An Enhanced Framework of Subjective Logic for Semantic Document Analysis

Sukanya Manna, B. Sumudu. U. Mendis, and Tom Gedeon

School of Computer Science, The Australian National University, Canberra, ACT 0200, Australia
{sukanya.manna,sumudu.mendis,tom.gedeon}@anu.edu.au

Abstract. Unlike propositional logic which works on truth or falsity of statements, human judgements are subjective in nature having certain degree of uncertainty. Two different people will analyse and interpret a document in two different ways based on their background and current focus. In this paper we present an enhanced framework of subjective logic for automated single document analysis where each sentence in the document represents a proposition, and 'opinions' are constructed about this proposition to focus the degree of uncertainty associated with it. The 'opinion' about a sentence determines the significance of that sentence in a document. The input arguments are built automatically from a document in the form of evidence; then they are analyzed based on subjective logic parameters. Two different approaches are described here. The first utilises "bag of words" concept. However, this approach tends to miss the underlying semantic meanings of the context, so we further enhanced it into the latter approach which incorporates semantic information of the context, by extending the basic definitions of subjective logic.

1 Introduction

Subjective logic [1] is a logic which operates on subjective beliefs about the world, and uses the term *opinion* to denote the representation of a subjective belief. An 'opinion' can be interpreted as a probability measure containing secondary uncertainty, and as such subjective logic can be seen as an extension of both probability calculus and binary logic. It is suitable for modeling and analysing situations involving uncertainty and incomplete knowledge [1], [2].

Jøsang et al. [2] claims that, subjective logic is mainly designed to apply and interpret different real world problems in artificial intelligence reliability analysis [3], authentication [4], and legal reasoning [5] where evidence is gathered from multiple sources with manual intervention like the case of open systems. Subjective logic also seems very suitable for reasoning about intrusion attacks because on the one hand an attack can be considered to be a crisp event, i.e. an attack either takes place or not, while on the other beliefs about intrusion can have varying degrees of certainty [6]. By analogy we can infer that any kind of decision making process, which works on crisp event but has uncertainty associated with its judgement or consequence can be dealt with subjective logic.

V. Torra, Y. Narukawa, and M. Daumas (Eds.): MDAI 2010, LNAI 6408, pp. 231–242, 2010.

In a document computing area, the picture is quite different; where only source of information is the document itself. This represents more of a closed system where the information source is restricted to a particular origin; which is a document in this case. When analysing single documents using subjective logic, the sets of arguments are generated automatically as evidence from the information available in it. It is mainly done by exploiting the structure and semantics of the text being considered.

When a document is read by a human, they analyse it by identifying the main idea of the source text and filtering what is essential in the information conveyed by the text. This step further involves differentiating complementary or superfluous information according to the intended purposes of the writers, with respect to what they aim at the readers to grasp. In [7], the authors have pointed out that the context of a given piece of text is interpreted and understood by a different person in a different fashion. Thus we see that human understanding and reasoning is *subjective* in nature unlike propositional logic which deals with either truth or falsity of a statement. Furthermore information provided by different persons can be either linguistically or factually different, with a prevalent degree of impreciseness and uncertainty.

In this paper, our main aim is to formulate an enhanced model based on subjective logic to analyse documents in a way which more similar to human judgements capturing uncertainty. Each sentence of a document represents specific facts about the document; we consider them to be propositions and define 'opinions' about these propositions. Thus we present a framework for automatically determining opinions about a sentence, using subjective logic because of its property of 'uncertain probability' measure. We portray two different concepts; 'bag of words' and further enhancement of the model with semantic information from the document; as 'bag of words' tend to lose the semantic binding of the context.

2 Representing Uncertain Probabilities: Subjective Logic (SL) Basics

In subjective logic, first order measure of evidence are expressed as belief mass distribution functions over frame of discernment. All these belief measure representations in subjective logic, which are called '*opinions*', also contain a base rate parameter which express the a priori belief in the absence of evidence. Philosophically, 'opinions' are quantitative representations of evidence as perceived by humans or by other intelligent agents [8]. This portraits a scenario which is an open system where evidence are gathered from different sources.

