Enhancement of Subjective Logic for Semantic Document Analysis Using Hierarchical Document Signature

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

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

Abstract. In this paper, an extension of Subjective Logic (SL) is presented which uses semantic information from a document to find 'opinions' about a sentence. This method computes semantic overlap of events (words or sentences) using Hierarchical Document Signature (HDS) and uses it as evidence to formulate SL belief measures to order sentences according to their importance. Stronger the opinion, more is the significance. These significant sentences then form extractive summaries of the document. The experimental results show that summaries generated by this method are more similar to human generated ones have outperformed the baseline summaries on average over all the data sets considered.

Keywords: subjective logic, opinions, evidence, Hierarchical Document Signature, semantic, summarization.

1 Introduction

Subjective logic is a type of probabilistic logic that explicitly takes uncertainty and belief ownership into account. In general, subjective logic is suitable for modeling and analysing situations involving uncertainty and incomplete knowledge. In subjective logic (SL) [4], first order measure of evidence is 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 [5].

In document computing context, SL is used by Manna et al. [15], to find 'opinion' about a sentence where a sentence is regarded as a proposition; considering the fact that determining importance of a sentence from a document is very subjective and depends on a reader's interest and motivation. The approach discussed in [15] is based on 'bag of words' concept where events are represented by words and their co-occurrences. This is enhanced in [14] where a theoretical framework is proposed using the semantic information.

K.W. Wong et al. (Eds.): ICONIP 2010, Part I, LNCS 6443, pp. 298–306, 2010.

Aim of this paper is to look at the practical application of the theoretical framework presented in [14] for extracting significant sentences from a document using the underlying semantic information of the context. Here, semantic overlap or similarity of concepts or events are computed using Hierarchical Document Signature (HDS) [12] and Fuzzy Word Similarity (FWS) measure [13] over simple 'bag of words' as in [15].

2 Measure of Semantic Similarity

Two words are contextually similar, if they share similar senses. To perform this automatically, an online lexicon database, WordNet [17] is required 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. The word 'hit' when is a *verb*, then it means collision; again when it is a *noun*, then it means something popular as in cases of 'movie hit' and so on. Only considering root form of any word misses out the semantic meaning of it. To overcome this problem of 'bag of words' concept, a similarity measure α is introduced in this paper for finding semantic similarity of concepts.

Definition 1 (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 $\alpha : SimScore(x, y) \rightarrow [0, 1]$ defined by*

$$\alpha(x, y) = SimScore(x, y) \quad x, y \in 2^{\Theta}. \tag{1}$$

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 [6], path based measures [11], [18], edge based measures, [13] or sentence similarity measure [12]. A threshold κ is used to define the degree of similarity. Thus,
$\kappa = 1$, x and y are identical
$\psi \leq \kappa < 1$, x and y are similar
$0 \leq \kappa < \psi$, x and y are dissimilar, where $\psi \in (0, 1)$.
Generally $\kappa = 0.5$ is taken as a standard value for similarity scores [2]. But in this paper, similarity of events is computed using HDS and $\kappa = 0.4$ is used throughout due to fuzzy approximations in the model and by thorough testing on thousands of data.

2.1 Similarity Computation Using Hierarchical Document Signature (HDS)

Hierarchical Document Signature (HDS) [12] is a special form of Fuzzy Signature (FS) [16] meant for document analysis. For computing semantic similarity of contexts, HDS is generally used. Its levels are based on the natural hierarchy of a document and can vary depending on applications. In this case, parts of speech (POS) is a level which deals with the semantic information of the sentences of a document. The aggregations at different level contribute to the final similarity score. It uses Fuzzy Word Similarity (FWS) [13] at the word level to deal with the similarity or relatedness of a word pair; which then propagates to the next higher level using proper aggregations and continues till it reaches the document level. In this application, sentence-sentence similarity, sentence-word similarity are computed using HDS; which can be called as the similarity between events where events can be either atomic or composite.

For this case, the membership functions of the fuzzy sets to compute FWS as used in [12] are further tuned. In [13] it is seen that triangular membership functions performed better than trapezoidal at the output. Here after tuning it is seen that trapezoidal can also be used at the output and it also gives reasonably good results. The new membership functions for the input and output are:

Table 1. Co-ordinates of trapezoidal membership functions of input and output fuzzy sets used here

INPUT low	(0, 1)	(0.2, 1)	(0.3, 0)	
INPUT medium	(0.2, 0)	(0.3, 1)	(0.4, 1)	(0.5, 0)
INPUT high	(0.4, 0)	(0.5, 1)	(1, 1)	
OUTPUT low	(0, 1)	(0.1, 1)	(0.3, 0)	
OUTPUT medium	(0.1, 0)	(0.3, 1)	(0.5, 1)	(0.7, 0)
OUTPUT high	(0.5, 0)	(0.7,1)	(1, 1)	

Besides the membership functions, similarity threshold is determined to be 0.4 after repeated experiments.

