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

Adolfo Guzman-Arenas, Adriana Jimenez-Contreras
a.guzman@acm.org, dyidyia@yahoo.com

Centro de Investigación en Computación (CIC), Instituto Politécnico Nacional (IPN), México

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 "3m." Another person with the help of a meter said "3.13m". A lady with a micrometer reported "3.1427m." From these, it is possible to obtain the most likely value (μ=3.09m, 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 (Figure 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.

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

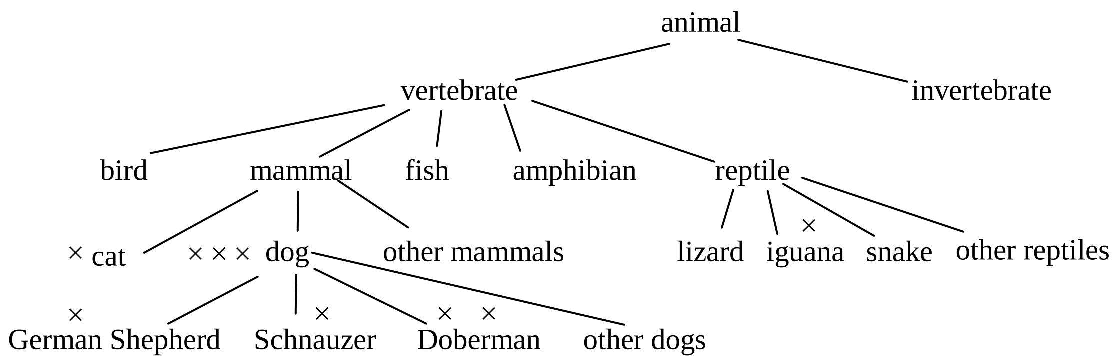

Figure 1. A hierarchy of symbolic values. It is a tree where every node is either a symbolic value or, if it is a set, then its descendants form a partition. Hierarchies make possible to compute the confusion conf(*r*, *s*) that results when value *r* is used instead of *s*, the true or intended value. The confusion (§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 × refer to Section 3

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, Shafer] 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 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 ≠ cat ≠Doberman ≠ …, the set is inconsistent, and the sentence (the killer is a dog) ∧ (the killer is a cat) ∧ … 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] 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

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

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]".[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]."[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]. 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 §2 quantifies this.

(f) Using some kind of high-order Logic, such as para-consistent logic or non-monotonic logic.

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 (§2). But, as we shall see (§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, §3 finds the r^* that minimizes them, and that is the result.

An important remark is that our solutions to Problems 1 and 2 do not address the full Inconsistency Problem in Belief Revision [Gardenfors]. 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) to (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 2005, 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

[3] Citations are from [Hunter].

distances. The hierarchies introduced in Figure 1 allows 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

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

To measure CONF, count the descending links from r (the replacing value) to s (the intended or real value). CONF is not a distance, nor an ultradistance. It is not symmetric.

We can normalize CONF by dividing into h, the height of H (the number of links from the root of H to the farthest element of H), yielding the following

Definition. The **confusion** of using r instead of s is

$$\mathrm{conf}(r, s) = \mathrm{CONF}(r, s)/h.$$

Notice that $0 \le \mathrm{conf}(r, s) \le 1$. It is not symmetric: $\mathrm{conf}(r, s) \ne \mathrm{conf}(s, r)$, in general.

Examples for the hierarchy of Figure 1. CONF (cat, mammal)=0; if I ask for a mammal and I am given a cat instead, I am happy, and CONF = 0. But CONF (mammal, cat) = 1; if I ask for a cat and are given a mammal, I am somewhat unhappy with CONF = 1. For the same reason, CONF(cat, vertebrate)=2. Being given a vertebrate when I ask for a cat makes me more unhappy than when I was handed a mammal.

Remark. Since conf is a function on a hierarchy, it is not possible for a value (such as *rabbit*) to have two ascendants, to have more than one path from it towards the root. That is, *rabbit* may not be both a mammal and a bird.

Example. In the hierarchy of Figure 1, conf(cat, mammal)=0; conf(cat, dog)=1.

2.1 Is conf a distance?

The function conf is not a distance, since

(A) $\mathrm{conf}(r, s) \ne \mathrm{conf}(s, r)$ in general. Nevertheless,

(B) conf $(r, s) \ge 0$, and $\mathrm{conf}(r, r) = 0$. But conf $(r, s) = 0$ does not imply $r = s$. But

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

(C) conf (a, b) + conf $(b, c) \le$ conf(a, c). The triangle inequality is obeyed by conf.

To prove (C), we first notice that Figure 2 represents all possible distributions of a, b and c in a hierarchy, where m, n, p and q represent the number of links (perhaps 0) in their respective arc segments. Then,

For T1 we have conf(a, b) + conf(b, c) = (m + n) + q $\le$ conf(a, c) which is m + n.
For T2 we have conf(a, b) + conf(b, c) = (n + m) + q $\le$ conf(a, c) which is n + q.
For T3 we have conf(a, b) + conf(b, c) = m + (n + q) $\le$ conf(a, c) which is q.
For T4 we have conf(a, b) + conf(b, c) = m + (n + q) $\le$ conf(a, c) which is q.
For T5 we have conf(a, b) + conf(b, c) = m + (n + q) $\le$ conf(a, c) which is q.
For T6 we have conf(a, b) + conf(b, c) = m + q $\le$ conf(a, c) which is q.