A frame of discernment Θ contains the set of possible states. It is assumed that the system cannot be in more than one elementary state at the same time. However, if an elementary state is assumed to be true then all the superstate can be considered true as well. In fact Θ is by definition always true because it contains a true state.

The elementary states in the frame of discernment Θ will be called atomic states because they do not contain any substates. The powerset of Θ, denoted by 2^Θ, contains atomic states, and all possible combinations of atomic states, including Θ. A frame of discernment can be finite or infinite, in which cases the corresponding powerset is also finite or infinite.

An observer assigns a belief mass to various states based on its strength of belief that the state (or one of its substates) is true. We have directly taken the basic definitions from the original paper [2] which we have used to build up evidence from a document in our study.

Definition 1 (Belief Mass Assignment). *Let Θ be a frame of discernment. If with each substate $x \in 2^\Theta$ a number $m_\Theta(x)$ is associated such that:*
1. $m_\Theta(x) \geq 0$
2. $m_\Theta(\emptyset) = 0$
3. $\sum_{x \in 2^\Theta} m_\Theta(x) = 1$
then m_Θ is called a belief mass assignment in Θ, or BMA for short. For each substate $x \in 2^\Theta$, the number $m_\Theta(x)$ is called the belief mass of x.

Definition 2 (Belief Function). *Let Θ be a frame of discernment, and let m_Θ be a BMA on Θ. Then the belief function corresponding with m_Θ is the function $b: 2^\Theta \rightarrow [0, 1]$ defined by:*

$$b(x) = \sum_{y \subseteq x} m_\Theta(y), \quad x, y \in 2^\Theta \tag{1}$$

Definition 3 (Disbelief Function). *Let Θ be a frame of discernment, and let m_Θ be a BMA on Θ. Then the disbelief function corresponding with m_Θ is the function $d: 2^\Theta \rightarrow [0, 1]$ defined by:*

$$d(x) = \sum_{y \cap x = \emptyset} m_\Theta(y), \quad x, y \in 2^\Theta. \tag{2}$$

Definition 4 (Uncertainty Function). *Let Θ be a frame of discernment, and let m_Θ be a BMA on Θ. Then the uncertainty function corresponding with m_Θ is the function $u: 2^\Theta\ [0, 1]$ defined by:*

$$u(x) = \sum_{\substack{y \cap x \neq \emptyset \\ y \not\subseteq x}} m_\Theta(y), \quad x, y \in 2^\Theta. \tag{3}$$

From Josang's concept, we can get the **Belief Function Additivity** which is expressed as:

$$b(x) + d(x) + u(x) = 1, \quad x \in 2^\Theta,\ x \neq \emptyset. \tag{4}$$

Definition 5 (Relative Atomicity). *Let Θ be a frame of discernment and let $x, y \in 2^\Theta$. Then for any given $y \neq \emptyset$ the relative atomicity of x to y is the function $a : 2^\Theta \rightarrow [0, 1]$ defined by:*

$$a(x/y) = \frac{|x \cap y|}{|y|}, \quad x, y \in 2^\Theta,\ y \neq \emptyset. \tag{5}$$

It can be observed that $x \cap y = \emptyset \Rightarrow = 0$ and that $y \subseteq x \Rightarrow a(x/y) = 1$. In all other cases relative atomicity will be a value between 0 and 1. The relative atomicity of an atomic state to its frame of discernment, denoted by $a(x/\Theta)$, can simply be written as $a(x)$. If nothing else is specified, the relative atomicity of a state then refers to the frame of discernment.

Definition 6 (Probability Expectation). *Let Θ be a frame of discernment with BMA m_Θ then the probability expectation function corresponding with m_Θ is the function $E: 2^\Theta \rightarrow [0,1]$ defined by:*

$$E(x) = \sum_y m_\Theta(y) a(x/y), \quad x, y \in 2^\Theta. \tag{6}$$

Definition 7 (Opinion). *Let Θ be a binary frame of discernment with 2 atomic states x and $\neg x$, and let m_Θ be a BMA on Θ where $b(x)$, $d(x)$, $u(x)$, and $a(x)$ represent the belief, disbelief, uncertainty and relative atomicity functions on x in 2^Θ respectively. Then the opinion about x, denoted by w_x is the tuple defined by:*

$$w(x) \equiv (b(x), d(x), u(x), a(x)). \tag{7}$$

For compactness and simplicity of notation we will in the following denote belief, disbelief, uncertainty and relative atomicity functions as b_x, d_x, u_x and a_x respectively.