3 Extension of SL Belief Measures Using Semantic Information

In this section, extension of subjective logic formulation for document analysis using semantic information. The equations of [15] are redefined using the similarity score as computed in (1).

Definition 2 (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.

The states/substates are also called events. In this paper, events can be atomic or composite depending on whether they are single words or a sentence (as a sentence is composed of many words).

3.1 BMA Calculation

BMA is explained in def.2. Now, for a document, BMA for each event is calculated by,

$$m(x) = \frac{F(x)}{Z}, \tag{2}$$

where $F(x) = \sum_{k=1}^{N} f_{x_k}$, where N 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 words, it is the total frequency of that event in all the sentences (or the whole document).

$$Z = \sum_{\substack{\forall x \neq \emptyset \\ f_x \neq 0}} F(x), \quad x \in 2^{\Theta} \tag{3}$$

Z is the total frequency of the all the existing events whose frequency is non zero. In calculating frequency of atomic states, the POS of words are considered instead of root form as in [15]. For example, suppose the word ‘crash’ is in two different pars of speech (POS) in two different sentences, so these are considered to be two separate atomic events. Likewise, for the word ‘plane’ when it occurs in two sentences in the form of noun then a total count of 2 will be considered for that state.

Definition 3 (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:*

$$b^s(x) = \sum_{\forall y | \alpha(x,y) \leq 1} m_{\Theta}(y), \quad x, y \in 2^{\Theta},\ y \subseteq x \tag{4}$$

Definition 4 (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:*

$$d^s(x) = \sum_{\alpha(x,y) < \kappa} \alpha(x,y)\, m_{\Theta}(y), \quad x, y \in 2^{\Theta}. \tag{5}$$

Definition 5 (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:*

$$u^s(x) = \sum_{1 > \alpha(x,y) \geq \kappa} \alpha(x,y)\, m_{\Theta}(y), \quad x, y \in 2^{\Theta}. \tag{6}$$

In this situation, the **Semantic Belief Function** will no longer hold strict additivity like ([15]) and is thus expressed as:

$$b^s(x) + d^s(x) + u^s(x) \leq 1, \quad x \in 2^{\Theta},\ x \neq \emptyset. \tag{7}$$

Definition 6 (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:*

$$a^s(x/y) = \frac{\sum_{j=1}^{|y|} \bigvee_{i=1}^{|x|} \alpha(x_i, y_j)}{|y|}, \quad x, y \in 2^{\Theta}, \quad \alpha(x_i, y_j) \geq \kappa \tag{8}$$

where x_i and y_i are atomic elements of x and y respectively.

Definition 7 (Semantic Probability Expectation). *Let Θ be a frame of discernment with BMA m_{Θ} then the semantic probability expectation function corresponding with m_{Θ} is the function $E : 2^{\Theta} \rightarrow [0, 1]$ defined by:*

$$E^s(x) = \sum_y m_{\Theta}(y) a^s(x/y), \quad x, y \in 2^{\Theta}. \tag{9}$$

Thus semantic **Opinion** about a proposition can be written as,

$$\omega^s(x) \equiv (b^s(x),\ d^s(x),\ u^s(x),\ a^s(x)). \tag{10}$$

A sentence is considered significant if it has stronger opinion. In other words, if it has greater probability expectation and lower uncertainty, more it is significant in the document.

4 Experiment

This section presents the experiment and evaluation results from evidence based subjective logic model discussed above. The significant sentences are extracted and then they are compared with human generated summaries provided with the dataset. As a benchmark, baseline summaries are also provided, which are compared in the same way with the human assessors and difference of results are noticed.

4.1 Data Set

In this experiment DUC2001 data set [1] is used for evaluation. The documents are grouped based on a specific topic. DUC2001 comes with human generated summaries and baseline summaries, providing a good platform for evaluation.

4.2 Generation of Summaries

Summaries are broadly classified into text extraction and text abstraction [9], [7]. For text extraction, sentences from the documents are used as summaries and for text abstraction important pieces of information are extracted and then stitched together to form summaries following some linguistic rules. This evidence based model can be used as a text extraction as the original sentences are used for summary generation.

Semantic evidence based model (SPEU): Section 3 describes how to compute opinion about a sentence. All the SL belief measures are computed for each sentence. Then the sentences are arranged based on their descending order of semantic Probability Expectation (SPE) and ascending order of uncertainty (U). Then 30% [3] of the top ranked sentences are extracted and formed extractive summary of the document.