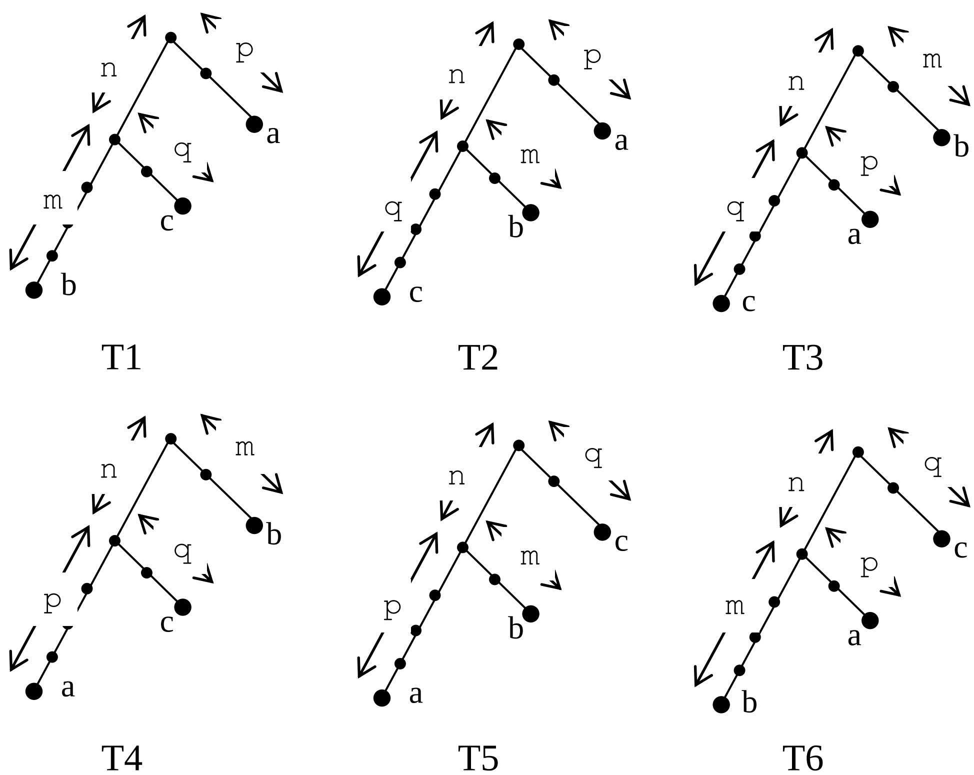

Figure 2. The cases used to prove that conf obeys the triangle inequality

3. Measuring the degree of inconsistency

Problem 1. Given a bag of observations reporting non-numeric values, how can me measure their inconsistency? What is the value that minimizes this inconsistency? We shall call r^* this value and σ the inconsistency that r^* produces.

Restrictions to the solution to Problem 1:

(A) All the reported values are about the same *fact* or property. One reporter can not report about the identity of the killer, while other reporter tells about the weather in London.

(B) The fact or feature that the observers are gauging, has a single value. There is only one killer. The weather in London (for a particular date and corner of the city) is unique.

(C) All reporters use the same *context* expressed in a vocabulary arranged in the same hierarchy –the same hierarchy for all observations in a bag. It is clear that for observers with other conceptions about the animals and their differences, the consensus r^* will differ. Thus, r^* and σ are a function of the bag and the hierarchy.

Intuitively, r^* is the value most likely to be true, given the available information, and taking into account observation errors. Intuitively too, σ is the average dispersion, "variance," discrepancy or degree of disagreement of this bag of observed values. We find r^* by finding the qualitative value r^* that minimizes the sum of discrepancies of each element in the bag with respect to such value. The inconsistency of the bag, called σ, is such minimum divided by the number of elements of the bag. Thus, σ is the minimum average discrepancy of the bag.

Let us solve problem 1 for an specific case. Assume we have a bag {dog, dog, cat, German Shepherd, Schnauzer, Doberman, Doberman, iguana} of observations about the killer: observers report that the killer a dog, a dog, a cat, and so on, respectively. These testimonies are marked with × in Figure 1. We wish to determine how divergent or discrepant they are –how inconsistent they are. Notice that inconsistency is a property of a bag of assertions. We first measure the total confusion that occurs when a given value r (one of the ×) is used instead of all the reported values (×). For instance, for r = Doberman, we obtain

Total confusion when Doberman is selected as the "representative" of the bag of observations = conf(Doberman, cat) +conf(Doberman, dog) +conf(Doberman, dog) +conf (Doberman, dog) +conf(Doberman, iguana) +conf(Doberman, German Shepherd) +conf (Doberman, Schnauzer) +conf(Doberman, Doberman) +conf(Doberman, Doberman) = ¼+0+0+0+½+¼+¼+0+0 = 5/4. The first term of the sum, conf(Doberman, cat)= ¼, means that the observer that reported "cat" will be at discomfort = ¼ when he finds that the consensus is Doberman. The second term, conf(Doberman, dog)=0 means that the observer that reported a dog will agree or show no discomfort (conf=0) when he finds that the representative is "Doberman" The fifth term of the sum, conf(Doberman, iguana) = ½ means that the observer that saw an iguana will be at high discomfort = ½ when he finds that the consensus is Doberman. The sum of confusions thus measures the total disagreement with the chosen representative value.

