

Processing of Beliefs extracted from Reported Speech in Newspaper Articles

Ralf Krestel and René Witte
Institut für Programmstrukturen und Datenorganisation (IPD)
Universität Karlsruhe (TH), Germany
krestel|witte@ipd.uka.de

Sabine Bergler
Department of Computer Science and Software Engineering
Concordia University, Montréal, Canada
bergler@cs.concordia.ca

Abstract

The growing number of publicly available information sources makes it impossible for individuals to keep track of all the various opinions on one topic. The goal of our *artificial believer* system presented in this paper is to extract and analyze statements of opinion from newspaper articles.

Beliefs are modeled using a fuzzy-theoretic approach applied after NLP-based information extraction. A fuzzy believer models a human agent, deciding what statements to believe or reject based on different, configurable strategies.

Keywords

Belief Processing, Textual Entailment, Artificial Believer

1 Introduction

With the possibility to gain access to huge amounts of information, for example via the Internet, the Natural Language Processing (NLP) research community has developed whole branches that deal explicitly with vast amounts of information encoded in written natural language¹. One goal is to gain knowledge about irrefutable facts like “The number of inhabitants of city X” or the “Name of the president of country X.” But a lot of information, especially within newspaper articles, are not hard facts that could be easily proven right or wrong. Often newspaper articles contain different views of the same event, or state controversial opinions about a certain topic. In this case the notion of *belief* becomes relevant.

For humans this is a daily task, sometimes a conscious act, sometimes unconsciously adopted. Depending on context information and background knowledge, together with other belief structures, humans tend to believe certain statements while other statements get rejected. Although everybody uses the term *belief*, the definition is rather vague, and the processes taking place inside the brain while “someone is believing something” are not understood. The process of believing

also varies between different humans, not only depending on different background knowledge but on different attitudes towards a coherent worldview or importance and ability of logical thinking.

A computational system, whose task should be to simulate a human newspaper reader by imitating his belief processing, must take into account not only the beliefs (of others) stated in an article, but also the existing beliefs held by the system. Such an *artificial believer*² should also be able to distinguish between different belief strategies, modeling the different human approaches.

Our *Fuzzy Believer* system models a human newspaper reader who develops his own point of view for current events described in newspaper articles. More specifically, we only rely on information stated within the grammatical construct of *reported speech*. This allows a clear assignment of statements to sources and enables the system to judge according to different degrees of reliability in a source.

Our approach differs from existing work by addressing two different problems usually dealt with in isolation: opinion extraction/mining and recognizing textual entailment. Solving these two tasks is necessary to implement an artificial believer. The area of opinion mining [3, 5, 7] is dominated by systems limited to extraction, where the processing of the extracted opinions is rather rudimentary. On the other side are systems that deal with the relation of two sentences to each other [6, 10, 12, 14]. The Pascal RTE Challenge [2, 4] has led to the development of a number of new systems dealing with inference or entailment.

Our fuzzy believer combines these approaches and thereby presents an application capable of “reading” and evaluating newspaper articles. To internally represent the extracted statements and process the different beliefs, we employ fuzzy set theory techniques [18]. Fuzzy set theory explicitly expresses the intrinsic fuzziness in natural language, and the handling of ambiguities and similarities in natural languages is done in a more robust way than crisp approaches. Another reason we chose a fuzzy approach are the existing fuzzy operations for representation and belief revision [15].

To summarize, the system we present in this paper

¹ for example, *Information Extraction, Summarization, or Information Retrieval*

² this term was coined by [1]

