Bridging Gaps Between Subjective Logic and Semantic Web

Davide Ceolin(✉), Archana Nottamkandath, and Wan Fokkink

VU University, Amsterdam, The Netherlands
{d.ceolin,a.nottamkandath,w.j.fokkink}@vu.nl

Abstract. Subjective logic is a powerful probabilistic logic which is useful to handle data in case of uncertainty. Subjective logic and the Semantic Web can mutually benefit from each other, since subjective logic is useful to handle the inner noisiness of the Semantic Web data, while the Semantic Web offers a means to obtain evidence useful for performing evidential reasoning based on subjective logic. In this chapter we describe three extensions and applications of subjective logic in the Semantic Web, namely: the use of deterministic and probabilistic semantic similarity measures for weighing subjective opinions, a way for accounting for partial observations, and "open world opinion", i.e. subjective opinions based on Dirichlet processes, which extend multinomial opinions. For each of these extensions, we provide examples and applications to prove their validity.

Keywords: Subjective logic · Semantic similarity · Dirichlet process · Partial observations

1 Introduction

Subjective logic [12] is a probabilistic logic widely adopted in the trust management domain, based on evidential reasoning and statistical principles. This logic focuses on the representation and the reasoning on assertions of which the truth value is not fully determined, but estimated on the basis of the observed evidence. The logic comes with a variety of operators that allow to combine such assertions and to derive the truth values of the consequences.

Subjective logic is well-suited for the management of uncertainty within the Semantic Web. For instance, the incremental access to these data (as a consequence of crawling) can give rise to uncertainty issues which can be dealt with using this logic. Furthermore, the fact that the fulcrum of this logic is the concept of "subjective opinion" (which represents a logic proposition, its corresponding belief and the source of this evidence), allows correctly representing how the estimated truth value of an assertion is bound to the source of the corresponding evidence and allows to easily keep lightweight provenance information. Finally, evidential reasoning allows to limit the typical noisiness of Semantic Web data. On the other hand, we also believe that the Semantic Web can be

F. Bobillo et al. (Eds.): URSW 2011-2013, LNAI 8816, pp. 242–264, 2014.
DOI: 10.1007/978-3-319-13413-0_13

beneficial to this logic, as an immeasurably important source of information: since the truth value of assertions is based on availability of observations, the more data is available (hopefully of high enough quality), the closer we can get to the correct truth value for our assertions. We believe that this mutual relationship can be improved. This chapter proposes extensions and applications of subjective logic that aim at the Semantic Web, namely: the use of deterministic and probabilistic semantic similarity measures for weighing subjective opinions, a method for accounting for partial observations, and "open world opinions", that are subjective opinions based on Dirichlet processes. Open world opinions allow modeling categorical data for which categories are partially known. Only the latter is a proper extension of the logic, while the first two items are representations within the logic of external elements, with proper mappings and, when necessary, specific representations.

This chapter revises and extends the paper "Subjective Logic Extensions for the Semantic Web" [5], presented at the 8th International Workshop on Uncertainty Reasoning for the Semantic Web, at the 11th International Semantic Web Conference 2012. Here, we add to that paper two methods to map probabilistic semantic similarity measures and subjective opinions.

The rest of the chapter is organized as follows. Section 2 gives an overview of subjective logic. Sections 3 and 4 show how to combine subjective logic with deterministic and probabilistic semantic similarity measures respectively. Section 5 introduces a method for dealing with partial observations of evidence, Sect. 6 describes the concept of "open world opinion". Section 7 describes related work, and Sect. 8 provides a final conclusion about the work presented.

2 Subjective Logic

In subjective logic, so-called "subjective opinions" express the belief that source x owns with respect to the value of proposition y. The values of y are chosen among the elements of the set Θ ("frame of discernment"). For instance, if y is a binomial proposition, then $\Theta = \{true, false\}$. A subjective opinion describes the belief in the elements of the power set of Θ (2^Θ). In symbols, an opinion is represented as

$$\omega_y^x(b, d, u, a)$$

when $|\Theta| = 2$ (binomial opinion) or as

$$\omega_y^x(\overrightarrow{B}, u, \overrightarrow{A})$$

when $|\Theta| > 2$ (multinomial opinion). Throughout the paper we refer to vectors with the following notation: $\overrightarrow{B}$. Its elements are represented as b_x. In the binomial opinion, b represents the belief in y being *true* and d the belief in y being *false*, i.e., the disbelief. The uncertainty u represents a part of probability mass that we are unable to assign to either *true* or *false* and it therefore corresponds to the belief in Θ. In the case of the multinomial opinion there is no disbelief because there is no specific *false* value, since y can assume multiple ones. a represent the prior probability that y has to be true, while A represents the vector of prior probabilities for each of the possible truth values of y. The values b, d, u are determined by observing pieces of evidence. a is given a priori. The positive and negative evidence is represented as p and n respectively. The belief (b), disbelief (d), uncertainty (u), and a priori values (a) for binomial opinions are computed as:

$$b = \frac{p}{p+n+2} \quad d = \frac{n}{p+n+2} \quad u = \frac{2}{p+n+2} \quad a = \frac{1}{2}. \tag{1}$$

The value 2 indicates the cardinality of Θ, i.e., the number of values that y can take. A subjective opinion is equivalent to a Beta probability distribution (binomial opinion) or to a Dirichlet distribution (multinomial opinion). This probability distribution describes the most likely probability values that y can take. If y has Pr probability to be true, since we determine Pr starting from a limited set of evidence, we estimate the most likely value of Pr by means of a Beta (or Dirichlet) probability distribution.

Opinions can be contextualized. For example, source x provides an observation about assertion y in context c (e.g. about an agent's expertise). The most likely value for y in context c, represented as $t(x, y : c)$, is the expected value of the Beta distribution corresponding to the opinion and computed as:

$$E = t(x, y : c) = b + a \cdot u. \tag{2}$$

The reason why we rely on this logic is the fact that it makes use of a double probabilistic layer. The probability of each proposition can be represented by means of a Binomial distribution (or by means of a Multinomial distribution if the proposition is multivalued). However, we base our truth estimations on samples of Web data so the parameter p of the Binomial distribution (or the vector of parameters $\overrightarrow{P}$ of the Multinomial) is rather uncertain. In fact, the Web data sample is possibly unreliable, uncertain and partially representative of the entire Web data population. Subjective logic uses a second-order distribution based on the distribution and size of the sample at our disposal to estimate the most likely value that the p (or $\overrightarrow{P}$) value can take. This is the primary reason why we adopt this logic. Also, on the Web, data are exposed by different sources presenting different reliability levels. The ability to keep track of the source that exposes a given piece of data or a subjective opinion is crucial to be able to assess the trust in that piece of data or subjective opinion. Subjective logic allows keeping track of such provenance information and reasoning on subjective opinions weighing them on the reputation of their source. Lastly, subjective logic offers a variety of operators that allow combining subjective opinions in several manners. For instance, operators allow "discounting" an opinion based on the reputation of the source that exposes it, or computing the truth value (expressed as a subjective opinion) of the logical disjunction or conjunction of two opinions held by the same source. This makes the logic a useful tool to reason upon data extracted from the Web. One important remark is that this logic allows reasoning on binomial or multinomial data, that include, for instance URIs. The Beta and the Dirichlet distributions are used because they are “conjugated” [11] with the Binomial and Multinomial distributions respectively, i.e., their computation is particularly manageable. Other kind of data and other probability distributions are outside the scope of this logic.