Total confusion when cat is selected as the "representative" of the bag of observations = conf(cat, cat) +conf(cat, dog) +conf(cat, dog)+ conf(cat, dog) + conf(cat, iguana) + conf(cat, German Shepherd) + conf(cat, Schnauzer) + conf(cat, Doberman) + conf(cat, Doberman) = 0+¼+¼+¼+½+½+½+½+½=3 ¼.

Total confusion if dog were the "representative" of the observations = conf(dog, cat) + conf(dog, dog) +conf(dog, dog) +conf(dog, dog) +conf(dog, iguana) +conf(dog, German Shepherd) +conf (dog, Schnauzer) +conf(dog, Doberman) +conf(dog, Doberman) = ¼ +0+0+0+½+¼+¼+¼+¼ =7/4.

Total confusion for German Shepherd as the representative=¼+0+0+0+½+0+¼+¼+¼= 3/2.

Total confusion for Schnauzer as the representative = ¼+0+0+0+½+¼+0+¼+¼ = 3/2.

Total confusion for iguana as the representative = ½+½+½+½+0+3/4+3/4+3/4+3/4= 5. Iguana is the most unlikely representative, producing the largest discomfort in the bag.

It makes sense to take as the best representative (consensus value) the animal that minimizes the total confusion. Such animal is Doberman for a total confusion of 5/4. This is the best "consensus", because it minimizes the confusion or "discomfort" of the observers when they see that Doberman was selected, instead of their observed animal. We can call this the "centroid" of the observed facts (× in Figure 1), the r^* of Problem 1.

The average confusion for r* is total confusion / number of observations = (5/4)/9 = 5/36. This is called the *inconsistency* σ of the bag of observations. It is the average confusion produced by its centroid r^*. Therefore, we have the following

Definitions. The *centroid* or *consensus* r^* of a bag B of observations reporting qualitative values $\{s_1, s_2, \ldots, s_k\}$ is the $r_j \in$ B that minimizes

$$\sum_{i=1}^{k} \mathrm{conf}(r_j, s_i) \qquad \text{for } j = 1,\ldots, k$$

The *inconsistency* σ of B is the minimum that such r^* produces, divided by k:

$$\sigma = (1/k)\ \min \sum_{i=1}^{k} \mathrm{conf}(r_j, s_i) = (1/k) \sum_{i=1}^{k} \mathrm{conf}(r^*, s_i) \qquad j\in[1,k]$$

Remarks.

I. The above centroid and inconsistency are the solutions to Problem 1.
II. r^* and σ depend on the context of use –represented by the hierarchy employed. The role played by the hierarchy in the solution to Problem 1 is to provide a *common vocabulary* for all observations. See restriction 3.(C).
III. The inconsistency $\sigma\in[0, 1)$. In fact, for a bag B of size |B|, $0 \le \sigma \le (|B|-1)/|B|$, since at least one of the members of B, namely r^*, has confusion 0 with the centroid r^* (by property IX, $r^*\in$B). For instance, given a hierarchy whose root is Person and having a single partition {male} {female}, then we have for bag {male, female} that r^* = male and σ = ½. Another solution is r^* = female, with the same value for σ, of course.
IV. There may be more than one value r^* that minimizes the total confusion, as the above example shows.
V. To compute the inconsistency of a bag, we resort to finding r^* first. In other words, the inconsistency of a bag is the average discomfort (average confusion) produced by r^*. This is the lowest inconsistency attainable; any other element different from r^* will give a larger or equal inconsistency (by definition of r^*).
VI. r^* is not necessarily the most popular value (the mode), which in the example for elements marked with (×) in Figure 1 is dog, while r^* = Doberman, as computed above.
VII. The *least common ancestor* (vertebrate in our example) produces a total confusion larger or at best equal than the total confusion produced by r^*. It is "too general" for everybody.
VIII. Discarding outliers, such as iguana in the bag of ×'s in Figure 1, we have a new σ = (4/4)/8 = 1/8, a tighter result than the σ =5/36 of the unexpunged bag. Details in §3.2.1.

Examples. For bag3={air, airplane, land, floor, subway, subway, motorcycle} (Figure 4), the inconsistency for air is (0 + 1/3 + 1/3 + 3/3 + 3/3 + 2/3 + 3/3)/7 = 13/21; for airplane, is σ=3/7; for airplane, is σ=3/7; for land, σ=20/21; for floor, σ=8/21; for subway, σ=2/7. Thus, the *largest outlier* (the observation that generates most distrust in bag3. §3.2, that with the largest inconsistency) is air, and the consensus is subway. For bag1 = {animal, vertebrate, bird, mammal, cat, dog, dog, iguana, German Shepherd} (marked with × in Figure 3), the centroid *r** is German Shepherd, and the inconsistency of the bag is (4/9)/5 = 4/45. For bag2 = {animal, amphibian, amphibian, reptile, reptile, snake} marked with • , *r** = snake, σ = (2/6)/5 =1/15. Taking into account all the findings × and •, we obtain for bag1∪bag2 a consensus *r** = German Shepherd with σ = (10/15)/5 = 2/15.