Definition 8 (Ordering of Opinions). *Let ω_x and ω_y , be two opinions. They can be ordered according to the following criteria by priority:*
1. The opinion with the greatest probability expectation is the greatest opinion.
2. The opinion with the least uncertainty is the greatest opinion.
3. The opinion with the least relative atomicity is the greatest opinion.

3 Subjective Logic in Document Analysis

How can we define evidence in a document related to its overall meaning[1]? This is what we are building here automatically. We consider words, phrases or co-occurrence of words, semantic associations, or a sentence itself to be evidence present in a document. Now, based on this, our basic motivation is to formulate 'opinion' about a proposition, which is a sentence in this case. Stronger the opinions about a sentence, more is its significance in the document. These opinions are measured by probability expectation of a sentence as defined in (6). Greater the probability expectation, more significant is the sentence.

3.1 Representation of a Document

Assumptions. We propose the following framework for the practical application of subjective logic in a document computing context.

[1] From here, we simply write 'evidence' to express that the "evidence in a document related to its overall meaning".

1. All the words or terms (removing the stop words) in a document are atomic. However, some sentences can have single word.
2. The sentences are unique, i.e., each of them occur only once in a given document.

A document consists of sentences. In this paper, a sentence is considered to be a set of words. In a document, sentences are separated by stop marks (".", "!", "?"). Terms (stop words excluded) are extracted and the frequencies (i.e. number of occurrences) of the words in each sentence are calculated.

Let us now define the notations which we will be using in the paper. Θ is the frame of discernment. We represent a document as a collection of words, which is

$$\Theta = D_w = \{w_1, w_2, ..., w_n\} \tag{8}$$

where, D_w is a document consisting of words. w_1, $w_2 ... w_n$ and $|D_w| = n$. Now,

$$\rho(\Theta) = \{\{w_1\}, \{w_2\}, ..., \{w_1, w_2, w_3, ..., w_n\}\} \equiv 2^{\Theta} \tag{9}$$

$$|\rho(\Theta)| = 2^n \tag{10}$$

Since a document is a collection of sentences, it can also be represented as

$$D_s = \{s_1, s_2, ..., s_t\} \tag{11}$$

where t is a finite integer and each s_i is an element of $\rho(\Theta)$. Each sentence is comprised of words, which belong to the whole word collection of the document D_w. We thus represent each sentence by,

$$S_l = \{w_i w_k ... w_r\} \in \Theta \tag{12}$$

where, $1 \leq i,\ k,\ r \leq n$ and $S_l \in \rho(\Theta)$.

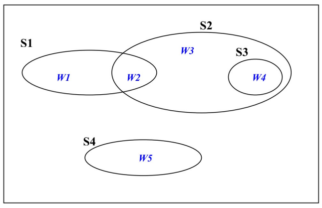
3.2 Example of Documents

Eg:1- A generic example. In fig.1, we illustrate a generic document D with four sentences $D_s = \{s_1, s_2, s_3, s_4\}$ and a list of unique words $D_w = \{w_1, w_2, w_3, w_4, w_5\}$. Atomic events are the single words w_1 to w_5 and the non atomic events are the sentences from s_1 to s_4; but in this case s_3 and s_4 are atomic. Each sentence is composed of both atomic and non atomic events. These are used as evidence for subjective logic formulation in this study.

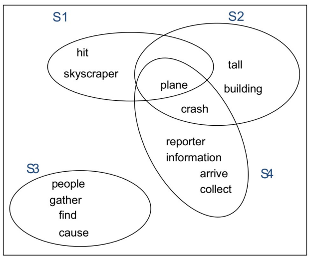
Eg:2- A specific example. Here is another sample document which consists of four different real sentences, $D_s = \{s_1, s_2, s_3, s_4\}$.

1. A plane hits a skyscraper.
2. A plane crashed into a tall building.
3. People gathered to find out the cause.
4. Reporters arrived to collect information about the crash.

We will refer to this example in the following sections for explaining our representations of subjective logic.