2.1 Base Rate Discounting Operator in Subjective Logic

An important class of operators of subjective logic is the so-called “discounting” operator. In fact, a subjective opinion allows keeping track of the source of the opinion itself. This permits the reuse of the opinion by third parties, because these third parties, knowing where the opinion comes from, can decide to use it. However, before using it, these third parties may require to “smoothen” the opinion to take into account the limited reliability of the source or its possible maliciousness. Therefore, in subjective logic there exists a variety of discounting operators: for instance, one favors disbelief (to be used if the source is known to be malicious), and one that favors uncertainty (to be used when no specific intention of the source is known). We can also make use of the base-rate sensitive discounting operator in the case we just have a probability (i.e., the expected value of an opinion), instead of having at our disposal a complete subjective opinion for a source. The base-rate sensitive discounting of opinion of source B on y by opinion of source A on B ω_B^A,

$$\omega_y^B = (b_y^B, d_y^B, u_y^B, a_y^B)$$

by opinion

$$\omega_B^A = (b_B^A, d_B^A, u_B^A, a_B^A)$$

of source A produces transitive belief

$$\omega_y^{A:B} = (b_y^{A:B}, d_y^{A:B}, u_y^{A:B}, a_y^{A:B})$$

where

$$\begin{aligned} b_y^{A:B} &= E(\omega_B^A) b_y^B \\ d_y^{A:B} &= E(\omega_B^A) d_y^B \\ u_y^{A:B} &= 1 - E(\omega_B^A)(b_y^B + d_y^B) \\ a_y^{A:B} &= a_y^B. \end{aligned} \quad (3)$$

3 Combining Subjective Logic with Deterministic Semantic Similarity Measures

We saw in the previous section that opinions can be contextualized. Setting the context is important, because it allows delimiting the validity of an opinion and increasing the precision of the corresponding evaluation. For instance, if we gather evidence about the expertise of a user in a given topic, let us say, flowers, then it is important to delimit the validity of the corresponding opinion to the topic "flowers". However, contexts may also impede the use of evidence about a given subject, if the context differs from the context where the evidence was collected. Therefore, we propose to "bridge" contexts by using semantic similarity measures to import evidence from a context to another, after having weighed them on the similarity of the two contexts. Many semantic similarity measures have been developed (see the work of Budanitsky and Hirst [2]). These basically split into two main classes: deterministic and probabilistic semantic similarity measures. The deterministic ones are based on deterministic computations made, for example, on word graphs (e.g. WordNet [19]). The probabilistic ones apply probabilistic reasoning to derive semantic relatedness between words based, for instance, on the occurrence and co-occurrence of these two words in large document corpora. We extend the logic to incorporate these measures by representing semantic similarity measures by means of subjective opinions (discounting), or by using the similarity measures to weigh items of evidence before using them to build subjective opinions (weighing). The first extension of subjective logic that we propose regards the use of deterministic semantic similarity measures and is described as follows.

3.1 Wu and Palmer Semantic Similarity Measure

Among all deterministic semantic similarity measures, our attention focuses on those computed from *WordNet*. *WordNet* groups words into sets of synonyms called synsets that describe semantic relationships between them. It is a directed and acyclic graph in which each vertex v, is an integer that represents a synset, and each directed edge from v to w represents that w is a hypernym of v. In particular, we use the Wu and Palmer similarity measure [28], which calculates semantic relatedness in a deterministic way by considering the depths between two synsets in the WordNet taxonomies, along with the depth of the Least Common Subsumer (lcs) as follows:

$$score(s1, s2) = \frac{2 \cdot depth(lcs(s1, s2))}{depth(s1) + depth(s2)}. \quad (4)$$

This means that $score \in [0, 1]$. For deriving the opinions about a concept where no evidence is available, we incorporate $score$, which represents the semantic similarity ($sim(c, c')$) in our trust assessment, where c and c' are concepts belonging to synset $s1$ and $s2$ respectively which represent two contexts.

3.2 Using Semantic Similarity Measures Within Subjective Logic

We propose two means to import deterministic semantic similarity measures into subjective logic, by mapping them with subjective opinions.

Deriving an Opinion About a New or Unknown Context. Since we compute opinions based on contexts, it is possible that evidence required to compute the opinion for a particular context is unavailable. For example, suppose that source x owns observations about a proposition in a certain context (e.g. the expertise of an agent about tulips), but needs to evaluate them in a new context (e.g. the agent's expertise about sunflowers), of which it owns no observations. The semantic similarity measure between two contexts, $sim(c, c')$, can be used for obtaining the opinion about an agent y on an unknown or new context through two different methods. We can weigh the evidence at our disposal, and for every piece of evidence, use only the part that corresponds to the semantic similarity between the two contexts. If I have one observation in the known context c' and the similarity between the two contexts is 0.5, then I can use that observation in the new context c as 0.5 piece of evidence. Otherwise, I could compute a subjective opinion in the known context c' and then use it in the unknown context c after having "discounted" it (using the subjective logic discounting operator). The discounting factor would be a subjective opinion that represents the semantic similarity between the two contexts. The reason why we have these two different approaches is that weighing operates directly on the evidence, while discounting applies on the subjective opinion. In the latter case uncertainty has already been quantified (and therefore some probability mass has been assigned to it), while in the first case not. Hence, the choice between the two alternative depends on the strategy chosen (it could be that operating on the opinion is more computationally efficient, and hence discounting is preferable), or on case study constraints (e.g., if the evidence from a given context are already expressed as a subjective opinion, then it is simpler to use it than to revert it to the pieces of evidence on which it is computed). Below we provide more details about the two methods.

Evidence weighing. We weigh the positive and negative evidence belonging to a certain context (e.g. *Tulips*) on the corresponding semantic similarity to the new context (e.g. *Sunflowers*), *sim(Tulips, Sunflowers)*. We then perform this for all the contexts for which source x has already provided an opinion, $\forall c' \in C$, by weighing all the positive (p) and negative (n) evidence of c' with the similarity measure $sim(c, c')$ to obtain an opinion about y in c (see the work of Ceolin et al. [4]).

Opinion discounting. In the second approach, every opinion source x has about other related contexts c', where $c' \in C$, is discounted with the corresponding semantic similarity measure $sim(c, c')$ using the Discounting operator in subjective logic. The discounted opinions are then aggregated to form the final opinion of x about y in the new context c.

Discounting Operators and Semantic Similarity. Subjective logic offers a variety of operators for "discounting", i.e. for smoothing opinions given by third parties, provided that we have at our disposal an opinion about the source itself. "Smoothing" is meant as reducing the belief provided by the third party, depending on the opinion on the source (the worse the opinion, the higher the reduction). Moreover, since the components of the opinion always sum to one, reducing the belief implies an increase of (one of) the other components: hence there exists a discounting operator favoring uncertainty and one favoring disbelief. Finally, there exists a discounting operator that makes use of the expected value E of the opinion. Following this line of thought, we can use the semantic similarity as a discount factor for opinions imported from contexts related to the one of interest, in case of a lack of opinions in it, to handle possible variations in the validity of the statements due to the change of context.

So, we need to choose the appropriate discounting operator that allows us to use the semantic similarity value as a discounting factor for opinions. The disbelief favoring discounting is an operator that is employed whenever one believes that the source considered might be malicious. This is not our case, since the discounting is used to import opinions own by ourselves but computed in different contexts than the one of interest. Hence we do not make use of the disbelief favoring operator.