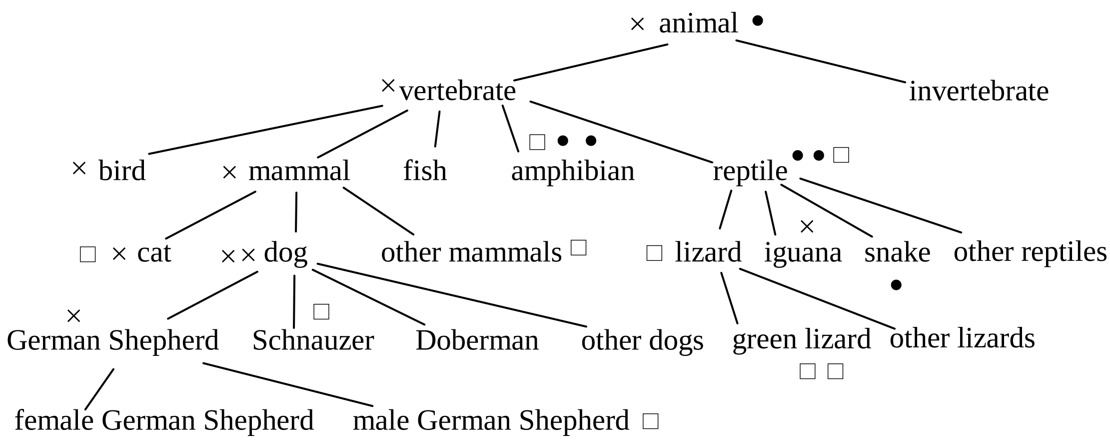

Figure 3. The consensus of the observations with × is *r** = German Shepherd, with inconsistency σ = 4/45. For observations with •, *r** = snake, σ = 1/15. For observations with □, *r** = green lizard, σ = (12/9)/5 = 4/15

Notice that we have found a way of adding (and averaging) apples and oranges, and a quantity (σ) to quantify out how disperse or divergent a bag of symbolic values is. In §4 we will introduce negative findings, and we will be able to add and average apples and "negative oranges," so to speak.

3.1 Properties of the consensus *r** and the inconsistency σ

It can be easily proved that:

IX. *r** is always inside the bag. Given a bag B, there is no other value *r*+ ≠ *r** in the hierarchy that provokes a lower value for σ. For instance, for bag1 in 2, all ascendants of German Shepherd will yield inconsistencies larger than 4/45. We could have used Female German Shepherd instead of German Shepherd, which also yields σ=4/45. But it will be strange to report that the consensus is Female German Shepherd, since no observer reported this value! Hence, in the definition we force *r** ∈ B.

X. The more specialized, the better. Consensus tends to go to the precise values (those deep in the hierarchy), unless of course overruled by several other less-precise values. For instance, in Figure 3, the most precise value of {cat, dog, dog, mammal, lizard, green lizard}is green lizard, but the consensus is dog with σ = (1/5+0+0+0+2/5+3/5)/6 =1/5, while for green lizard that value would be (2/5+2/5+2/5+1/5+0+0)/6 = 7/42.

XI. The consensus of a bag whose elements can be totally ordered by the "descendant" relation, is the lowest or more specific element in the bag. Example: In Figure 3, the consensus of {Doberman, dog, mammal} is Doberman. In general, if the elements of a bag can be ordered such that $\{a \leq b \leq c \ldots\}$ where $\leq$ is the descendant relation of the hierarchy, then the consensus of such bag is a, and its inconsistency is 0.

XII. Given the consensus r^* of a bag B, there is no $r \in$B such that r' is a descendant of r^*. Because if r' existed, then conf(r', r^*) = 0 but conf (r^*, r') ≥ 0, which means that r' would produce lower total confusion than r^*, contradicting the fact that the r^* is the consensus.

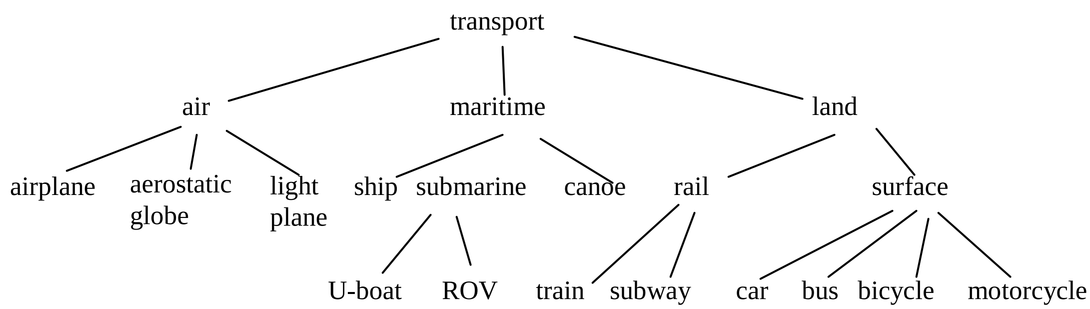

Figure 4. A hierarchy of types of transportation. The observations of the bag of example 3 are shown with ×

3.1.1 The consensus tends to the most precise observation

The consensus is the value closest to truth (to whatever is true in real life, to the real killer in our example) that we could have, given the information in the bag. Property 3.1.X is telling us that in general the consensus is closer to the greatest accuracy, and not to the mode (the most popular value –the truth is not proved by majority voting). Example: 22 miles North of Quito, Ecuador, there exists a place called *La Mitad del Mundo* (Middle of the World), with a monument, built in the early 1980's, that marks the place where the equator crosses the area. Findings were done by about 15 geographers. Later, with better accuracy, it was determined that the actual equator is about 200m north of the monument.