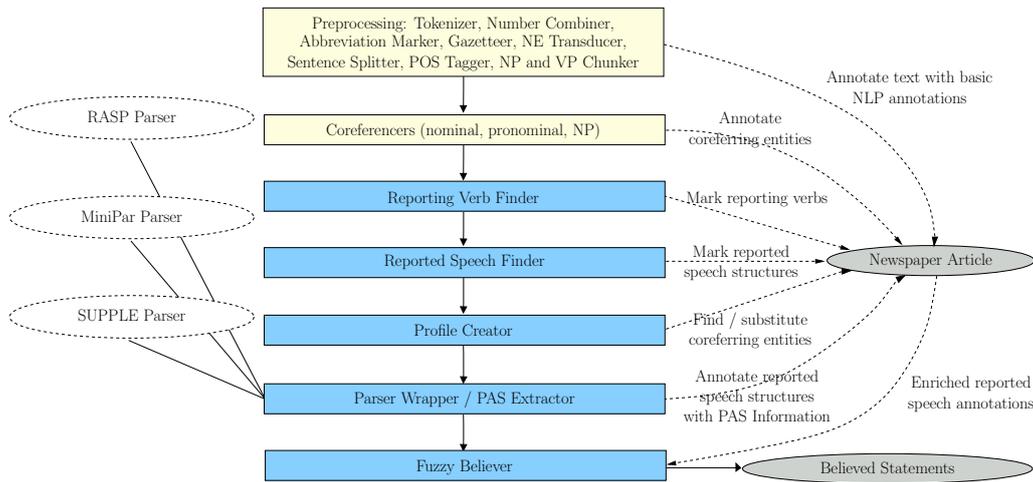


Fig. 1: Overview of the Fuzzy Believer system components

addresses various problems within the NLP domain. Our main contributions making our research significant are: 1. Developing rules to identify and extract reported speech from newspaper articles; 2. processing the gained information by applying fuzzy set theory to natural language processing; 3. creating a working implementation of these ideas, together with an evaluation environment.

The remainder of this paper is structured as follows: In the next section, we give an overview of our fuzzy believer system, followed by a more detailed description of the main component in Section 3. An evaluation of our approach, using different corpora and evaluation methods, is presented in Section 4. Section 5 discusses related work, followed by conclusions in Section 6.

2 System Overview

The starting point of our system is a selection of newspaper articles. Different components are used to realize specific tasks within the system to process the input documents, as can be seen in Fig. 1.

After preprocessing an input document, a first important step is to identify noun phrases. These structures are important to identify acting entities, like persons within a text. We do full noun phrase coreference resolution making use of an existing coreferencer [17].

The next step is to identify and *extract reported speech* within the document (see Fig. 1: *Reporting Verb Finder*, *Reported Speech Finder*). For this part, patterns had to be developed representing the different ways to express reported speech.

Afterwards, we have to combine the results found in the last two steps. The coreference component can identify the same source of two different reported speech utterances enabling us to *build profiles* (see Fig. 1: *Profile Creator*).

The core of our system is the processing of the information encoded in the profiles. We use external parsers to extract predicate-argument structures (see Fig. 1: *PAS Extractor*) as basis for further processing. Our focus lies thereby on the analysis of the extracted PASs of the reported speech utterance and the gener-

ation of held beliefs from it in the last step.

Finding the *entailment* relation between two sentences is the most complex part, and an active research field [2, 4]. Do they express the same, similar things, contradicting things, or are they totally independent? Our approach uses fuzzy set theory and WordNet³ to tackle this question.

The final step is, after trying to “understand” what has been said and by whom, to define what the system should actually believe. The Fuzzy Believer thus has to do processing on the created belief structure. To model different human “believers,” the Fuzzy Believer component (see Fig. 1: *Fuzzy Believer*) uses different *believe strategies*.

The result of the system is a set of propositions the system “believes,” and a set of propositions the system rejected.

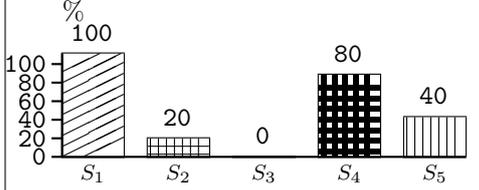
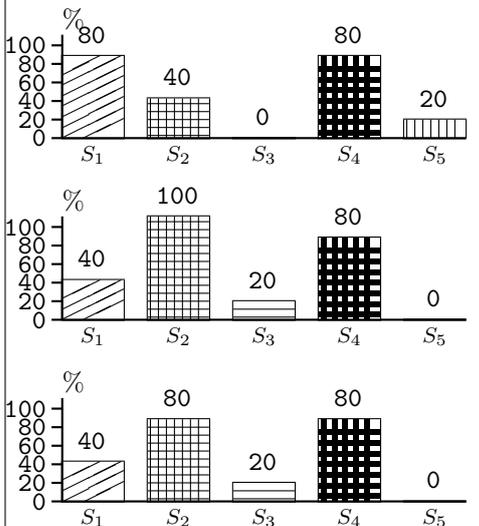
In the next section, we will describe the main component of our system. For more details about the other components concerned with steps 1 to 3 covered in the example shown in Fig. 2, we direct the reader to [9].