In principle, we would have no specific reason to choose one between the uncertainty favoring discounting and the base rate discounting. Basically, having that only rarely the belief (and hence the expected value) is equal to 1, the two discounting operators decrease the belief of the provided opinion, one by multiplying it by the belief in the source, the other one by the expected value of the opinion about the source. In practice, we will see that, thanks to Theorem 1 these two operators are almost equivalent in this context.

Theorem 1 (Semantic relatedness measure is a dogmatic opinion). *Let $sim(c, c')$ be the semantic similarity between two contexts c and c' obtained by computing the semantic relatedness between the contexts in a graph through deterministic measurements (e.g. [28]). Then, $\forall$ $sim(c, c') \in [0,1]$,*

$$\omega_{c=c'}^{measure} = (b_{c=c'}^{measure}, d_{c=c'}^{measure}, u_{c=c'}^{measure}, a_{c=c'}^{measure})$$

is equivalent to a dogmatic opinion in subjective logic, i.e., a subjective opinion with uncertainty equal to zero.

Proof. A binomial opinion is a dogmatic opinion if the value of *uncertainty is 0.* The semantic similarity measure can be represented as an opinion about the similarity of two contexts c and c'. However, since we restrict our focus on *WordNet*-based measures, the similarity is inferred by graph measurements, and not by probabilistic means. This means that, according to the source, this is a "dogmatic" opinion, since it does not provide any indication of uncertainty: $u_{c=c'}^{measure} = 0$. The opinion is not based on evidence observation, rather on actual deterministic measurements.

$$E(\omega_{c=c'}^{measure}) = b_{c=c'}^{measure} + u_{c=c'}^{measure} \cdot a = sim(c, c'), \tag{5}$$

where *measure* indicates the procedure used to obtain the semantic relatedness, e.g. Wu and Palmer Measure. The values of belief and disbelief are obtained as:

$$b_{c=c'}^{measure} = sim(c, c') \qquad d_{c=c'}^{measure} = 1 - b_{c=c'}^{measure}. \;\square \tag{6}$$

Corollary 1 (Discounting an opinion with a dogmatic opinion). *Let A be a source who has an opinion about y in context* c' *expressed as*

$$\omega^{A}_{y:c'} = (b^{A}_{y:c'}, d^{A}_{y:c'}, u^{A}_{y:c'}, a^{A}_{y:c'})$$

and let the semantic similarity between the contexts c and c' be represented as a dogmatic opinion

$$\omega^{measure}_{c=c'} = (b^{measure}_{c=c'}, d^{measure}_{c=c'}, 0, a^{c'}_{c=c'}).$$

Since the source A does not have any prior opinion about the context c, we derive the opinion of A about c represented as

$$\omega^{A:c'}_{c} = (b^{A:c'}_{c}, d^{A:c'}_{c}, u^{A:c'}_{c}, a^{A:c'}_{c})$$

using the base rate discounting operator on the dogmatic opinion.

$$\begin{gathered} a^{A:B}_{y} = a^{B}_{y} \qquad b^{A:B}_{y} = sim(c,c') \cdot b^{B}_{y} \\ u^{A:B}_{y} = 1 - sim(c,c') \cdot (b^{B}_{y} + d^{B}_{y}) \qquad d^{A:B}_{y} = sim(c,c') \cdot d^{B}_{y}. \end{gathered} \tag{7}$$

Definition 1 (Weighing operator). *Let C be the set of contexts* c' *of which a source A has an opinion derived from the positive and negative evidence in the past. Let c be a new context for which A has no opinion yet. We can derive the opinion of A about facts in c, by weighing the relevant evidences in set C with the semantic similarity measure* $sim(c, c')\ \forall c' \in C$. *The belief, disbelief, uncertainty and a priori obtained through the weighing operation are expressed below.*

$$\begin{gathered} b^{A}_{c} = \frac{sim(c,c') \cdot p^{A}_{c'}}{sim(c,c')(p^{A}_{c'} + n^{A}_{c'}) + 2} \qquad d^{A}_{c} = \frac{sim(c,c') \cdot n^{A}_{c'}}{sim(c,c')(p^{A}_{c'} + n^{A}_{c'}) + 2} \\ u^{A}_{c} = 1 - \frac{sim(c,c') \cdot (p^{A}_{c'} + n^{A}_{c'})}{sim(c,c')(p^{A}_{c'} + n^{A}_{c'}) + 2} \qquad a^{A}_{c} = a^{A}_{c'}. \end{gathered} \tag{8}$$

Theorem 2 (Approximation of the weighing and discounting operators). *Let*

$$\omega^{A:c'}_{y:c} = (b^{A:c'}_{y:c}, d^{A:c'}_{y:c}, u^{A:c'}_{y:c}, a^{A:c'}_{y:c})$$

be a discounted opinion which source A has about y in a new or unknown context c, derived by discounting A's opinion on known contexts c' $\in C$ *represented as* $\omega^{A}_{c'} = (b^{A}_{c'}, d^{A}_{c'}, u^{A}_{c'}, a^{A}_{c'})$ *with the corresponding dogmatic opinions (e.g. sim(c,c')). Let source A also obtain an opinion about the unknown context c based on the evidence available from the earlier contexts c', by weighing the evidence (positive and negative) with semantic similarity between c and c', sim(c,c')* $\forall c' \in C$. *Then the difference between the results from the weighing and from the discount operator in subjective logic are statistically insignificant.*

Proof. We substitute the values of belief, disbelief, uncertainty values in Eq. (9) for Base Rate Discounting with the values from Eq. (1) and expectation value from Eq. (5). We obtain the new value of the discounted base rate opinion as follows:

$$\begin{array}{ll} b_c^{A:c'} = \frac{sim(c,c')\cdot p_{c'}^A}{(p_{c'}^A + n_{c'}^A + 2)} & d_c^{A:c'} = \frac{sim(c,c')\cdot n_{c'}^A}{(p_{c'}^A + n_{c'}^A + 2)} \\ u_c^{A:c'} = 1 - \frac{sim(c,c')\cdot(p_{c'}^A + n_{c'}^A)}{(p_{c'}^A + n_{c'}^A + 2)} & a_c^{A:c'} = a_{c'}^A. \end{array} \quad (9)$$

Equations (9) and (8) are pretty similar, except for the $sim(c, c').(p_{c'}^A + n_{c'}^A)$ factor in the weighing operator. In the following section we use a 95 % t-student and Wilcoxon signed-rank statistical test to prove that the difference due to that factor is not statistically significant for large values of $sim(c, c')$ (at least 0.5).

3.3 Evaluations

We show empirically the similarity between the weighing and the discounting.[1]

First Experiment: Discounting and Weighing in a Real-Life Case. We propose here a first validation of the similarity between weighing and discounting by using both of them in the process of estimation of the trustworthiness of a series of tags derived from a cultural heritage crowdsourcing project.

Steve social tagging project dataset. For the purpose of our evaluations, we use the "Steve Social Tagging Project" [25] data (in particular, the "Researching social tagging and folksonomy in the ArtMuseum"), which is a collaboration of museum professionals and others aimed at enhancing social tagging. In our experiments, we used a sample of tags which the users of the system provided for the 1784 images of the museum available online. Most of the tags were evaluated by the museum professionals to assess their trustworthiness. The tags can be single words or a string of words provided by the user regarding any objective aspect of the image displayed to them for the tagging. We used only the evaluated tags for our experiments.