But, as the example given in Property 3.1.X shows, enough not-so-accurate observations can overcome a more accurate finding.

3.2 Finding the largest outlier

Definition. The largest outlier o of a bag B is the assertion s_j that contributes most to the inconsistency of the bag. Thus, o is given by the $s_j \in$ B that maximizes

conf(r^*, s_j).

Of all the assertions of B, o is the furthest away from the consensus; the one that provokes most discomfort among all the assertions of B. To have an idea of how far this largest outlier is, compare conf(r^*, o), the largest disagreement, with σ, the average disagreement. Removing o from the bag B contributes most to the consistency of B (reduces σ by the largest amount possible in a single removal).

Caution: If there are several repeated values inside the bag, removing one of them will not remove the others. *Hint:* try first candidates away from the consensus. *Remark.* There may be more than one o. *Remark:* removal of o, in addition to producing a smaller redundancy, may result in a new centroid (for the smaller bag).

Examples. ♦ The outermost outlier for the bag of objects marked with × in Figure 1 is iguana; expunging it produces Doberman as consensus, with a new inconsistency of (¼+ 0+0+0+0+¼+¼+¼)/8 = 1/8, as opposed to 5/36 when undeleted (as computed in the first example of §3). ♦ For bag1 (values marked × in Figure 3), the centroid is German Shepherd, the inconsistency is 4/45 (according to Figure 3). Its most conspicuous outlier is iguana; removing this value reduces the inconsistency of the smaller bag to (1/5)/8 =1/40; the new centroid is still German Shepherd (The only contributor to the inconsistency is the observation *bird*). ♦ Bag2 shown with • in Figure 3 has an inconsistency of 1/15, with centroid = snake. Its most prominent outlier is amphibian. Without it (that is, without one of them; since there are two "amphibian" observations), the smaller bag2 drops its inconsistency to (1/5)/5 = 1/25; the new centroid is still snake. ♦ Removing *air*, the largest outlier in bag3 (elements marked in Figure 3 with ×) leaves a smaller bag with still the same consensus r^* = subway, but with smaller $\sigma = 5/18$.

3.2.1 Removal of the largest outlier

As in the numeric case, care should be taken when removing large outliers from a bag. Suppose a large outlier lies deep in the hierarchy. It means that it was observed (the results were obtained) with care, with accuracy. Why could it be wrong, if the observer did a careful job, and he is not a liar? Just because his finding was not liked by r^*? Just because *o* does not agree with "most other observations"? It could be that the observer "missed the mark", he was watching of the *wrong* murder. Or it could be that we are just trying to adjust the observations to our preferences, to embellish the data, to tweak the data to our taste.

Remark. This removal is not a *contraction*, in the sense used in Belief Revision [Gärdenfors], since no additional formulae (additional assertions) are removed from B; just *o* is removed. Thus, the AGM postulates (Alchourrón, Gärdenfors, and Makinson) need not be obeyed.

3.3 Incremental computation of the inconsistency

Let B a bag of assertions with centroid r^* and inconsistency σ. If a new assertion *s* is added to B, we would like to compute the new centroid and inconsistency without having to process all the elements again, specially if B is large.

It is not possible, just knowing *s*, r^* and σ for a bag, to compute its new σ when a new observation *s* is added. We must also know |B|, the number of observations in the bag. Then, the solution can be found by computing the new total confusion $\sigma\,|B| + \mathrm{conf}(r^*, s)$ and dividing it among the new number of elements of B, |B| +1. That is:

$$\sigma_{NEW} = [\sigma\,|B| + \mathrm{conf}(r^*, s)] / [|B| + 1]$$
$$r^*_{NEW} = r^* \qquad \text{(APPROXIMATION 1)}$$

Unfortunately, this new inconsistency is computed against the old consensus r^*, which may not be the new consensus anymore. Approximation 1 is suitable when the number |B| of observations is large (say, >10), or when the new observation *r* is not too far from r^* (say, $\mathrm{conf}(r^*, r) < \sigma$). In these cases, it is reasonable to assume that the consensus will remain unchanged. If it remains unchanged, then the s computed by Approximation 1 is accurate (*is* the correct value).

3.3.1 Exact computation of r^*_{NEW}

Exact computation of r^*_{NEW}, σ_{NEW} requires to know all the observations of bag B, and then apply to B ∪ {*s*} the solution given in Section 3 to Problem 1. The inconsistency of the new bag may increase, decrease or stay unchanged. Examples in Table 1 give the new inconsistency and consensus when a new observation *bird* is added to the bag in the first column.

Table 1. Incremental change of the inconsistency and the consensus when a new value *s* = bird is added

	Old consensus and inconsistency	New inconsistency (by approximation 1)	New consensus and inconsistency (correct values)	Was Approx. 1 good?
Bag × of Figure 1	Doberman, 5/36	(9*5/36+¼)/10=3/20	Doberman, (6/4)/10=3/20	✓
Bag × of Figure 3	German Shepherd, 4/45	(9*4/45+1/5)/10=1/10	German Shepherd, (5/5)/10= 1/10	✓
Bag • of Figure 3	snake, 1/15	(6*1/15+1/5)/7=3/35	snake, (3/5)/7 = 3/35	✓
Bag {×}∪{•} of Figure 3	German Shepherd, 2/15	(15*2/15+1/5)/16 = 0.137	German Shepherd, (11/5)/16 = 0.137	✓
Bag {□} of Figure 3	green lizard, 4/15	(9*4/15 + 1/5)/10 = 0.26	green lizard, (13/5)/10 =0.26	✓

4. Inconsistency in the presence of negative assertions

Sometimes, there are examinations or tests that conclude that the observed value *is not* one of several values, but it does not tell us what the value *is.* For instance, a paternity test using ADN strands of a person and her alleged fathers can tell us that the father ∉ {Albert, Bob, Carl} but it does not tell us who the real father is. How can we compute the consensus r^* and the inconsistency σ of several observations that include negative findings?