(a) Generic example



(b) Real example

Fig. 1. Representation of 'bag of words' form of sentences in a document

3.3 Modeling 'Opinions' about a Sentence in a Document

In this section we present the formulation of 'opinion' about a sentence in a document using subjective logic explained in sec.2. Now, let us explain step-wise computation of opinion based on sec.2 equations for the examples considered. Examples 1 and 2 as shown in subsec.3.2 are of same kind expect the fact that eg.2 represents real words in place of symbols of eg.1. So, in this section, we illustrate the computation for any one of these, i.e, eg.1.

BMA calculation: BMA is explained in def.1. Now, for a document, we calculate BMA for each event by,

$$m(x) = \frac{F(x)}{Z}, \tag{13}$$

where $F(x) = \sum_{k=1}^{t} f_{x_k}$, where t is the total number of sentences in the document, $x \in 2^{\Theta}$, and f_{x_k} is the frequency of occurrence of event x in sentence k. In other words, it is the total frequency of that event in all the sentences (or the whole document).

$$Z = \sum_{\forall x \neq \Phi} F(x), \quad x \in 2^{\Theta} \tag{14}$$

Z is the total frequency of the all the existing events (whose frequency is non zero). In the given example 1, we have 7 valid states and their corresponding frequencies in the document are: $F(\{\{w_1\}) = 1,\ F(\{w_2\}) = 2,\ F(\{w_3\}) = 1,\ F(\{w_4\}) = 2,\ F(\{w_5\}) = 1, F(\{w_1, w_2\}) = 1,\ F(\{w_2, w_3, w_4\}) = 1\}$. Therefore, $Z = 9$ in this case. Using (13), we calculate BMA for each of the states (or events) in the given example shown in fig.1. So, for eg.1, we have $m(w_1) = \frac{1}{9},\ m(w_2) = \frac{2}{9},\ m(s_1) = m(w_1, w_2) = \frac{1}{9}$ ... $m(s_4) = m(w_5) = \frac{1}{9}$

Figure 1(b) is the diagrammatic representation of example 2 of subsec.3.2. The words shown in the diagram are processed by stemming and stop words removed. This is a 'bag of words' representation of the document. Here, the number of atomic states (or events) are 14 and non-atomic states are 4. Now total frequency for all of these 18 states is 21 (which means $Z = 21$) (calculated exactly in the same way as the generic example). Now, using (13), we get the BMA for each of these states respectively; provided the frequency of each non stop words in each sentence is 1 as per example 2; such as, $m(hit) = \frac{1}{21}$, $m(skyscraper) = \frac{1}{21}$, $m(plane) = \frac{2}{21}$, $m(crash) = \frac{2}{21}$...and so on.

Belief, Disbelief, and Uncertainty. Using definitions from sec. 2, we use equations (1), (2), and (3) to calculate the belief, disbelief and uncertainty of a sentence respectively. We illustrate the computation using eg.1's s_1 by,
$b(s_1) = m(w_1) + m(w_2) + m(w_1, w_2) = \frac{4}{9}$
$d(s_1) = m(w_3) + m(w_4) + m(w_5) = \frac{4}{9}$
$u(s_1) = 1-(b(s_1)+d(s_1)) = \frac{1}{9}$; using (4). For eg.2, we calculate these parameters in the same way as shown for eg.1.

Calculation of relative atomicity, probability expectation and 'opinion' about a sentence. Here inorder to calculate probability expectation, we first need to find relative atomicities. Again, using equations (5), (6), and (7) of sec. 2, we compute relative atomicity for sentence s_1 of eg.1 as:
$a(s_1/w_1) = \frac{|s_1 \cap w_1|}{|w_1|} = \frac{1}{1} = 1$
$a(s_1/w_2) = \frac{|s_1 \cap w_2|}{|w_2|} = \frac{1}{1} = 1$
... $a(s_1/w_5) = a(s_1/s_4) = \frac{|s_1 \cap w_5|}{|w_5|} = \frac{0}{1} = 0$
Likewise, we calculate the atomicity for other sentences. So, the probability expectation is then obtained by, $E(s_1) = m(w_1)a(s_1/w_1) + m(w_2)a(s_1/w_2) + m(\{w_1, w_2\})a(s_1/\{w_1, w_2\}) + ... + m(w_5)a(s_1/w_5)$ Thus $E(s_1) = \frac{13}{27} = 0.48$. Thus opinion ($\omega_{s_1}$ or $\omega(s_1)$)about a sentence s_1 can be expressed using these four parameters by (7) as, $\omega(s_1) = (0.44,\ 0.44,\ 0.11,\ 4.33)$. Likewise, we compute the parameters in the same way for eg.2.

