert Systems with Applications 37 (2010) 158–164

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Obtaining the consensus and inconsistency among a set of assertions on a qualitative attribute

Adolfo Guzman-Arenas *, Adriana Jimenez-Contreras

Centro de Investigación en Computación (CIC), Instituto Politécnico Nacional (IPN), 07738 Mexico City, DF, Mexico

ARTICLE INFO

Keywords:
Confusion
Inconsistency
Consensus
Hierarchy
Qualitative value
Inconsistency measurement
Outlier
Assertion

ABSTRACT

It is well understood how to compute the average or centroid of a set of numeric values, as well as their variance. In this way we handle inconsistent measurements of the same property. We wish to solve the analogous problem on qualitative data: How to compute the "average" or consensus of a set of affirmations on a non-numeric fact, as reported for instance by different Web sites? What is the most likely truth among a set of inconsistent assertions about the same attribute?

Given a set (a bag, in fact) of statements about a qualitative feature, this paper provides a method, based in the theory of confusion, to assess the most plausible value or "consensus" value. It is the most likely value to be true, given the information available. We also compute the *inconsistency* of the bag, which measures how far apart the testimonies in the bag are. All observers are equally credible, so differences arise from perception errors, due to the limited accuracy of the individual findings (the limited information extracted by the examination method from the observed reality).

Our approach differs from classical logic, which considers a set of assertions to be either consistent (True, or 1) or inconsistent (False, or 0), and it does not use Fuzzy Logic.

1. Previous work and problem statement

Assume several measurements are performed on the same property (for instance, the length of a table). One measurer some distance away asserted "3 m." Another person with the help of a meter said "3.13 m". A lady with a micrometer reported "3.1427 m". From these, it is possible to obtain the most likely value ($\mu = 3.09$ m, the average length) as well as the dispersion of these measurements (σ, the variance), perhaps disregarding some outliers. For quantitative measurements we know how to take into account contradicting facts, and we do not regard them necessarily as inconsistent. We just assume that the observers' gauges have different precisions or accuracies.

Let us now consider several asseverations on a single-valued non-numeric variable (such as *the killer is*) that ranges on qualitative values (such as dog, cat, German Shepherd, Schnauzer) that can be arranged in a hierarchy (Fig. 1). That is, observer 1 reports that the killer is a dog, observer 2 reports that the killer is a cat . . . Can we find the consensus value or most likely value for the assassin? The "centroid" or "average" of the reported animals?[1] Or, we know that Ossama Bin Laden is reported to hide in {Afghanistan; Beirut; Irak; Kabul; Middle East; Afghanistan; Syria}. What is the most likely value to be true? Intuitively, this is the value that minimizes the sum of disagreements or discomforts for all the observers when they learn of the value chosen as the consensus value.

Section 2 of the paper tells us how to measure the discomfort that an observer has when using *r* instead of (his reported value) *s*. Section 3 of this paper solves the following:

Problem 1. Given a bag[2] of statements reporting non-numeric values, what is the most plausible value? How can me measure their inconsistency?

In Problem 1 we assume that all observers are equally credible, so the discrepancy in observed values is due only to inaccuracy in observations. Section 4 solves Problem 1 in the presence of negative findings (negative assertions).

1.1. Previous work

The Plausibility Theory of Dempster–Shafer (Dempster, 1968; Shafer, 1979) solves Problem 1 assuming that each observer has a given *confidence*, that their findings are independent – they do not influence each other, and that all observers have the same precision. We assume, instead, that all observers have the same confidence. But, in distinction to Plausibility Theory, the discrepancy in

* Corresponding author. Tel.: +52 (55) 5595 5075; fax: +52 (55) 5658 1250.
E-mail addresses: a.guzman@acm.org (A. Guzman-Arenas), dyidyia@yahoo.com (A. Jimenez-Contreras).

[1] We shall assume that only one value is possible – no two or more killers in our example.

[2] A *bag* is a set where repeated elements are allowed.

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.05.010

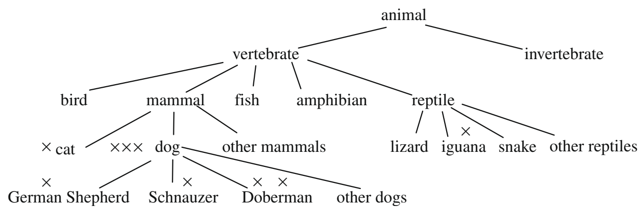

Fig. 1. A hierarchy of symbolic values. It is a tree where every node is either a symbolic value or, if it is a set, then its descendants form a partition. Hierarchies make possible to compute the confusion conf (r, s) that results when value r is used instead of s, the true or intended value. The confusion (Section 2) is the number of *descending* links in the path from r to s, divided by the height of the hierarchy. For instance, conf(dog, Doberman) = 1/4, conf(Doberman, dog) = 0, conf(Doberman, German Shepherd) = 1/4, conf(Doberman, iguana) = 2/4, conf(iguana, Doberman) = 3/4. conf $\in [0, 1]$. Refer to Section 2. Values marked with $\times$ refer to Section 3.

the values reported is due to the different methods used to perform the examinations (observer 1 saw the killer at a distance, observer 2 saw it at night, observer 3 heard it bark...).