To handle this case, let s be the forbidden value. The definition of conf in §2 should be expanded to account for negative evidence. How best to do this? First, let us try:

conf(r, ¬s) = conf (s, ¬s) = conf(r, any descendant of s) = 1;
conf(r, ¬s) = 0 otherwise.

Example: conf(snake, ¬reptile) = 1; If I want something which is not a reptile and they give me a snake, my discomfort is 1 since a snake is, after all, a reptile.

Example: The values of conf(r, ¬reptile) for different r are shown in.Figure 5.

This extension to conf seems to be fine, since the brothers of reptile (which are bird, mammal, fish, amphibian) are definitely ¬reptile, so for instance conf(dog, ¬reptile) = 0 (If I want something which should not be a reptile, and they give me a dog, I am satisfied). But, what to do with vertebrate, which is partially a reptile, partially a bird, etc? Are most vertebrates non-reptiles? If the set vertebrate is partitioned in sons of equal size, then CONF(vertebrate, ¬reptile) should be 1/5 [Only one of the five sons of vertebrate is reptile]. Therefore, conf(vertebrate, ¬reptile) = 1/5*h* (*h* is the height of the hierarchy). Or, better, conf(vertebrate, ¬reptile) = (total population of reptiles) / (total population of vertebrates), which is more precise but harder to estimate [Guzman & Levachkine 2004]. If we use the approximation 1/5, then CONF(animal, ¬reptile) = ½ * 1/5, since animal has two sons, and only one of them has CONF = 1/5. Hence, conf(animal, ¬reptile) == ½ * 1/5 * 1/*h*. [Remember that conf could be small even if CONF is large, if the subtree that *animal* spans is a small part of the larger hierarchy in which it is immersed].

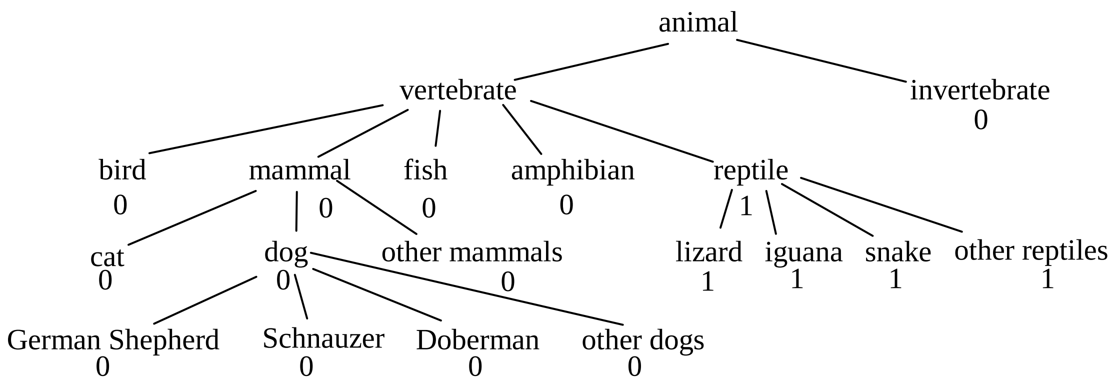

Figure 5. conf(*r*, ¬reptile) for different *r*'s. The number under each value represent the confusion conf of using that value instead of "not reptile." The text argues that the number under vertebrate should be 1/5, and under animal 1/10

Then, the extension to CONF for negative assertions is as follows:

CONF(*r*, ¬*s*) = *h* if *r* = *s* or if *r* is any descendant of *s*;
CONF(*r*, ¬*s*) = [1/(1+number of brothers of *r'*)]* CONF(*r',* ¬*s*) if *r* and *r'* are ascendants of *s* and *r* =father_off(*r*');
CONF(*r*, ¬*s*) = 0 otherwise.

Here, we give the value *h* to CONF(*r*, ¬*s*) when *r* = *s* or if *r* is any descendant of *s*, in order that conf(*r*, ¬*s*) for the same cases will become 1.

The transition from CONF to conf is as usual:

conf(*r*, ¬*s*) = CONF(*r*, ¬*s*)/*h*

Intuitively, we can insert a new node ¬reptile in the precise place in the hierarchy, to help us understand this extension to CONF. The hierarchy of Figure 5 now becomes that of Figure 6. But this new node does not represent all non reptiles, since invertebrates are also non reptiles. The place where such node is inserted is "not completely accurate."