3 Fuzzy Believer

The Fuzzy Believer component uses predicate-argument structures extracted from the output of a parser to group the statements of the newspaper articles according to common topics. To process these predicate-argument structures using fuzzy set theory later on, we need to consider a few constraints: The basic set for fuzzy operations to work on has to be limited to statements dealing with the same topic or fact in the world. This is due to the character of fuzzy processing always considering the whole set to perform its operations on. And with beliefs having nothing to do with each other stored in only one single set, we could not use a similarity measure between statements to perform our computations, because this would, for example, lead to the deletion of dissimilar statements dealing with independent topics.

The fuzzy processing task therefore has to consist of four steps:

³ WORDNET, <http://www.wordnet.princeton.edu>

| Step | Example | Description |
|------|---|--|
| 1. | “Preisig worked as a consultant”, one of the employees said. | Sentence in a newspaper article. |
| 2. | [Preisig worked as a consultant](content) [one of the employees](source) [said](reported verb) | Identified reported speech structure. |
| 3. | [Preisig](subject) [work](verb) [consultant](object) | Extracted predicate-argument structure from the output of a parser. |
| 4. | Preisig – work – consultant ($S_1, domain_{38}$) | predicate-argument structure assigned to a domain according to the topic. |
| 5. |  <p style="text-align: right;">Atoms $A_{S_{1,j}}$</p> | Atoms $A_{S_{1,j}}$ for Predicate-argument structure S_1 with correlation grades for all statements in $domain_{38}$ (S_1, \dots, S_5) as computed by heuristic H_1 . (S_1, S_1) = 100 means a 100% possibility that the two statements express the same meaning. (S_1, S_2) = 20 indicates on the other hand that the two statements have probably different meanings. |
| 6. |  <p style="text-align: right;">Formula F_{S_1}</p> <p style="text-align: right;">Formula F_{S_2}</p> <p style="text-align: right;">Formula $F_{S_1} \oplus_{\gamma} F_{S_2}$</p> | Fuzzy belief revision: Result of γ -revision with $\gamma = 0.8$ of the two formulas on top. The first formula represents the existing statements within a domain by combining the different atoms to form literals, then clauses, and then formulas of each statement. The second formula represents the new statement added to the domain. The resulting formula contains only these clauses that are not contradicting the new one, or in other words having a similarity degree of at least 0.8. The interpretation of the result is that the system believes the new statement and all older statements about the same topic that are not contradicting the new one. |

- (1) Grouping statements into domains (topics),
- (2) Finding a fuzzy representation for the statements,
- (3) Identifying the polarity of statements,
- (4) Computing beliefs according to a strategy.

The different strategies make it necessary to identify the *topics* statements deal with. And the Fuzzy Believer has to identify the *polarity* of the statements to detect opposite opinions.

The first task is handled by two heuristics (semantic, based on WORDNET, and syntactic, based on string similarity) that compare the extracted predicate-argument structures (PAS) of two statements. If the heuristics recognize a similarity degree higher than a given threshold between two statements, they are grouped into one domain (topic).

The second and third task is solved by using fuzzy set theory and representing all statements as degrees of similarity between the verbs of the statements in one domain. This similarity is again computed using WORDNET together with the detection of negations and antonyms.

For the fourth task we use three fuzzy set operations (*Union*, *Expansion*, and *Revision*, see [16]) to model various belief strategies.

Domain Finding. The task of this component is to group similar statements together according to their topics to form a *domain*. We use a WORDNET distance measure to find similar, related words and assign a score to each word pair. The threshold of this score can be adjusted as a run-time parameter, allowing a more lenient or a more strict domain classification. As another run-time parameter, the maximum WORDNET distance⁴ can be defined.