Logic solves this problem by

(a) Declaring that, since dog $\neq$ cat $\neq$ Doberman $\neq \cdots$, the set is inconsistent, and the sentence (the killer is a dog) $\wedge$ (the killer is a cat) $\wedge \cdots$ evaluates to F; no agreement is possible.[1] This approach is unacceptable since, in practice, there are "small" inconsistencies, such as that between "It is a Doberman" and "It is a Female Doberman".
(b) Postulating a (small) set of predicates that must all be true (Byrne & Hunter, 2005) for this set of observations to be consistent, and declaring that the degree of inconsistency of the set is the percentage of predicates that become false. This is unacceptable since this set of predicates can be shortened (by ANDing some of the predicates) or lengthened (by dividing a complex predicate in parts), thus artificially varying the amount of inconsistency measured. This solution is syntax-sensitive.
(c) "Counting the minimal number of formulae needed to produce the inconsistency in a set of formulae. This idea rejects the possibility of a more fine-grained inspection of the (content of the) formulae. In particular, if one looks to singleton sets only, one is back to problem to the initial problem (a), with only two values: consistent or inconsistent (Knight, 2001)".[3]
(d) "Looking at the proportion of the language that is touched by the inconsistency of a set of formulae. This allows us to look *inside* the formulae (Konieczny, Lang, & Marquis, 2003)".[3] Disadvantage: two formulae can have different inconsistency measures. It is not sensitive to the syntax of the formulae.
(e) Measuring inconsistency through minimal inconsistent sets (Hunter & Konieczny, 2008). Here, subsets that are minimally inconsistent are defined and considered as the "relevant sets" that measure inconsistency. This approach does not take into account that often it is possible to perceive degrees of inconsistency among two logical *constants* (Doberman, Dog, Mammal, Iguana). That is, Doberman is more different (more inconsistent, informally speaking) to Iguana than to Dog. Function conf of Section 2 quantifies this.
(f) Using some kind of high-order Logic, such as para-consistent logic or non-monotonic logic.

[3] Citations are from Hunter and Konieczny (2008).

Fuzzy Logic does not by itself solve Problem 1. It can be used to give fuzzy confidence values to observers, and then fall into Plausibility Theory. Or you can assign a fuzzy membership function to the set Doberman, another fuzzy membership function to the set Dog, and so on, and then fall into Confusion Theory (Section 2). But, as we shall see (Section 3), Problem 1 can be solved without resort to Fuzzy Logic.

Our solution uses hierarchies of qualitative values and the confusion conf(r, s), to measure how r^* (the yet unknown result) differs from each of the reported values. Once these measurements are known, Section 3 finds the r^* that minimizes them, and that is the result. This paper is a summary of (Jiménez, in press).

An important remark is that our solutions to Problems 1 and 1 do not address the full Inconsistency Problem in Belief Revision (Gärdenfors, 1992). Formulae in a theory (which may or may not be consistent) can use several constants on several variables (several attributes), but we are dealing just with assertions on one characteristic or attribute (such as *who was the killer?* or *the place of birth of Juárez*).

Solutions (b)–(e) still regard a set of formulae as consistent or inconsistent, and they try to ascertain, given an inconsistent set, how many causes or reasons for inconsistency it contains. In some sense, they measure how much work is needed to make consistent an inconsistent set. Our solution does not measure the inconsistency of a set by how much work is needed to bring it back to consistency. Instead, it measures the "intrinsic discrepancies" among the members of the set.

2. Measuring the confusion among two qualitative values

This section is an extract from our work in Levachkine, Guzman-Arenas, and de Gyves (2005) and Levachkine and Guzman-Arenas (2007). How close are two numeric values v_1 and v_2? The answer is $|v_2 - v_1|$. How close are two symbolic values such as cat and dog? The answer comes in a variety of similarity measures and distances. The hierarchies introduced in Fig. 1 allow us to define the confusion conf(r, s) on two symbolic values. We assume that the observers of a given fact (such as *the killer*) share a set of common vocabulary, best arranged in a hierarchy. This hierarchy can be regarded as the "common terminology"[4] for the observers of the bag, their *context*. Observers reporting in other bag may share a different context, that is, another hierarchy. The function conf will open the way to evaluate in Section 3 the inconsistency among a bag of symbolic observations.

What is the capital of Germany? *Berlin* is the correct answer; *Frankfurt* is a close miss, *Madrid* a fair error, and *sausage* a gross error. What is closer to a *cat*, a *dog* or an *orange*? Can we measure these errors and similarities? Can we retrieve objects in a database that are close to a desired item? Yes, by arranging these symbolic (that is, non-numeric) values in a hierarchy. More precisely, qualitative variables take symbolic values such as *cat*, *orange*, *California*, *Africa*. These values can be organized in a hierarchy H, a mathematical construct among these values. Over H, we can define the function *confusion* resulting when using a symbolic value instead of another.

Definition. For $r, s \in H$, the **absolute confusion** of using r instead of s, is

$$\mathrm{CONF}(r, r) = \mathrm{CONF}(r,\ \text{any ascendant of } r) = 0;$$
$$\mathrm{CONF}(r, s) = 1 + \mathrm{CONF}(r, \mathrm{father\_of}(s)).$$

[4] If the symbolic values become full *concepts*, it is best to use an *ontology* instead of a *hierarchy* to place them.