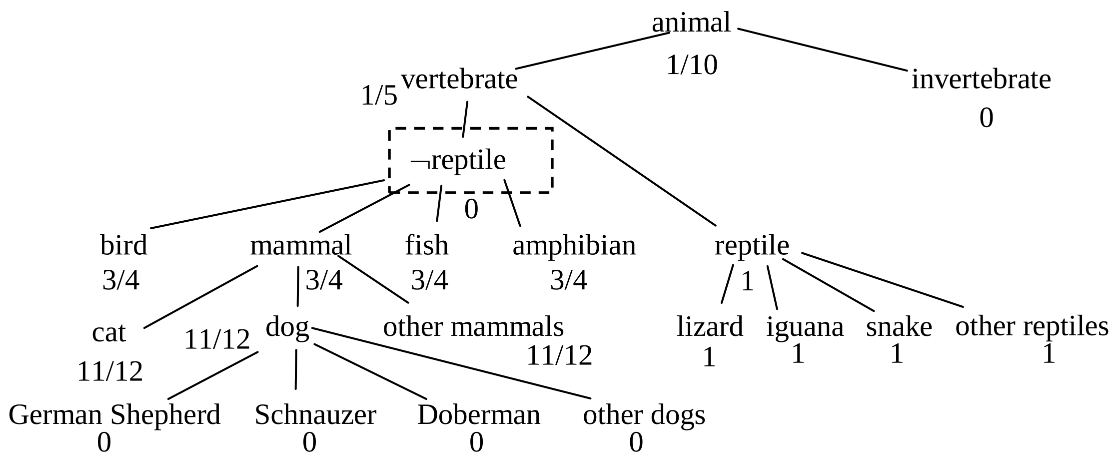

Figure 6. The introduction of a virtual node ¬reptile helps to understand the confusion with negative findings. The numeric labels under each concept *r* show conf(*r*, ¬reptile). For instance, conf(dog, ¬reptile) = 0. If I want a non reptile and they give me a dog, I am satisfied. Notice that the node ¬reptile in the figure misses the invertebrates, which are also non reptiles!

4.1 The confusion in using a negative finding instead of other value *s*

Does it make sense to report as the consensus of a bag of observations a *negative finding*? For instance, given a set of observations, could the consensus be "not a dog"? Depending on the situation at hand, this may or may not make sense. Let us consider some examples; refer to Figure 7.

If I want a reptile and they give me a ¬reptile, my discomfort as measured by conf is clearly 1. The same discomfort I have if I want an iguana and I am given a ¬reptile, conf(¬reptile, iguana) = 1. Thus, we have

conf(¬reptile, reptile) = conf(¬reptile, any descendant of (reptile)) = 1.

Also, it is clear that conf(¬reptile, ¬reptile) = 0. I am given what I want.

What could be conf (¬reptile, bird)? I want a bird, and they give me a "non reptile." My discomfort is close to 1, since there is only a small probability that the "non reptile" they give me is indeed a bird. Only one of the four sons of ¬reptile is a bird. All the invertebrates are also members of ¬reptile. My discomfort is 1 - (total number of birds) / (total number of non reptiles). Using some approximations,[5] we have:

conf (¬reptile, bird) = 1-(total number of birds)/(total number of non reptiles) ≈ 1 – ([1/5] / [4/5*½ + ½]) = 1 - 0.2/0.9 = 1 – 0.222 = 0.778

conf (¬reptile, mammal) = 1-(total number of mammals)/(total number of non reptiles) ≈ 0.778

[5] It is difficult to assess the total number of animals, vertebrates, birds, etc. We can grossly approximate them from the hierarchy (Figure 7), as follows: For each animal, we have 0.5 vertebrates and 0.5 invertebrates. For each vertebrate we have 0.2 birds, 0.2 mammals, 0.2 fish, 0.2 amphibians and 0.2 reptiles. So, the percentages of animals that are invertebrates is 50%; the percentages of vertebrates is 50%, too; the percentage of reptiles is 20% of 50% = 10%, and the same number holds for birds, mammals, fishes and amphibians. Thus, the percentage of non reptiles is 0.1 + 0.1 +0.1 +0.1 + 0.5 = 0.9, and (total number of birds) / (total number of non-reptiles) = 0.2/0.9 = 0.222.

By the same reasoning, conf (¬reptile, cat) = 1-(total number of cats)/(total number of non reptiles) ≈ 1 – ([1/5* 1/3] / [4/5*½+ ½]) = 1- 0.0667/0.9 = 1-0.074= 0.925

Also, conf (¬reptile, Schnauzer) = 1 – (total number of Schnauzers)/(total number of non reptiles) ≈ 1 – ([1/5* 1/3] / [4/5*½+ ½]) = 47/48 = 0.979

We have intuitively computed the values for nodes in the tree spanned by reptile and by ¬reptile. Let us go beyond them. What could be the value for conf (¬reptile, vertebrate)? That is, I want a vertebrate, and they give me a ¬reptile.

conf(¬reptile, vertebrate) = 1 – (total number of vertebrates) / (total number of non reptiles). From footnote 5, this is 1 – 0.5/0.9 = 1 – 0.555 = 0.445

.