4 Extension of Subjective Logic with Semantic Information of a Document

In this section, we extend basic subjective logic model explained in the previous sec.3 where we have already shown, how to define 'opinion' about a sentence in a document considering words, phrases and sentences to be atomic or composite events as different sources of evidence. But we used 'bag of words' for formulating this measure, which is a superficial approach according to information retrieval context. Only root form of words are used for frequency measure where the underlying semantic relations between events are ignored. Hence, here we use semantic similarity as a measure to find relatedness of concepts of sentences whose 'opinions' are desired.

4.1 Why Do We Need Semantic Information?

What we write or say are very context sensitive. A same word can be linguistically expressed differently in different contexts; at the same time, different words can linguistically express same thing at a particular context. If we look at our example 2 of sec.3,

sentence 1: "A plane hits a skyscraper."
sentence 2: "A plane crashed into a tall building."
Anyone can easily infer that both the sentences are similar in their context though different words are used to express it. Similarly, if we look at other two sentences,
sentence 3: "People gathered to find out the cause."
sentence 4: "Reporters arrived to collect information about the crash."
the inference will again be same.

In fig.2, we illustrate 4 sentences with overlap only if the words in them are exactly same with same parts of speech (POS) tag. The dotted lines show which words are most similar in their meanings in sentences. In sentence 1 and 2, phrases 'hits a skyscraper' is similar as 'crashed into a tall building' or the word pairs like 'hits' and 'crash', 'people' and 'reporters', 'gathered' and 'arrived' etc have great similarity in their meanings. Index terms are not enough to find this kind of analogies as they look for only exact matches between words, which in this case failed to find any kind of relations among the sentences of eg.2. We thus extend and redefine subjective logic belief measures by incorporating semantic information about word, phrase, and sentence similarities from the document. To accomplish this, we used WordNet [9] as a lexical dictionary to gather semantic information about each word of sentences; thus making the whole decision making process context sensitive.

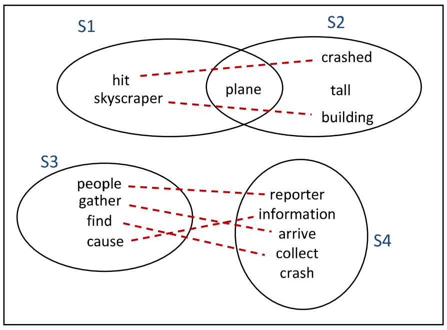


Fig. 2. An example of a document with semantic overlap

4.2 Measure of Semantic Similarity

Two words are contextually similar, if they share similar senses. To perform this automatically, we require WordNet [9], an online lexicon database, to compute this measure. Each word can have one or more synsets based on different senses of their existence also in different parts of speech like noun, verb, adjective, and adverb. Same word in different parts of speech convey different meaning to the context in which they are used. In both sentences 2 and 4, the word 'crash' occurs but in two different parts of speech; *verb* for former and *noun* for latter; obviously imparting different sense to the context. So, considering only root form of any word, misses out the semantic meaning of it. To overcome this problem of 'bag of words' concept, we introduce a similarity measure α.

Definition 9 (Semantic Similarity). *Let Θ be a frame of discernment, and let $x, y \in 2^{\Theta}$. Then for any x and y, semantic similarity is the function $\alpha : 2^{\Theta} \rightarrow [0,1]$ defined by*

$$\alpha(x,y) = SimScore(x,y) \quad x,y \in 2^{\Theta}. \tag{15}$$

where $SimScore(x,y)$ is a function which determines the semantic similarity measure between x and y provided the elements of x and y are in the same parts of speech. This can be any kind of similarity score like gloss overlap [10], path based measures [11], [12], edge based measures or sentence similarity measure [13]. We use a threshold κ to define the degree of similarity. Thus we can say,
$\alpha = 1$, x and y are identical
$\kappa \leq \alpha < 1$, x and y are similar
$0 \leq \alpha < \kappa$, x and y are dissimilar, where $\kappa \in [0,1]$.
Generally $\kappa = 0.5$ is taken as a standard value for similarity scores [14].