Gathering evidence for evaluation. We select a very small set of semantically related tags, by using a Web-based *WordNet* interface [23]. We then gather the list of users who provided the tags regarding the chosen words and count the number of positive and the negative evidence. The chosen tags are only three (Asian, Chinese and Buddhist), and they correspond to 206 entries in total (i.e., they are associated 206 times to one or more pictures by one or more users). This represents a small sample compared to the total number of tag entries (0.5 %). However, this experiment is meant only to exemplify the use of the semantic similarity measure when one needs to compute an opinion about a new context (e.g., "Chinese"), given two existing ones (e.g., "Asian" and "Buddhist"). Therefore, we consider the *Chinese-Asian* pair (semantic similarity 0.933) and the *Chinese-Buddhist* pair (semantic similarity 0.6667). We refer to the second experiment for a

[1] Complete results are available at http://trustingwebdata.org/books/URSW_III/slsw.zip.

more indicative evaluation. The opinions are calculated using two different methods. First by weighing the evidence with the semantic relatedness using Eq. (8) and the second method is by discounting the evidence with the semantic relatedness using Eq. (9).

Results. We employ the Student's t-test and the Wilcoxon signed-rank test to assess the statistical significance of the difference between two sample means. At 95 % confidence level, both tests show a statistically significant difference between the two means. This difference, for the *Chinese-Asian* pair is 0.025, while for the *Chinese-Buddhist* pair is 0.11, thanks also to the high similarity (higher than 0.5) between the considered topics. Having removed the average difference from the results obtained from discounting (which, on average, are higher than those from weighing), both the tests assure that the results of the two methods distribute equally.

Second Experiment: Discounting and Weighing on a Large Simulated Dataset. In the Steve.Museum dataset, the average number of annotations provided by a given user is limited (about 20). To check if the two methods for building subjective opinions using semantic similarity measures are significantly different, we built a large dataset consisting of 1000 sample tags and we treated the tags as if they were contributed by the same user. In this manner, we could check if the two methods present relevant differences both when the evidence amount is small or large. We perform the Student's t-test and the Wilcoxon signed-rank test to evaluate the hypothesis that the two methods are not statistically significantly different. For semantic relatedness values $sim(c, c') > 0.7$, the mean difference between the belief values obtained by weighing and discounting is 0.092. Thus with 95 % confidence interval, both tests assure that both the weighing operator and the discounting operator produce similar results. The semantic similarity threshold $sim(c, c') > 0.7$ is relevant and reasonable, because it becomes more meaningful to compute opinions for a new context based on the opinions provided earlier for the most semantically related contexts, while also in case of lack of evidence for a given context, evidence about a very diverse context can not be very significant.

4 Combining Probabilistic Semantic Similarity Measures Within Subjective Logic

The second extension that we propose regards the use of probabilistic semantic similarity measures within subjective logic.

Wikipedia Relatedness Measure. The Wu and Palmer measure introduced above is a deterministic semantic similarity measure, because it is deterministically computed based on the position of the two examined words in WordNet. We want to exemplify the adoption in subjective logic of semantic similarity measures belonging to another class, that is the probabilistic class of measures.

These measures determine the semantic similarity between two words in a statistical manner, by checking the occurrence and co-occurrence of the two words within a large corpora of documents. A famous example of this kind of similarity measures is the Normalized Google Distance [9], which uses Google as a corpus of documents.

We use the Wikipedia [27] relatedness measure, as defined by Milne et al. [20,21] because of its easiness of use. This distance adapts the Normalized Google Distance to use Wikipedia as a corpus of reference for computation. The Wikipedia similarity distance is defined as follows:

$$sim(c, c') = \frac{log(max(|A|, |B|) - log(|AB|)}{log(|W|) - log(min(|A|, |B|)} \tag{10}$$

where $|A|$ and $|B|$ are the cardinalities of the set of documents containing s_1 and s_2 respectively, and $|W|$ is the size of Wikipedia.

Moreover, Milne et al. provide a disambiguation confidence score for the measure, that ranges between zero and one.

4.1 Wikipedia Relatedness Measure as a Subjective Opinion

As in the previous section, given two synsets (s_1 and s_2), we name c and c' the respective context identified by them. To differentiate from the previous section, we use $measure'$ as a placeholder for probabilistic similarity measures.

The elements at our disposal from the Wikipedia distance are:

- $sim(c, c') \in [0, 1]$ is the semantic relatedness between two synsets a and b;
- $conf(c, c') \in [0, 1]$ is the confidence in the semantic relatedness between a and b.

To represent the Wikipedia distance in subjective logic, we need to map all its elements to specific elements (or combinations of elements) of subjective logic, while taking into account the logic's constraints and mechanisms (e.g., the fact that $b + d + u = 1$). We provide a mapping for each of the elements above, and we provide a motivation for them as follows.

1. $conf(s1, s2) = 1 - u_{c=c'}^{measure'}$ because the confidence value determines exactly the portion of probability mass that is certain. Therefore, the remaining part of the probability mass is assigned to the uncertainty element of subjective opinions.
2. $E_{c=c'}^{measure'} = sim(c, c')$. That is, the expected value of the subjective opinion should coincide with the similarity between the two synsets considered.
3. $b_{c=c'}^{measure'} = conf(c, c') \cdot sim(c, c')$ because the certain part of an opinion $(1 - u)$ is assigned $b + d$. Thus, we assign this mass proportionally to the value of the similarity measure, to represent our belief in the two synsets being semantically related.

However, given the constraints of subjective logic, by virtue of Eq. (2) that we report as follows,

$$E_{c=c'}^{measure'} = b_{c=c'}^{measure'} + a_{c=c'}^{measure'} \cdot u_{c=c'}^{measure'}$$

we obtain

$$sim(c, c') = a_{c=c'}^{measure'}$$

which is, of course, wrong. The similarity value might depend on the subjective opinion's prior, but if the equation above holds, then we do not even need to compute the opinion, since the a priori value would already give the similarity value.

We propose, then, two mappings between subjective opinions and probabilistic semantic similarity measures, each of them satisfying two of the three requirements above. Of the three requirements, only the first one is considered as unavoidable, because of the definition of the uncertainty of subjective opinions.

Definition 2 (Wikipedia relatedness measure of two synsets as a subjective opinion (expected value as semantic similarity)). *We define a subjective opinion capturing the similarity between* $synset_a$ *and* $synset_b$ *using the Wikipedia distance as follows:*

$$sim(c, c') \equiv \omega_{c=c'}^{mesaure'}(b_{c=c'}^{measure'}, d_{c=c'}^{measure'}, u_{c=c'}^{measure'}) \tag{11}$$

where

$$\begin{aligned} b_{c=c'}^{measure'} &= sim(c, c') - a_{c=c'}^{measure'} + a_{c=c'}^{measure'} \cdot conf(c, c') \\ d_{c=c'}^{measure'} &= sim(c, c') + a_{c=c'}^{measure'} - a_{c=c'}^{measure'} \cdot conf(c, c') + conf(c, c') \\ u_{c=c'}^{measure'} &= 1 - conf(c, c'), \end{aligned} \tag{12}$$

hence

$$E_{c=c'}^{mesaure'} \equiv sim(c, c'). \tag{13}$$

We provide here motivation for the mapping that we propose. We treat the confidence value $conf(c, c')$ as the inverse of the uncertainty of a subjective opinion. In fact, we interpret the confidence as the percentage of probability mass confidently assigned by the semantic relatedness: the semantic relatedness ranges between zero and one, but we are confident on only $conf(c, c')\%$ of that mass. The rest of the probability mass $(1 - conf(c, c'))$ is, indeed, uncertain.