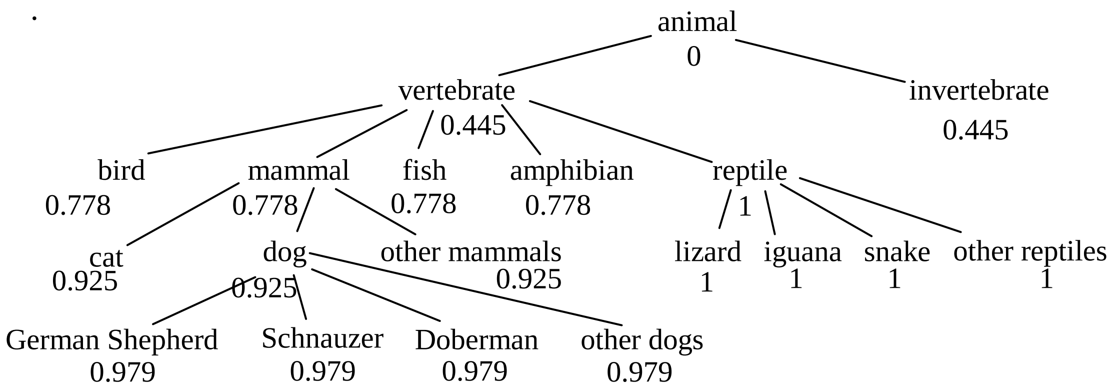

Figure 7. The values under each node *s* are the confusion in using ¬reptile instead of such node *s*. For instance, conf(¬reptile, Schnauzer) = 0.979

conf (¬reptile, invertebrate) = 1 – (total number of invertebrates) / (total number of non reptiles) ≈ 1 – 0.5/0.9 = 0.445

But notice that conf (¬reptile, animal) = 1 – (total number of animals) / (total number of non reptiles) ≈ 1 – 1/0.9 = -0.111, which does not make sense. This is because if I want an animal and they give a non reptile, I am satisfied, since that "non reptile" is also an animal. Thus, conf (¬reptile, animal) = 0.

In summary, the formulas for selecting a negative assertion ¬*r* instead of *s* are:

conf(¬*r*, *r*) = conf(¬*r*, any descendant of *r*) = 1.
conf(¬*r*, ¬*r*) = 0.
conf(¬*r*, *s*) = 0 when *s* is the root of the hierarchy.
conf(¬*r*, *s*) = 1 – (total number of *s*'s)/(total number of non *r*'s) otherwise.

Conclusions. Just like when different measurements produce inconsistent numeric values, it is possible to obtain the most likely or consensus value for a bag of non-numeric observations. While classic Logic tells us that, since dog ≠ Doberman ≠ cat, then the affirmation (*v* = dog) ∧ (*v* = Doberman) ∧ (*v* = cat) is false and inconsistent, we "really know" that dog and Doberman are not so different, and that Doberman and cat are more different, but still not "completely inconsistent." In other words, given that v = dog, the observation v = Doberman contradicts "just a little" the previous value. Until now, inconsistency (and the a

computation of the most likely value) in a set of observations was possible by assigning a confidence to each of the observers (Dempster-Shafer Theory), or by counting [Hunter] how many predicates were violated by the set of observations. But a composite predicate could be decomposed in several smaller ones (or several predicates could be consolidated into one by AND), thus giving a different value for the inconsistency computed in this way, while the inconsistency of the bag remains unchanged, since neither the bag nor the predicates testing it have changed.

We have found another way of measuring inconsistency: by considering that observers reporting the same fact differ because their observation instruments or means vary in precision. In order to measure this, the *consensus* or most likely value is obtained first.

The paper extends the notion of consensus and inconsistency to negative assertions.

Acknowledgments. Work was in part supported by CONACYT grant 43377, and by SNI.

References

Byrne, E., & Hunter, A. (2005) Evaluating violations of expectations to find exceptional information, *Data and Knowledge Engineering*, 54(2):97-120.

Dempster, Arthur P. (1968) A generalization of Bayesian inference, *Journal of the Royal Statistical Society*, Series B, Vol. 30, pp. 205-247.

Gärdenfors, M. (1992) *Belief Revision*, Cambridge University Press.

Guzman, A., and Levachkine, S. (2004) Hierarchies Measuring Qualitative Variables. *Lecture Notes in Computer Science* LNCS **2945** (Computational Linguistics and Intelligent Text Processing), Springer-Verlag. 262-274. ISSN 0372-9743. http://www.cic.ipn.mx/aguzman/papers/151%20lncs-h1(levachkine-guzman)-12(cicling04)5no03.pdf

Jiménez, Adriana. Characterization and measurement of logical properties of qualitative values organized in hierarchies. Ph. D. Thesis in progress. CIC-IPN. In Spanish.

Hunter, A., and Konieczny, S. Measuring inconsistency through minimal inconsistent sets. *Proceedings of the* 11th *International Conference on Knowledge Representation* (KR'08), AAAI Press (in press).

Knight, K. (2001). Measuring inconsistency. *Journal of Philosophical Logic* **31**:77–98.

Konieczny, S.; Lang, J.; and Marquis, P. (2003). Quantifying information and contradiction in propositional logic through epistemic actions. In *Proceedings of the 18th International Joint* Conference *on Artificial Intelligence* (IJCAI' 03), 106–111.

Levachkine, S, and Guzman-Arenas, A. and de Gyves, V.P. (2005) The semantics of confusion in hierarchies: from theory to practice. In *Contributions to ICCS 05 13th International Conference on Conceptual Structures: common semantics for sharing knowledge*, July 18-22 2005, Kassel, Germany. 94-107 http://www.cic.ipn.mx/aguzman/papers/167%20The%20semantics%20of%20confusion%20in%20hierarchies%20Copia%20de%20ICCS05final.pdf

Levachkine, S., and Guzman-Arenas, A. (2007) Hierarchy as a new data type for qualitative variables. *Journal Expert Systems with Applications* **32**, 3, June 2007

Shafer, Glenn. (1979) A Mathematical Theory of Evidence. Princeton University Press.