A second heuristic currently in use compares the string representation of two words. This is particularly useful for proper nouns that do not occur within the WORDNET database. The score of this heuristic depends on the character overlap of the two words, thus a perfect match is not necessary to gain a score.

To ensure that we compare the appropriate words, an analysis of the main verb is mandatory. We have to differentiate between active and passive constructs, exchanging the syntactic subject and the syntactic object.

The requirements for two predicate-argument structures to match are that at least two element pairs have a matching score of at least the defined threshold. This threshold can be set as a run-time parameter, and al-

⁴ we use the same WORDNET distance as [17]

lows for more strict or more lenient domain classification.

An advantage of dividing the domain classification and the actual matching finding process is that we can use different thresholds for the fuzzy process of assigning statements to different domains and discover supporting and conflicting statements. One statement can belong to more than one domain, exploiting the possibilities of a fuzzy set representation again. The result of this component is shown in Fig. 2 in step 4.

Fuzzy Representation. Every predicate-argument structure is presented as its degrees of similarity with other PASs in the same domain. Fig. 2 step 5 gives an example of the representation of one PAS that is an element of a domain containing five PASs.

Polarity Identification. To identify opposing statements, the fuzzy representation of the PASs is evaluated. If the heuristics yielded small values for the degree of similarity, the meaning of the two statements are considered opposing. A threshold makes it possible to decide whether two statements are similar enough to be considered as expressing the same sense or are likely to contain opposing views.

Computing Beliefs. The phenomena of opposing opinions compelling the human reader to take either one side or to believe neither has to be reflected in our fuzzy believer system as well. To model different human believe behavior, the fuzzy believer makes the decision which statements to believe based on different strategies. The result is a set of held beliefs and rejected beliefs after processing newspaper articles. The strategies used to model different human behavior are: (1) Believe everything, (2) believe old news, (3) believe new news, (4) believe majority, (5) believe certain source/reporter/newspaper, and (6) believe weighted majority – a combination of (4) and (5). Let’s take a closer look at one of the strategies. The “believe new news” strategy uses a fuzzy operation called *revision*. The result of a revision of formula 1 with another formula 2 depends on the order of the formulas, as well as on an ordering of the clauses of the formulas. The ordering can be chronological depending on the timestamp of the insertion of the clause into the formula, or any other ordering like ordering according to degrees of certainty or an order relying on the reporter or newspaper. We chose as an order the first way enabling us to model a belief strategy concerned with the chronological order of news.

The revision process compares statement sets, formally represented by fuzzy formulas in conjunctive normal form [16], with each other. If the two statements sets are compatible, the revision process results in a new set containing the fuzzy union of both sets. However, in case some of the statements are contradicting to a degree that is larger than the prescribed minimal consistency γ , the revision operator will remove individual, inconsistent statements from the first set, according to a preference ordering [16]. In the example in Fig. 2 at step 6, we can see the formula generated in previous steps containing two clauses, and below it, the new formula, with which we start the revision. The

result shown at the bottom in Fig. 2 is a new formula containing two clauses. The ordering of the clauses, which determines the sequence of processing, is again defined by the date of the statements.

4 Evaluation

So far, we performed a detailed evaluation of the individual components of our system. For the evaluation of the reported speech component, as well as for a more detailed evaluation, see [9].

Domain Finding. The evaluation of the domain finding component includes the comparison of the results obtained with RASP, MiniPar, and manually annotated predicate-argument structures. The test data we use is taken from the MSR corpus⁵ and comprises 116 paraphrase pairs. We treat all sentences as content of a reported speech construct. The best result for recall is 81% and best precision value obtained 52% with a different configuration. Detailed results also including manual PAS annotated test data can be found in [9].