We also set the expected value of the opinion to coincide with the similarity value, that is:

$$E_{c=c'}^{measure'} = sim(c, c')$$

From this, given Eq. (2), and having set $u_{c=c'}^{measure'} = 1 - conf(c, c')$, follows that:

$$\begin{aligned} b_{c=c'}^{measure'} &= E_{c=c'}^{measure'} - a_{c=c'}^{measure'} \cdot (1 - conf(c, c')) = \\ &= sim(c, c') - a_{c=c'}^{measure'} \cdot (1 - conf(c, c')) = \\ &= sim(c, c') - a_{c=c'}^{measure'} + a_{c=c'}^{measure'} \cdot conf(c, c') \end{aligned}$$

and

$$d_{c=c'}^{measure'} = 1 - b_{c=c'}^{measure'} - u_{c=c'}^{measure'} =$$
$$= 1 - (sim(c,c') - a_{c=c'}^{measure'} + a_{c=c'}^{measure'} \cdot conf(c,c')) - (1 - conf(c,c')),$$

so

$$d_{c=c'}^{measure'} = sim(c,c') + a_{c=c'}^{measure'} - a_{c=c'}^{measure'} \cdot conf(c,c') + conf(c,c').$$

In this manner we define an opinion that reflects our constraints, that is: (1) uncertainty as inverse of the confidence of the semantic similarity value and (2) semantic similarity value as expected value of the subjective opinion. However, this mapping has the undesirable consequence that the belief $b_{c=c'}^{measure'}$ and the disbelief $d_{c=c'}^{measure'}$ depend on the a priori value $a_{c=c'}^{measure'}$. So, we propose an alternative mapping.

Definition 3 (Wikipedia relatedness measure of two synsets as a subjective opinion (belief as semantic similarity times confidence)). *We propose here an alternative mapping that allows a subjective opinion to capture the similarity between synset a and b using the Wikipedia distance. The mapping is defined as follows:*

$$sim(c,c') \equiv \omega_{c=c'}^{measure'}(b_{c=c'}^{measure'}, d_{c=c'}^{measure'}, u_{c=c'}^{measure'}), \tag{14}$$

where

$$\begin{aligned} b_{c=c'}^{measure'} &= conf(c,c') \cdot sim(c,c') \\ d_{c=c'}^{measure'} &= conf(c,c') \cdot (1 - sim(c,c')) \\ u_{c=c'}^{measure'} &= 1 - conf(c,c'). \end{aligned} \tag{15}$$

Again, we set the constraint $u_{c=c'}^{measure'} = 1 - conf(c,c')$, however we do not bind the expected value of the opinion to be equal to $sim(c,c')$.

4.2 Using Wikipedia Relatedness Measure as a Subjective Opinion

We have shown in the previous subsection that we can represent Wikipedia relatedness measures between two synsets or two words by means of subjective opinion. As with many other subjective logic operators [17], we propose two possible mappings for the probabilistic semantic similarity measure. In particular, the second mapping that we propose does not present the undesirable characteristic shown by the first one, that is a dependency between a priori value and belief in the mapped opinion. Of course, these two mappings are different, so we do not check their equivalence, like we did in the previous section for the mapping between subjective logic and probabilistic semantic similarity measures.

Our goal is to show how to represent semantic similarity measures in subjective logic, to import externally defined elements in the logic and increase its capabilities. The choice of the mapping is dependent on the specific constraints given by a domain or an application where the logic is used in combination with the similarity measure, although our preference goes for the second mapping, because the first one presents an already mentioned undesirable dependency between belief and a priori value. The same reasoning applies to the choice of the semantic similarity measure to adopt. Each semantic similarity measure has specific limitations, like the requirement that words are present in a given graph or corpus of documents. The choice of a specific semantic similarity measure is beyond our focus, because it is a very domain- and application-dependent choice.

5 Partial Evidence Observation

The Web and the Semantic Web are pervaded of data that can be used as evidence for a given purpose, but that constitute partially positive/negative evidence for others. Think about the *Waisda?* tagging game [22]. Here, users challenge each other about video tagging. The more users insert the same tag about the same video within the same time frame, the more the tag is believed to be correct. Matching tags can be seen as positive observations for a specific tag to be correct. However, consider the orthogonal issue of the user reputation. User reputation is based on past behavior, hence on the trustworthiness of the tags previously inserted by him/her. Now, the trustworthiness of each tag is not deterministically computed, since it is roughly estimated from the number of matching tags for each tag inserted by the user. The expected value of each tag, which is less than one, can be considered as a partial observation of the trustworthiness of the tag itself. Vice-versa, the remainder can be seen as a negative partial observation. After having considered tag trustworthiness, one can use each evaluation as partial evidence with respect to the user reliability: no tag (or other kind of observation) is used as a fully positive or fully negative evidence, unless its correctness has been proven by an authority or by another source of validation. However, since only rarely the belief (and therefore, the expected value) is equal to one, these observations almost never count as a fully positive or fully negative evidence. We propose an operator for building opinions based on indirect observations, i.e., on observations used to build these opinions, each of which counts as an evidence.

Theorem 3 (Partial evidence-based opinions). *Let $\overrightarrow{P}$ be a vector of positive observations (e.g. a list of "hits" or "match" counts) about distinct facts related to a given subject s. Let l be the length of $\overrightarrow{P}$. Let each opinion based on each entry of $\overrightarrow{P}$ have an a priori value of $\frac{1}{2}$. Then we can derive an opinion about the reliability of the subject in one of the following two manners.*

– *By cumulating the expected values (counted as partial positive evidence) of each opinion based on each element of p:*

$$b_s = \frac{1}{l+2}\Sigma_{i=1}^{l}\frac{p_i+1}{p_i+2} \quad d_s = \frac{1}{l+2}\Sigma_{i=1}^{l}\frac{1}{p_i+2} \quad u_s = \frac{2}{l+2}. \qquad (16)$$

– *By averaging the expected values of the opinions computed on each of the elements of p:*

$$b_s = \frac{1}{3l}\Sigma_{i=1}^{l}\frac{p_i+1}{p_i+2} \quad d_s = \frac{1}{3} - \frac{1}{3l}\Sigma_{i=1}^{l}\frac{1}{p_i+2} \quad u_s = \frac{2}{3}. \tag{17}$$

Proof. For each "fact" about s we have at our disposal a count of positive pieces of evidence. We treat each fact as an observation about the trustworthiness of s. Examples of these observations are tags inserted by s in a crowdsourcing platform, and the items of evidence are the approvals or matches that these tags obtain. We do not set an upper limit to the amount of positive evidence. Rather, we convert it into a subjective opinion and we compute its expected value as follows (remember that no negative evidence is registered):

$$E_i = b_i + a_i \cdot u = \frac{p_i}{p_i+2} + \frac{1}{2}\cdot\frac{2}{p_i+2} = \frac{p_i+1}{p_i+2}. \tag{18}$$

E is considered as partial positive evidence. If p is an extremely high number, then E is approximated to 1. Otherwise, $1 - E$ is considered as partial negative evidence. Given that we have l pieces of partial evidence (because we have l distinct elements in $\overrightarrow{P}$), we compute the opinion about s following Eq. (1). Here we have two possibilities. If l contains evidence about distinct and independent facts, then we can cumulate all the pieces of evidence (represented as $E_i, 1 - E_i$) and by setting:

$$p_s = \Sigma_{i=1}^{l}\frac{p_i+1}{p_i+2} \qquad n_s = \Sigma_{i=1}^{l}\frac{1}{p_i+2},$$

we obtain Eq. (16). In fact, we consider each item of $\overrightarrow{P}$ as providing an observation about s.