4.3 Enhanced Belief Measures Using Semantic Information

In this section, we present an extension of subjective logic formulation for document analysis using semantic information. The equations are redefined using the similarity score α as shown in (15).

Computation of BMA: We compute belief mass assignment in the same way as shown in (13). The only difference is in the frequency calculation of atomic states; where we consider parts of speech of the words as well instead of only the root forms. For example, in eg.2, the word 'crash' is in two different parts of speech (POS) in s_2 and s_4, so these belong to two separate atomic events. Likewise, 'plane' being in same POS (noun) for both sentences 1 and 2 will have a total frequency count of 2 for that state.

Now, for example 2, there are 18 different states existing, and the frequency of each state can be represented as: $F(\{plane^{noun}\}) = 2$, $F(\{hit^{verb}\}) = 1$, $F(\{crash^{noun}\}) = 1$, $F(\{crash^{verb}\}) = 1$, $F(\{building^{noun}\}) = 1$, ..., $F(\{plane^{noun}, hit^{verb}, skyscraper^{noun}\}) = F(s_1) = 1$. Thus we get $Z = 19$ for this case. We compute BMA by (13) using these values computed, $m(plane^{noun}) = \frac{2}{19}$, $m(building^{noun}) = \frac{1}{19}$ and likewise for other events.

Similarity scores for example 2: For different belief measures, we need to use similarity score between two events. Let us assign similarity scores for each word pair belonging to same parts of speech (using example 2). Suppose, sentence 1 be the proposition we considered. So some of the similarity scores which are necessary for finding opinion about s_1 can be:
$\alpha(plane^{noun}_{s_1}, plane^{noun}_{s_2}) = 1$, $\alpha(plane^{noun}_{s_1}, building^{noun}_{s_2}) = 0.1$,
$\alpha(hit^{verb}_{s_1}, crash^{verb}_{s_2}) = 0.7$, $\alpha(skyscraper^{noun}_{s_1}, plane^{noun}_{s_2}) = 0.08$,
$\alpha(skyscraper^{noun}_{s_1}, building^{noun}_{s_2}) = 0.85$, $\alpha(plane^{noun}_{s_1}, people^{noun}_{s_3}) = 0.03$,

$\alpha(plane^{noun}_{s_1}, cause^{noun}_{s_3}) = 0.01$, ...
Likewise we compute α for all other word pairs. For this analysis, we also require similarity between composite events which can be computed using hierarchical document signature [13]. Using this method, word-sentence similarity and sentence-sentence similarity can be computed. Now, let us present some similarities of composite events for s_1,
$\alpha(s_1, s_2) = 0.5,\ \alpha(s_1, plane^{noun}_{s_2}) = 0.8,\ \alpha(s_1, crash^{verb}_{s_2}) = 0.6,\ \alpha(s_1, tall^{adj}_{s_2}) = 0.01$, ..., $\alpha(s_1, people^{noun}_{s_3}) = 0.02,\ \alpha(s_1, find^{verb}_{s_3}) = 0.01,\ \alpha(s_1, gather^{verb}_{s_3}) = 0.01$, ..., $\alpha(s_1, reporter^{noun}_{s_4}) = 0.01,\ \alpha(s_1, s_4) = 0.2$. The values of α shown here are solely based on intuitions and general understanding of semantics of the text considered.

Definition 10 (Semantic Belief Function). *Let Θ be a frame of discernment, m_Θ be a BMA and α be semantic similarity on Θ respectively. Then the belief function corresponding with m_Θ and alpha is the function $b^s : 2^\Theta \rightarrow [0, 1]$ defined by:*

$$b^s(x) = \sum_{\forall y|\alpha(x,y)\leq 1} m_\Theta(y), \quad x, y \in 2^\Theta,\ y \subseteq x \tag{16}$$