4.3 Evaluation by ROUGE

ROUGE [8] stands for Recall-Oriented Understudy for Gisting Evaluation. It includes measures to automatically determine the quality of a summary by comparing it to other (ideal) summaries created by humans. In this experiment, the result with ROUGE-1 (n-gram, where n=1) at 95% confidence level are presented. ROUGE is sensitive to the length of the summaries [10]; hence the length is fixed to 100 words for the evaluation.

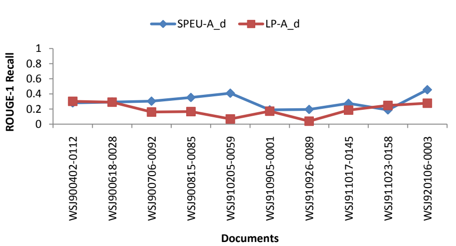
4.4 Results

DUC2001 dataset is used for this experiment. Among different document sets, the evaluation with 'daycare', 'healthcare', 'pres92' and 'robert.gates' data sets are presented. Our method (SPEU) and baseline summaries (denoted by LP) with two different human assessors are compared. For each set the assessors are different. The average table. 2 results show that our method out performs the baseline summaries which is rather good.

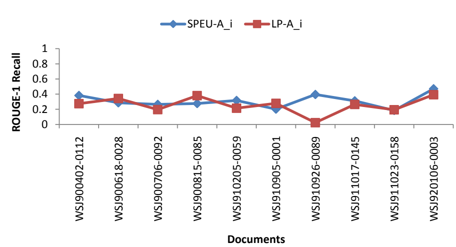
Figures 1 to 4 illustrate per document wise comparison for each data set. It is known that human judgements are subjective, whose evidence is seen through the summaries generated by human assessors. For *daycare* data, SPEU performs better than LP. Though in some documents, the results varied slightly.

For *healthcare* data similar results are noticed except for assessor j, where SPEU generated summaries are not as good as baseline summaries.

Again, for figures 3 and 4, SPEU performs better than LP for most of the cases showing higher similarity with the human generated summaries. For some documents, it is seen that SPEU's performance dropped. The main reason for this is the coverage pitfall of the WordNet used. If a document has lots of unknown

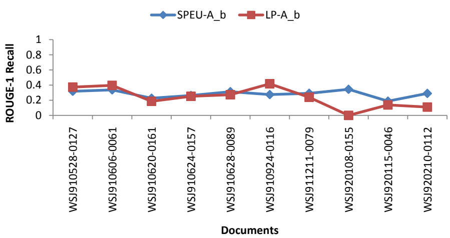


(a) Comparison with Assessor d

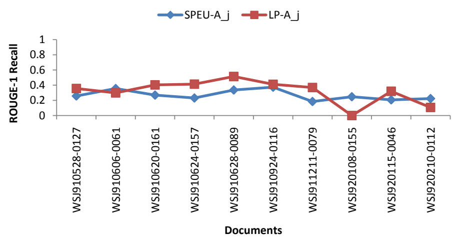


(b) Comparison with Assessor i

Fig. 1. Daycare dataset

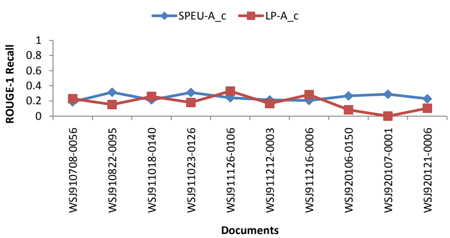


(a) Comparison with Assessor b

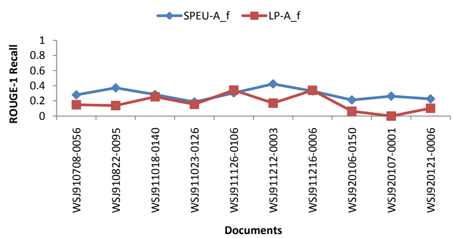


(b) Comparison with Assessor j

Fig. 2. Healthcare dataset

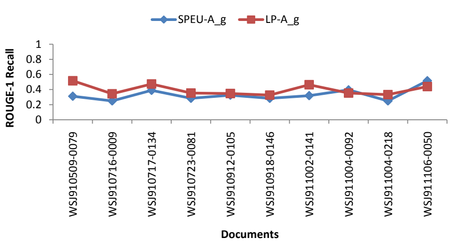


(a) Comparison with Assessor c

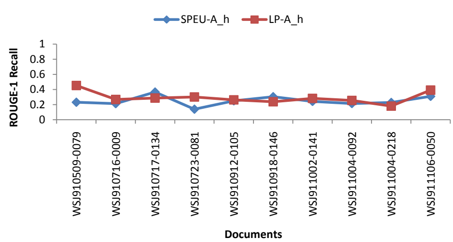


(b) Comparison with Assessor f

Fig. 3. Pres92 dataset



(a) Comparison with Assessor g



(b) Comparison with Assessor h

Fig. 4. Robert.Gates dataset

words, and if they are not found in WordNet, then SPEU cannot function properly as HDS and FWS are WordNet dependent for its semantic analysis. Besides this, it is also noticed that POS tagger which is used here failed to tag properly in many cases, as a result of which it affected HDS' performance, which directly affected SPEU method described here.

