SINAI at TAC-KBP BeST 2017: evaluating the impact of modal verbs in the classification of beliefs

Eladio Blanco López, Fernando Martínez Santiago, and Alberto Marín[1]

University of Jaén, Jaén 23081, Spain,
emblanco@ujaen.es,dofer@ujaen.es,aandreu@ujaen.es
SINAI Group, Computer Science Department, University of Jaén, Campus Las we are usLagunillas s/n, Jaén, SPAIN

Abstract. The aim for SINAI group is twofold in the task of belief classification for English. On one hand, we continue with the development of a hierarchical Bayesian model as a formalism to describe and classify categories, and its comparison with other popular classification algorithms. On the other hand, since the belief-sentiment annotation guidelines highlight the leverage of modal verbs, firstly, we have studied the distribution of this verbs over the corpus and, secondly, those events and relations that are under the scope of a modal verb have been accordingly labelled. We have found that this issue does not increase the performance of the classifiers. Since we have found that the scope of modal verbs and the category of the corresponding facts are correlated, it is necesary a more in-depth analysis of the results and possibly a more fine-grained classification of modal verbs. For example, by making explicit the distinction between deontinc and epistemic senses of such verbs.

1 Introduction

The objective of the present work is to identify and classify the belief of an entity toward another relation, or event[8]. In order to classify every belief we have developed a system following a number of assumptions as design guides:

- It is obvious that it is not possible to encode the same knowledge used by the human annotators. Thus, information to automatically classify every belief is less than the necessary.
- The number of examples for training is rather scarce: 209 posts from discussion forums and 37 documents from newswire.
- The features that should be considered operate are very heterogeneous, with different levels of abstraction: the type of relation or event, the type of entities as arguments,the focus or scope of every verb, and so on.
- The system should make easy the addition of new issues as they become available.

As a consequence of this assumptions we propose a system using Bayesian programming language [4]. More concisely we have used ProBT [1]. It is a formalism, a methodology, an API and an inference engine to solve problems with incomplete and uncertain information.

Part of the aim of this work is to provide insights into the characterization of the type of the belief. Specifically, we have focused on modal auxiliary verbs with the aim to give light about the type of knowledge that an automatic classifier should extract from the text. Moreover, under the scope of modal verbs, it is frequent to find relations and events that are not labeled as beliefs or they are non committed beliefs. Thus, the hypothesis is that if we are able to identify beliefs according the modal verb when it is present, then we will improve the performance of the system for those relations and events different from committed beliefs.

The rest of the paper is organized as follows: section 2 describes the document test best from the point of view of modal verbs distribution over the beliefs. Following this section we propose our system based on HBN and a preliminar evaluation is accomplished. Finally, some conclusions are exposed.

2 Description of the document testbed

The experiments have been performed using the collections available in English in BeSt TAC-KBP2017(see Table 2) [7]. These collections are made up of a set of documents whose source are discussion forums (DF collection) and news reported by different agencies and media (NW collection). In addition, along with these documents the Organization has provided a set of XML documents which have been manually marked with entities, events, relations, entity, event and relation mentions, entity type and subtype, event type and subtype, core (trigger) of each relation or event, arguments of these, with their corresponding types and semantic roles. Every mention (relation or event), has also been tagged with its polarity (positive, negative, neutral), and with an attribute that indicates whether or not the intention of the source of such mention is sarcastic.

The distribution of the four possible values for each belief, CB, NCB, ROB or NA on the training corpus, is displayed in the Table 2. This table points out a first peculiarity of the corpus, and is that is far to be treated in a uniform distribution, 77% of relations or events whose source is a forum are tagged with a belief of the Committed Belief (CB) type. In case that the relation or event is part of a news (NW collection), this percentage rises to 91%, possibly due to the nature more impartial and "aseptic" of these documents.

2.1 Study of the correlation between the category of a belief and the information labeled in the corpus (gold standard)

In addition to the judgments of relevance, the documents are manually tagged with the following information (gold ERE documents):

- Type of relation or event.