Thus, as per the similarity values provided, belief of sentence 1 is computed as,

$$\begin{aligned} b^s(s_1) &= m(plane^{noun}) \times \alpha(s_1, plane^{noun}) + m(hit^{verb}) \times \alpha(s_1, hit^{verb}) \\ &\quad + m(skyscraper^{noun}) \times \alpha(s_1, skyscraper^{noun}) + m(s_1) \times \alpha(s_1, s_1) \\ &= \frac{2}{19} \times 0.8 + \frac{1}{19} \times 0.5 + \frac{1}{19} \times 0.4 + \frac{1}{19} \times 0.8 \end{aligned}$$

Definition 11 (Semantic Disbelief Function). *Let Θ be a frame of discernment, m_Θ be a BMA and α be semantic similarity on Θ respectively. Then the disbelief function corresponding with m_Θ and α is the function $d^s : 2^\Theta \rightarrow [0, 1]$ defined by:*

$$d^s(x) = \sum_{\forall y|\alpha(x,y)<\kappa} \alpha(x, y) m_\Theta(y), \quad x, y \in 2^\Theta. \tag{17}$$

Now for disbelief calculation, we look for $0 \leq \alpha < \kappa$. Here, $\alpha(s_1, people^{noun}_{s_3}) = 0.02$, $\alpha(s_1, gather^{verb}_{s_3}) = 0.01$, $\alpha(s_1, reporter^{noun}_{s_4}) = 0.01$, are all less than $\kappa = 0.5$, s_1 do not have significant semantic overlap with sentences s_3 and s_4. So, they are part of disbelief. Thus,

$$\begin{aligned} d^s(s_1) &= \alpha(s_1, people^{noun}_{s_3}) m(people^{noun}_{s_3}) + \\ &\quad \alpha(s_1, gather^{verb}_{s_3}) m(gather^{verb}_{s_3}) + \\ &\quad ... + \alpha(s_1, reporter^{noun}_{s_4}) m(reporter^{noun}_{s_4}) + \\ &\quad ... + \alpha(s_1, s_3) m(s_3) + ... \\ &= (0.02 \times \frac{1}{19}) + (0.01 \times \frac{1}{19}) + (0.01 \times \frac{1}{19}) + ... \end{aligned}$$

Definition 12 (Semantic Uncertainty Function). *Let Θ be a frame of discernment, m_Θ be a BMA and α be semantic similarity on Θ respectively. Then the disbelief function corresponding with m_Θ and α is the function $u^s : 2^\Theta \rightarrow [0,1]$ defined by:*

$$u^s(x) = \sum_{1>\forall y|\alpha(x,y)\geq\kappa} \alpha(x,y)m_\Theta(y), \quad x,y \in 2^\Theta. \tag{18}$$

In case of uncertainty calculation, we consider $1 > \alpha \geq \kappa$, where $\kappa = 0.5$. Here, $\alpha(s_1, s_2) = 0.5$, $\alpha(s_1, plane_{s_2}^{noun}) = 0.8$, $\alpha(s_1, crash_{s_2}^{verb}) = 0.6$,..., have $\alpha \geq 0.5$; so these implies that s_1 has substantial overlap with s_2. Thus,

$$\begin{aligned} u^s(s_1) = &\ \alpha(s_1, plane_{s_2}^{noun})m(plane_{s_2}^{noun}) + \\ &\ \alpha(s_1, crash_{s_2}^{verb})m(crash_{s_2}^{verb}) + ... \\ &\ \alpha(s_1, s_2)m(s_2) \\ = &\ (0.8 \times \frac{2}{19}) + (0.6 \times \frac{1}{19}) + ... + (0.5 \times \frac{1}{19}) \end{aligned}$$

In this situation, the *Semantic Belief Function* will no longer hold strict additivity like (4) and is thus expressed as:

$$b^s(x) + d^s(x) + u^s(x) \leq 1, \quad x \in 2^\Theta,\ x \neq \emptyset. \tag{19}$$