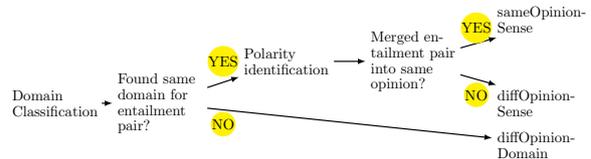


Fig. 2: Polarity identification evaluation strategy

Polarity Finding. The data that comes closest to the conditions we need are the entailment pairs of the PASCAL challenge corpus [2]. We tested different configurations and computed accuracy for two settings. For one experiment, we included all results in the evaluation counting the entailment pairs that were not grouped into the same domain by the domain classification as non-entailing. In Fig. 2, these are the pairs in the “diffOpinion-Domain” category. Here, the best results were around 55% accuracy. The other test setting only considered the sentence pairs that were actually grouped into the same domain by the domain classification component (in Fig. 2 the “same/diffOpinion-Sense” category) yielding an accuracy of 58% using MiniPar-extracted PASs. Table 1 gives an overview of the obtained results with the configuration settings in the table meaning, from left to right: Maximum WordNet Distance between (1) subjects, (2) verbs, (3) objects of two statements. (4) indicates whether a new statement has to match with one (lenient) or all (strict) statements within one domain and (5) is the threshold for assigning the same polarity to a statement.

5 Related Work and Discussion

The extraction of opinions from newspaper articles [3] or customers reviews [5, 7] has become an active re-

⁵ MSR-corpus, http://research.microsoft.com/research/nlp/msr_paraphrase.htm

| Configuration | Accuracy | | | |
|-------------------|----------------|---------|------------|---------|
| | Sense & Domain | | Sense Only | |
| | Rasp | MiniPar | Rasp | MiniPar |
| 3-3-3-strict-0.7 | 0.52 | 0.55 | 0.53 | 0.58 |
| 5-5-5-lenient-0.7 | 0.51 | 0.53 | 0.51 | 0.53 |
| 5-5-5-strict-0.3 | 0.52 | 0.53 | 0.55 | 0.51 |
| 5-5-5-strict-0.7 | 0.51 | 0.54 | 0.50 | 0.56 |
| 7-7-7-strict-0.7 | 0.51 | 0.52 | 0.51 | 0.52 |

Table 1: Polarity Identification: Accuracy values for different parse methods

search field. Those approaches are usually only concerned with the identification and extraction of information without processing it further, except for binary classification within a clearly specified domain.

In the wake of the PASCAL challenge [2, 4], systems have been developed to deal with the relation of sentences to each other. The different approaches include the recognition of false entailment [14], or learning entailment [10]. Others are concerned with relatedness between words and how to measure it [8]. We were not interested in concentrating on one of these areas but rather to develop an all-embracing system incorporating different aspects.

For the domain classification, our best results for 300 paraphrase pairs from the MSR-corpus are, for recall, 81% (with a precision of 38%), and for precision 52% (with a recall of 58%). These values can probably be improved by using more sophisticated heuristics, although there will be a ceiling set by the parser and by the use of language in general. The same meaning can be expressed by various different sentences whose words are not in close relations to each other and therefore hard to detect by current NLP tools. Keeping these facts in mind, the obtained numbers are rather satisfactory and promising for future development.

The rather shallow semantic approach sets a practical limit to the achievable results. This can be inferred by comparing the numbers obtained using manually parsed predicate-argument structures with the numbers obtained by the parsers. It shows that there is space for improvement on the side of the parsers, as well as on the side of the PAS extractor. Combining the results of different parsers could also lead to better results, but a precision of 55% and a recall of 85%, as obtained for the best configuration of the system using manually parsed PASs, shows that it needs more and/or better heuristics to get a really significant improvement.

The polarity identification task was expectedly the hardest one. This is illustrated by the rather poor results we obtained by trying to find different opinions within one domain. Best accuracy values were obtained using MiniPar and were around 58%. This task is very hard for computational systems. But with more elaborated heuristics it is possible to increase these numbers, comparable to the Pascal challenge [2, 4], where systems also started with around 50% accuracy and improved over time.

Testing of the different strategies revealed that the fuzzy processing operators perform in accordance to their assigned tasks. Further evaluation of the results would need some kind of measure to get quantitative,

comparable results. This is beyond the scope of this paper and deferred to future work.

6 Conclusions and Future Work

We developed an artificial believer system that can be applied in different scenarios: (1) companies evaluating product reviews on web sites or blogs, (2) governmental organizations interested in dispositions of people, or (3), as we demonstrated here, assist individuals in news analysis.