Now, in table 2, the results are averaged over the whole data sets. Here SPEU shows higher similarity with human assessors than LP.

Table 2. Summary of all four sets of results (ROUGE-1 Recall)

	LP-A_1	LP-A_2	SPEU-A_1	SPEU-A_2
daycare(d,i)	0.19	0.26	0.29	0.31
healthcare(b,j)	0.27	0.35	0.28	0.27
pres92(c,f)	0.20	0.19	0.24	0.29
robert.gates(g,h)	0.40	0.29	0.33	0.25
Average	0.26	0.27	0.29	0.28

5 Conclusion

This paper presented a semantic extension of Subjective Logic for document analysis. To formulate this, HDS was used to compute similarity between composite events. Using the semantic information, it is noticed that the opinions about a sentence are more subjective and similar to human decisions. The evaluation of the highly opinionated sentences when compared with baseline summaries of the DUC data set shows that they are more similar to human assessors. The possible reason is the use of semantic information for formulating. Noise entered in the model while using the NLP tools at the same time less coverage of information by WordNet have influenced the performance to a greater extent for some documents. Thus as a future work, these aspects needs to be taken care of by using fuzzy methods to reduce noise in using HDS, at the same time membership functions of FIS needs tuning at word level for further refinement of the fuzzy outputs using different machine learning methods.

References

1. DUC, The document understanding conference (2001), `http://duc.nist.gov/`
2. 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)
3. Dalianis, H.: SweSum-A Text Summarizer for Swedish (2000), `http://www.dsv.su.se/%7Ehercules/papers.Textsumsummary.html`
4. Jøsang, A.: A logic for uncertain probabilities. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9(3), 279–311 (2001)
5. Josang, A.: Belief Calculus. ArXiv Computer Science e-prints (June 2006)
6. 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 Press, New York (1986)
7. Liddy, E.D.: The discourse-level structure of empirical abstracts: an exploratory study. Information Processing and Management 27(1), 55–81 (1991)
8. Lin, C.-Y.: Rouge: A package for automatic evaluation of summaries. Association for Computational Linguistics, Barcelona, Spain, pp. 74–81 (July 2004)

9. Lin, C.Y., Hovy, E.: Identifying topics by position. In: Proceedings of the Fifth Conference on Applied Natural Language Processing, pp. 283–290. Morgan Kaufmann Publishers Inc, San Francisco (1997)
10. Lin, C.Y., Hovy, E.: Automatic evaluation of summaries using n-gram cooccurrence statistics. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, vol. 1, P. 78. Association for Computational Linguistics (2003)
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. Manna, S., Gedeon, T.: Semantic Hierarchical Document Signature In Determining Sentence Similarity. In: 2010 IEEE World Congress On Computational Intelligence, 19th International Conference on Fuzzy Systems, pp. 2780–2787. IEEE, Los Alamitos (2010)
13. Manna, S., Mendis, B.S.U.: Fuzzy Word Similarity: A Semantic Approach Using WordNet. In: 2010 IEEE World Congress On Computational Intelligence, 19th International Conference on Fuzzy Systems, pp. 1761–1768. IEEE, Los Alamitos (2010)
14. Manna, S., Mendis, B.S.U., Gedeon, T.: An Enhanced Framework Of Subjective Logic For Semantic Document Analysis (to appear). In: The 7th International Conference on Modeling Decisions for Artificial Intelligence, Springer, Heidelberg (2010)
15. Manna, S., Mendis, B.S.U., Gedeon, T.: Advances in knowledge discovery and data mining, 14th pacific-asia conference, pakdd 2010, hyderabad, india, proceedings. part ii. In: Zaki, M.J., Yu, J.X., Ravindran, B., Pudi, V. (eds.) Advances in Knowledge Discovery and Data Mining. LNCS, vol. 6119, Springer, Heidelberg (2010)
16. Mendis, B.S.U.: Fuzzy Signatures: Hierarchical Fuzzy Systems and Alpplications (PhD thesis). PhD thesis, College of Engineering and Computer Science, The Australian National University, Australia (2008)
17. Miller, G.A.: WordNet: a lexical database for English. Communications of the ACM 38(11), 41 (1995)
18. Resnik, P.: Using information content to evaluate semantic similarity in a taxonomy. In: International Joint Conference on Artificial Intelligence, vol. 14, pp. 448–453. Citeseer (1995)