Definition 13 (Semantic Relative Atomicity). *Let Θ be a frame of discernment, let $x, y \in 2^\Theta$, and let $\alpha(x,y)$ be semantic similarity of x and y. Then for any given $y \neq \emptyset$ the relative atomicity of x to y is the function $a : 2^\Theta \rightarrow [0,1]$ defined by:*

$$a^s(x/y) = \frac{\sum_{j=1}^{|y|} \bigvee_{i=1}^{|x|} \alpha(x_i, y_j)}{|y|}, \quad x,y \in 2^\Theta,\ x_i \in x,\ y_j \in y. \tag{20}$$

where x_i and y_i are atomic elements of x and y respectively. So, according to fig.2, $a^s(s_1/s_2) = \frac{1.0+0.7+0.85}{4}$, where $\alpha(plane_{s_1}^{noun}, plane_{s_2}^{noun}) = 1$, $\alpha(hit_{s_1}^{verb}, crash_{s_2}^{verb}) = 0.7$, and $\alpha(skyscraper_{s_1}^{noun}, building_{s_2}^{noun}) = 0.85$ (assuming α values based on meanings) respectively; but this is not the case when 'bag of words' are considered.

The **probability expectation** and **opinion** will remain same as (6) and (7) except the fact that the parameters will be replaced by the extended parameters based on semantic analysis, and hence represented as,

$$E^s(x) = \sum_y m_\Theta(y)a^s(x/y), \quad x,y \in 2^\Theta. \tag{21}$$

$$w^s(x) \equiv (b^s(x),\ d^s(x),\ u^s(x),\ a^s(x)). \tag{22}$$

Now, using the parameters like belief, disbelief, uncertainty, relative atomicity and BMA computed for s_1 we can get probability expectation (21) and opinion (22).

5 Conclusion

In this paper, we presented an enhanced framework of subjective logic for document analysis. Two different aspects of the model are shown. The former is simple computation of the original subjective logic [2] model using 'bag of words'. For the latter, we redefined all the definitions based on the semantic relatedness of concepts encountered in sentences and have shown how this approach is more significant for document analysis. As a future work we tend to determine the similarity threshold κ automatically by using some optimization algorithms.

References

1. Jøsang, A.: Artificial reasoning with subjective logic. In: Proceedings of the Second Australian Workshop on Commonsense Reasoning, vol. 48 (1997), Perth:[sn]
2. Jøsang, A.: A logic for uncertain probabilities. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9(3), 279–311 (2001)
3. Jøsang, A.: Reliability Analysis with Uncertain Probabilities. In: Proceedings of the 4th International Conference on Probabilistic Safety Assessment and Management (PSAM4). Springer, Heidelberg (1998)
4. Jøsang, A.: An algebra for assessing trust in certification chains. In: Proceedings of the Network and Distributed Systems Security Symposium (NDSS 1999). The Internet Society, San Diego (1999) (Citeseer)
5. Jøsang, A., Bondi, V.: Legal reasoning with subjective logic. Artificial Intelligence and Law 8(4), 289–315 (2000)
6. Svensson, H., Jøsang, A.: Correlation of Intrusion Alarms with Subjective Logic
7. Pardo, T., Rino, L., Nunes, M.: Extractive summarization: how to identify the gist of a text. In: The Proceedings of the 1st International Information Technology Symposium–I2TS, Citeseer, pp. 1–6 (2002)
8. Jøsang, A.: Belief Calculus. ArXiv Computer Science e-prints (June 2006)
9. Miller, G.: WordNet: a lexical database for English. Communications of the ACM 38(11), 41 (1995)
10. Lesk, M.: Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In: Proceedings of the 5th Annual International Conference on Systems Documentation, pp. 24–26. ACM, New York (1986)
11. Lin, D.: Using syntactic dependency as local context to resolve word sense ambiguity. In: Annual Meeting-Association For Computational Linguistics, vol. 35, pp. 64–71. Association For Computational Linguistics (1997)
12. Resnik, P.: Using information content to evaluate semantic similarity in a taxonomy. In: International Joint Conference on Artificial Intelligence, Citeseer, vol. 14, pp. 448–453 (1995)
13. Manna, S., Gedeon, T.: Semantic Hierarchical Document Signature In Determining Sentence Similarity. In: Proceedings of the 19th International Conference on Fuzzy Systems (accepted 2010)
14. Achananuparp, P., Hu, X., Shen, X.: The evaluation of sentence similarity measures. In: Song, I.-Y., Eder, J., Nguyen, T.M. (eds.) DaWaK 2008. LNCS, vol. 5182, pp. 305–316. Springer, Heidelberg (2008)