If, instead, $\overrightarrow{P}$ contains dependent observations, then it makes sense to average them in order to uniformly represent the evidence about s. In this case, we set:

$$p_s = \frac{1}{l}\Sigma_{i=1}^{l}\frac{p_i+1}{p_i+2} \qquad n_s = \frac{1}{l}\Sigma_{i=1}^{l}\frac{1}{p_i+2}.$$

Following again Eq. (1), we obtain Eq. (17). Note that, in this case, we use only the average of the observation as item of evidence. Therefore, we have only one item of evidence. This justifies the fact that in Eq. 17 we always have 3 as denominator: following Eq. (1), $p + n + 2 = 1 + 2 = 3$. □

More often, we use Eq. 16, because we consider the cases where evidence from different and independent facts about the same individual are provided.

For instance, consider the *Waisda?* tagging game [22]. Here, users challenge each other in tagging videos. Whenever two tag entries for the same video from two different users match, the users get a score. Indeed a matched tag has a higher probability to be correct, and the goal of the game is to collect reliable tags by incentivizing the users. How can we estimate a user reliability? Suppose that a user *user* added two different tags about two different videos. One of them got five matches, the other got two. We can compute a subjective opinion about *user* that represents his reputation using Eq. (16) and we obtain:

$$\omega_{user}\left(\frac{1}{l+2}\Sigma_{i=1}^{l}\frac{p_i+1}{p_i+2}, \frac{1}{l+2}\Sigma_{i=1}^{l}\frac{1}{p_i+2}, u_s=\frac{2}{l+2}\right)=$$
$$\omega_{user}\left(\frac{1}{4}\left(\frac{6}{7}+\frac{3}{4}\right), \frac{1}{4}\left(\frac{1}{7}+\frac{1}{4}\right), \frac{1}{4}\left(\frac{2}{4}\right)\right)=$$
$$\omega_{user}\left(\frac{55}{112}, \frac{11}{112}, \frac{1}{2}\right).$$

If, instead, the two tags got the same scores as before, but they were inserted for the same video in different matches, we can average their contribution, since they provide indications about the user reliability in the same situation. What we obtain using Eq. (17) is:

$$\omega_{user}\left(\frac{1}{3l}\Sigma_{i=1}^{l}\frac{p_i+1}{p_i+2}, \frac{1}{3}-\frac{1}{3l}\Sigma_{i=1}^{l}\frac{1}{p_i+2}, \frac{2}{3}\right)=$$
$$\omega_{user}\left(\frac{1}{6}\left(\frac{6}{7}+\frac{3}{4}\right), \frac{1}{3}-\frac{1}{6}\left(\frac{6}{7}+\frac{3}{4}\right), \frac{2}{3}\right)=$$
$$\omega_{user}\left(\frac{55}{168}, \frac{1}{168}, \frac{2}{3}\right).$$

This method, and in particular Eq. (16), has been adopted and implemented in a work of ours [3].

6 Dirichlet Process-Based Opinions: Open World Opinions

We present here an extension of subjective logic that allows using Dirichlet processes called "open world opinions". We start from introducing Dirichlet processes, and then describe the extension.

6.1 Preliminaries: Dirichlet Process

The Dirichlet process [10] is a stochastic process representing a probability distribution whose domain is a random probability distribution. As we previously saw, the binomial and multinomial opinions are equivalent to Beta and Dirichlet probability distributions. The Dirichlet distribution represents an extension of the Beta distribution from a two-category situation to a situation where one among n possible categories has to be chosen. A Dirichlet process over a set S is a stochastic process whose sample path (i.e. an infinite-dimensional set of random variables drawn from the process) is a probability distribution on S. The finite dimensional distributions are from the Dirichlet distribution: if H is a finite measure on S, α is a positive real number and X is a sample path drawn from a Dirichlet process, written as

$$X \sim DP(\alpha, H), \tag{19}$$

then for any partition of S of cardinality m, say $\{B_i\}_{i=1}^{m}$

$$(X(B_1), \ldots, X(B_m)) \sim Dirichlet(\alpha H(B_1), \ldots, \alpha H(B_m)). \tag{20}$$

Moreover, given n draws from X, we can predict the next observation as:

$$obs_{n+1} = \begin{cases} x_i^*(i \in [1 \ldots k]) & \text{with probability } \frac{n(x_i^*)}{n+\alpha} \\ H & \text{with probability } \frac{\alpha}{n+\alpha} \end{cases} \tag{21}$$

where x_i^* is one of the k unique values among the observations gathered.

6.2 Open World Opinions

Having to deal with real data coming from the Web, which are accessed incrementally, the possibility to update the relative probabilities of possible outcomes might not be enough to deal with them. We may need to handle unknown categories of data which should be accounted and manageable anyway. Ceolin et al. [8] show how it is important to account for unseen categories, when dealing with Web data. Here, we propose a particular subjective opinion called "open world opinion" which accounts for partial knowledge about the possible outcomes. A subjective opinion resembles personal opinion provided by sources with respect to facts. Open world opinions represent the case when something about a given fact has been observed, but the evidence allows also for some other (not yet observed) outcome to be considered as plausible. With this extension we allow the frame of discernment to have infinite cardinality. In practice, open world opinions allow to represent situations when the unknown outcome of an event can be equal to one among a list of already observed values (proportionally to the amount of observations for each of them), but it is also possible that (and so some probability mass is reserved for cases where) the outcome is different from what has been observed so far, and is drawn from an infinitely large domain.

Definition 4 (Open world opinion). *Let: X be a frame of infinite cardinality, $\alpha \in \mathbb{R}^+$, k be the number of categories observed, $\overrightarrow{P}$ be the array of evidence per category, $\overrightarrow{B}$ be a belief function over X, and H be a continuous function representing the prior probabilities for all the categories considered. New observations will belong to the previously observed categories with probability determined by the previous observations and to a new category with a probability determined by the parameter α (that determines the uncertainty u). New categories are drawn from H. We define the open world opinion ω_x as:*

$$\omega_x(\overrightarrow{B}, u, H)$$
$$b_{x_i} = \frac{p_{x_i}}{\alpha + \Sigma_{i=1}^{k} p_{x_i}} \quad u = \frac{\alpha}{\alpha + \Sigma_{i=1}^{k} p_{x_i}} \quad 1 = u + \Sigma_{x_i} b_{x_i}. \tag{22}$$

Definition 5 (Expected value of open world opinion). *The expected value of a category x_i given an open world opinion is computed as follows:*

$$E(x_i, \omega_x(\overrightarrow{B}, u, H))) = b_{x_i} + H(x_i) \cdot u = \frac{p_{x_i} + \alpha \cdot H(x_i)}{\alpha + \Sigma p_{x_t}} = \frac{p_{x_i}}{\alpha + \Sigma p_{x_t}}, \quad (23)$$

where $\overrightarrow{P}$ is the array of evidence observed, α is the concentration parameter, that determines how frequently observations belonging to new categories are likely to appear (we often set this parameter to 1 as default value; higher values imply that observations belonging to new categories are drawn with high probability), and H is the base distribution of the Dirichlet process, that is the probability distribution from which new categories are drawn with probability $\frac{\alpha}{\alpha+\Sigma p_{x_t}}$. Given that H is a continuous probability distribution (and hence with an infinite number of values), the probability of x_i determined by H is zero.