Apart from the evaluation described above, tests of the system on actual newspaper articles showed accepted and rejected beliefs that reflect the desired results. Embedding the system within an Internet agent and measuring its effectiveness for a real user will be the next major step.

References

- [1] A. Ballim and Y. Wilks. *Artificial Believers: The Ascription of Belief*. Lawrence Erlbaum Associates, Inc., 1991.
- [2] R. Bar-Haim, I. Dagan, B. Dolan, L. Ferro, D. Giampiccolo, B. Magnini, and I. Szpektor. The Second PASCAL Recognising Textual Entailment Challenge. In *Proc. of the Second PASCAL Challenges Workshop on Recognising Textual Entailment*, 2006.
- [3] S. Bethard, H. Yu, A. Thornton, V. Hatzivassiloglou, and D. Jurafsky. Automatic extraction of opinion propositions and their holders. In Qu et al. [13], pages 20–27.
- [4] I. Dagan, O. Glickman, and B. Magnini. The PASCAL Recognising Textual Entailment Challenge. In *Proc. of the PASCAL Challenges Workshop on Recognising Textual Entailment*, 2005.
- [5] M. Gamon, A. Aue, S. Corston-Oliver, and E. K. Ringger. Pulse: Mining customer opinions from free text. In A. F. Famili, J. N. Kok, J. M. Peña, A. Siebes, and A. J. Feelders, editors, *Advances in Intelligent Data Analysis VI, 6th International Symposium on Intelligent Data Analysis, IDA 2005, Madrid, Spain, September 8-10, 2005, Proceedings*, volume 3646 of *Lecture Notes in Computer Science*, pages 121–132. Springer, 2005.
- [6] S. Hahn, R. Ladner, and M. Ostendorf. Agreement/disagreement classification: Exploiting unlabeled data using contrast classifiers. In Moore et al. [11], pages 53–56.
- [7] S.-M. Kim and E. Hovy. Identifying and analyzing judgment opinions. In Moore et al. [11], pages 200–207.
- [8] B. B. Klebanov. Measuring Semantic Relatedness Using People and WordNet. In Moore et al. [11], pages 13–16.
- [9] R. Krestel, R. Witte, and S. Bergler. Creating a Fuzzy Believer to Model Human Newspaper Readers. In Z. Kobti and D. Wu, editors, *Proc. of the 20th Canadian Conference on Artificial Intelligence (Canadian A.I. 2007)*, LNAI 4509, pages 489–501, Montréal, Québec, Canada, May 28–30 2007. Springer.
- [10] B. MacCartney, T. Grenager, M.-C. de Marneffe, D. Cer, and C. D. Manning. Learning to recognize features of valid textual entailments. In Moore et al. [11], pages 41–48.
- [11] R. C. Moore, J. Billes, J. Chu-Carroll, and M. Sanderson, editors. *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*. Association for Computational Linguistics, New York City, USA, June 2006.
- [12] R. Nairn, C. Condoravdi, and L. Karttunen. Computing relative polarity for textual inference. In *Proceedings of the 5th Workshop on Inference in Computational Semantics (ICoS-5)*, Buxton, UK, April 2006.
- [13] Y. Qu, J. Shanahan, and J. Wiebe, editors. *Exploring Attitude and Affect in Text: Theories and Applications*. Technical Report SS-04-07. AAAI Press, Stanford, CA, USA, March 22–25 2004.
- [14] R. Snow, L. Vanderwende, and A. Menezes. Effectively using syntax for recognizing false entailment. In Moore et al. [11], pages 33–40.
- [15] R. Witte. *Architektur von Fuzzy-Informationssystemen*. BoD, 2002. ISBN 3-8311-4149-5, <http://rene-witte.net>.
- [16] R. Witte. Fuzzy Belief Revision. In *9th Intl. Workshop on Non-Monotonic Reasoning (NMR'02)*, pages 311–320, Toulouse, France, April 19–21 2002.
- [17] R. Witte and S. Bergler. Fuzzy Coreference Resolution for Summarization. In *Proc. of 2003 Intl. Symposium on Reference Resolution and Its Applications to Question Answering and Summarization (ARQAS)*, pages 43–50, Venice, Italy, June 23–24 2003.
- [18] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.