Theorem 4 (Equivalence between the subjective and Dirichlet process notation). *Let $\omega_X^{bn} = (\overrightarrow{B}, U, H)$ be an open world opinion expressed in belief notation, and $\omega_X^{pn} = (\overrightarrow{P}, \alpha, H)$ be an opinion expressed in probabilistic notation (i.e., using the notation from a Dirichlet process; $\overrightarrow{P}$ is an array of evidence), both over the same frame X. ω_X^{bn} and ω_X^{pn} are equivalent when the following mappings holds:*

$$\begin{cases} b_{x_i} = \dfrac{p_{x_i}}{\alpha + \Sigma_{i=1}^{k} p_{x_i}} \\ \\ u = \dfrac{\alpha}{\alpha + \Sigma_{i=1}^{k} p_{x_i}} \end{cases} \Leftrightarrow \begin{cases} p_{x_i} = \dfrac{\alpha b_{x_i}}{U} \\ \\ 1 = u + \Sigma b_{x_i}. \end{cases} \quad (24)$$

Proof. Each step of the Dirichlet process can be seen as a Dirichlet distribution. Hence the mapping between Dirichlet distributions and multinomial opinions [14] holds also here. □

Theorem 5 (Mapping between open world opinion and multinomial opinion). *Let $\omega 1_y^x(\overrightarrow{B}, u, H)$ be an open world opinion and let $\omega 2_y^x(\overrightarrow{B}, u, \overrightarrow{a})$ be a multinomial opinion. Let X_2 and Θ_2 be the frame and the frame of discernment of $\omega 2_y^x$. Let $\{b_i\}_{i=1}^k$ be the result of the partition of $dom(H)$ such that:*

1. $|\Theta_2| = |\{b_i\}|$;
2. $\bigcup\{b_i\}_{i=1}^k = dom(H)$;
3. $\forall\{x_i\}[(\{x_i\} \in X_2 \wedge |\{x_i\}| = 1 \wedge x_i \in b_j) \Rightarrow \nexists x_{k \neq j} \in b_i]$.
4. *$W = k$, where W is the non-informative constant of multinomial opinions*

Then there exists a function $D : Dom(H) \rightarrow \{b_i\}$ such that $D(\omega 1_y^x) = \omega 2_y^x$.

Proof. The equivalence between the discretized open world opinion and the multinomial opinion is proven by showing that:

- given Eq. (20), since the partition $\{b_i\}_{i=1}^k$ covers the entire $dom(H)$, then the partition distributes like the corresponding Dirichlet distribution;
- to each category of $\omega 2_y^x$ corresponds one and only one partition of $\{b_i\}$ as per item 2 of Theorem 5. □

In other words, open world opinions extend multinomial opinions by allowing the frame of discernment Θ to be infinite. However, by properly discretizing an open world opinion, what we obtain is an equivalent multinomial opinion.

6.3 Example: Using Open World Opinions

Here we illustrate an example of the use of open world opinions. Piracy at sea is a well-known problem. Every year, several ships are attacked, hijacked, etc. by pirates. The International Chamber of Commerce has created a repository of reports about ship attacks.[2] van Hage et al. [26] have created an enriched Semantic Web version of such a repository, the Linked Open Piracy (LOP).[3] On the basis of LOP, one might think to be able to predict the frequency of attacks from one year based on the previously available data. However, a problem arises in this situation, since new attack types appear every year and this makes that frequencies vary. Ceolin et al. [8] have shown how the Dirichlet process can be employed to model such situations. Having the possibility to represent this information by means of an open world opinion adds the power of subjective logic to the Dirichlet process based representation. We can merge contributions from different sources, taking into account their reliability. Moreover, we can combine these facts with others in a logical way and then estimate the opinion (and the corresponding probability to be true) of the consequent facts. By using open world opinions, we can easily apply usual subjective logic operators to these data and easily represent them in a way that takes into account basic provenance information (e.g. data source) when applying fusing or discounting operators. For instance, if according to LOP, in Asia in 2010 we had ten hijacking events and ten attempted boardings, then we would represent this as:

$$\omega^{LOP}_{Attacks\ in\ Asia\ in\ 2010}([0.48, 0.48], 0.04, U(0, 1)).$$

If our opinion about LOP is that is a reliable but not fully accountable source (e.g. $\omega^{us}_{LOP}(0.8, 0.1, 0.1)$), then we can take this information into account by weighing the opinion given by LOP as follows:

$$\omega^{us}_{LOP}(0.8, 0.1, 0.1) \otimes \omega^{LOP}_{Attacks\ in\ Asia\ in\ 2010}([0.48, 0.48], 0.04, U(0, 1)) =$$
$$= \omega^{us:LOP}_{Attacks\ in\ Asia\ in\ 2010}([0.384, 0.384], 0.232, U(0, 1)).$$

The resulting weighted opinion is more uncertain than the initial one, because, even though the two observed types are more likely to happen, the small uncertainty about the source reliability makes the other probabilities to rise.

[2] http://www.icc-ccs.org
[3] http://semanticweb.cs.vu.nl/lop

Having represented this information in subjective logic allows us to reason on the data that we gathered. For instance, we could estimate the cost of the insurance premium for a ship that goes along Asian routes, given that insurance companies relate their premiums to the attack predicted to happen. By means of Dirichlet processes these attacks are quite precisely predictable. The determination of the cost of the premium given the happening of some attacks is representable by means of logical statements. In the future, we will develop subjective logic operators that allow to combine logically open world opinions. However, representing these opinions is the first step towards the possibility to permit this kind of reasoning.

Another useful consequence of this representation is the fact that, as we saw in the example above, we take into account the reputation of the source when modeling the opinion. We could also merge contributions from different sources, once we have developed extensions of the fusion operators tailored for open world opinions.

So, in the future we plan to develop subjective operators that extend those currently existing in order to handle open world opinions. The logical operators will allow combining propositions in logical manners (conjunction, disjunction, etc.) and will probably allow ontological reasoning, although this needs to be investigated further, for instance, with respect to the feasibility of subsumption computation et simila. The fusion and discounting operators will allow handling opinions from different sources and accounting for their reliability.

7 Related Work

The core element of subjective logic is the concept of "opinion" that is, the representation that a given source holds with respect to the truth value of a given proposition. Subjective logic's operators allow combining opinions in different manners, and their development has been widely investigated. Remarkably, the averaging and cumulative fusion [13,14] (i.e., operators that allow averaging or cumulating opinions about the same proposition from different sources) and the discounting [16] (i.e., the operator that allows weighing a source's opinion based on the source's reputation) operators are among the most generic and useful operators for this logic. These operators provide the foundations for the work proposed in this chapter. The connections between subjective logic and the (Semantic) Web are increasing. Ceolin et al. [7] adopt this logic for computing trust values of annotations provided by experts, using DBpedia and other Web sources as evidence. Unlike this work, they do not use semantic similarity measures. Ceolin et al. [4,6] and Bellenger et al. [1] provide applications of the combination of evidential reasoning with semantic similarity measures and Semantic Web technologies. In this chapter we provide the theoretical foundations for these approaches, and we generalize them. Sensoy et al. [24] use semantic similarity in combination with subjective logic to import knowledge from one context to another. They use the semantic similarity measure to compute a prior value for the imported data, while we use it to weigh all the available evidence.

Kaplan et al. [18] focus on the exploration of uncertain partial observations used for building subjective opinions. Unlike their work, we restrict our focus on partial observations of Web-like data and evaluations, which comprise the number of "likes", links and other similar indicators related to a given Web item. The weighing and discounting based on semantic similarity measures can resemble the work of Jøsang et al. [13], although the additional information that we include in our reasoning (that is, semantic similarity) is related only to the frame of discernment in subjective logic, and not to the belief assignment function.

8 Conclusion

We show the potential for employing subjective logic as a basis for reasoning on Web and Semantic Web data. We show that it can be really powerful for handling uncertainty and how little extensions can help in improving the mutual benefit that Semantic Web and subjective logic obtain from cooperating together. We propose the use of semantic similarity measures, both deterministic (in particular, the Wu and Palmer similarity measure) and probabilistic ones (in particular, the Wikipedia semantic relatedness), within subjective logic. Part of this work is based on previously mentioned practical applications that show the usefulness of it, and here we provide theoretical foundations for it.

Second, we propose a means to represent subjective opinions on the basis of partial evidence, which is a common phenomenon on the Web (e.g. number of hits or number of tweets). This operator has been employed in a few empirical works already, but here we provide a formal definition for it.

Lastly, we extend subjective opinions to model Dirichlet processes. These have shown to be particularly useful to represent at least some Web datasets. We introduce open world opinions to incorporate Dirichlet processes in subjective logic.

We plan to investigate further the integration of semantic similarity measures in subjective logic, to make it more uniform, and possibly provide best practices that help choosing the right measure and mapping for a given set of requirements. Also, we will provide additional operators for managing open world opinions. We foresee that other extensions will be possible as well like, for instance, the usage of hyperopinions [15] to handle subsumption reasoning about uncertain data.

Acknowledgements. This research was supported by the Data2Semantics and SEALINC Media projects in the Dutch national program COMMIT.

References

1. Bellenger, A., Gatepaille, S., Abdulrab, H., Kotowicz, J.-P.: An evidential approach for modeling and reasoning on uncertainty in semantic applications. In: URSW, pp. 27–38. CEUR-WS.org (2011)
2. Budanitsky, A., Hirst, G.: Evaluating Wordnet-based measures of lexical semantic relatedness. Comput. Linguist. **32**(1), 13–47 (2006)

3. Ceolin, D., Groth, P., van Hage, W., Nottamkandath, A., Fokkink, W.: Trust evaluation through user reputation and provenance analysis. In: 8th International Workshop on Uncertainty Reasoning for the Semantic Web, pp. 15–26 (2012)
4. Ceolin, D., Nottamkandath, A., Fokkink, W.: Automated evaluation of annotators for museum collections using subjective logic. In: Dimitrakos, T., Moona, R., Patel, D., McKnight, D.H. (eds.) IFIPTM 2012. IFIP AICT, vol. 374, pp. 232–239. Springer, Heidelberg (2012)
5. Ceolin, D., Nottamkandath, A., Fokkink, W.: Subjective logic extensions for the semantic web. In: URSW, vol. 900, pp. 27–38. CEUR-WS.org (2012)
6. Ceolin, D., Nottamkandath, A., Fokkink, W.: Semi-automated assessment of annotation trustworthiness. In: PST, pp. 325–332. IEEE Computer Society (2013)
7. Ceolin, D., Van Hage, W., Fokkink, W.: A trust model to estimate the quality of annotations using the Web. In: WebSci10. Web Science Trust (2010)
8. Ceolin, D., van Hage, W., Fokkink, W.: Estimating the uncertainty of categorical web data. In: URSW, pp. 15–26. CEUR-WS.org (2011)
9. Cilibrasi, R.L., Vitanyi, P.M.B.: The Google similarity distance. IEEE Trans. Knowl. Data Eng. **19**(3), 370–383 (2007)
10. Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. Ann. Stat. **2**, 209–230 (1973)
11. Fink, D.: A compendium of conjugate priors. Technical report, Cornell University (1995)
12. Jøsang, A.: A logic for uncertain probabilities. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **9**(3), 212–279 (2001)
13. Jøsang, A., Daniel, M., Vannoorenberghe, P.: Strategies for combining conflicting dogmatic beliefs. In: FUSION, pp. 1133–1140. IEEE (2003)
14. Jøsang, A., Diaz, J., Rifqi, M.: Cumulative and averaging fusion of beliefs. Inf. Fusion **11**(2), 192–200 (2010)
15. Jøsang, A., Hankin, R.: Interpretation and fusion of hyper opinions in subjective logic. In: FUSION, pp. 1225–1232. IEEE (2012)
16. Jøsang, A., Marsh, S., Pope, S.: Exploring different types of trust propagation. In: Stølen, K., Winsborough, W.H., Martinelli, F., Massacci, F. (eds.) iTrust 2006. LNCS, vol. 3986, pp. 179–192. Springer, Heidelberg (2006)
17. Jøsang, A., McAnally, D.: Multiplication and comultiplication of beliefs. Int. J. Approximate Reasoning **38**(1), 19–51 (2005)
18. Kaplan, L., Chakraborty, S., Bisdikian, C.: Subjective logic with uncertain partial observations. In: FUSION, pp. 565–572. IEEE (2012)
19. Miller, G.A.: WordNet: a lexical database for English. Commun. ACM **38**(11), 39–41 (1995)
20. Milne, D., Witten, I.H.: An effective, low-cost measure of semantic relatedness obtained from Wikipedia Links. In: Wikipedia and Artificial Intelligence: An Evolving Synergy, pp. 25–30. AAAI Press (2008)
21. Milne, D., Witten, I.H.: Learning to link with Wikipedia. In: CIKM, pp. 509–518. ACM (2008)
22. Netherlands Inst. for Sound and Vision. Waisda? (2012). http://wasida.nl
23. Princeton University. Wordnet::Similarity (2012). http://marimba.d.umn.edu/cgi-bin/similarity/similarity.cgi
24. Sensoy, M., Pan, J., Fokoue, A., Srivatsa, M., Meneguzzi, F.: Using subjective logic to handle uncertainty and conflicts. In: TrustCom, pp. 1323–1326. IEEE (2012)
25. U.S Institute of Museum and Library Service. Steve Social Tagging Project (2012)
26. van Hage, W., Malaisé, V., van Erp, M.: Linked Open Piracy: a story about e-Science, linked data, and statistics. J. Data Semant. **1**(3), 187–201 (2012)

27. Wikimedia Foundation. Wikipedia: The Free Encyclopedia (2014). http://www.wikipedia.org
28. Wu, Z., Palmer, M.: Verbs semantics and lexical selection. In: ACL '94, pp. 133–138. ACL (1994)