Table 1. Set of documents used in experimentation (DF=forums, NW=news)

	EN		ES	
Training				
Source	DF	NW	DF	NW
Docs	192	33	121	33
Beliefs	6208	2403	8123	2991
Test(2016)				
Source	DF	NW	DF	NW
Docs	84	80	74	71
Beliefs	2644	2350	1921	1840
Test(2017)				
Source	DF	NW	DF	NW
Docs	84	83	83	83
Beliefs	7600	12430	6549	8979

Table 2. Frequency of each type of belief in the case of the English training dataset

	CB	NCB	ROB	NA	Total
DF	4754	116	42	1296	6208
NW	1855	21	16	151	2403

- Type and subtype of the entities involved in the relation or event.
- Semantic roles involves in every relation or event.
- Polarity: If the position of the source is positive or negative to the given fact.
- Sarcasm: If it is an expression that has interpreted sarcastically.

This information is available for each belief, with the exception of the last two (polarity and sarcasm), which are not available to those relations or events marked as NA. In any case, it is not clear that all or at least part of this additional information is related to the type of belief. To try to shed light on this issue, it has been done a hypothesis test based on the Pearson χ^2 test, considering the hypothesis of independence between the type of belief and each label available in the corpus. It is important to recall that this statistic is not a sufficient condition, but needed to make a relationship cause and effect between two variables. In addition, to perform with the approximate normality assumption, it has been grouped some categories to reach an estimated frequency equal to or greater than 5 for each co-occurrence for each pair of variables. For this reason, some types of entities, relations and semantic roles had to be grouped under a single label. For the same reason, and given the low frequency of NCB and ROB beliefs (show Table 2) it had to be grouped NCB, ROB and NA in a single category. Finally, we could not apply this statistic with guarantees on the polarity and the sarcasm, since both labels are not available for NCB or ROB, and consequently it is not possible to perform with the requirement of minimum estimated frequency in these cases. The results are displayed in the Table 3: in all cases the χ^2 test statistic calculated by using the training dataset is higher than the theoretical value. As a consequence, we must reject the null hypothesis of independence between each of the three variables considered and the type of belief. For this reason it is our view that these variables are candidates to be included in any model that explains the type of belief.

Table 3. Correlation between some manual labels and the type of belief. K=number of categories, $\chi^2_{K,1}exp$ is the experimental value and $\chi^2_{K,1}$ is the theoretical value with a significance level of 5%

Variable	$K-1$	$\chi^2_{k,1}$	$\chi^2_{k,1}exp$
Relation or event type	13	22.4	487.3
Entity type	94	117.6	325.6
Semantic roles	142	170.81	614.7

2.2 Using the syntactical scope to identify and classify facts under the focus of modal verbs

According to the annotation guidelines of the BeST corpus, it is expected to find relations and events that are not labeled as beliefs or they are non committed beliefs when they happens under the focus of a modal verb. For example, there is a rule that asses that "might" modal verb is used to express non-committed beliefs in the majority of the cases. Of course, it is true if the relation or event is located in the same context of such a verb. In order to be more precise in the definition of context, it is useful to distinguish between focus and scope such as, i.e., negation does[3,5]. Negation is the task of identifying which words are negated because of the presence of a negation construction in the sentence, such as "no" , "not" or "never". The scope of negation includes all the words affected by negation, while the focus corresponds to the part of the scope that is most directly negated. In this way, the focus corresponds with a fragment of the scope, which is most relevant and explicitly negated. In the same way, "might" is able to modify the interpretation as belief of a given fact under its focus or scope. Two examples are shown below.

- *John **might** be traveling/NCB to Turkey*
- *A store like this might **survive**/NCB in **my country**/CB*

In this examples, both "traveling" and "survive" are non-committed beliefs since the author is expressing a degree of certainty in the proposition. In contrast, "my country" is supposed to be a well-known fact and, consequently, it is a committed belief because in spite of "my country" is part of the scope of the modal verb, it is out of the focus.

In spite of the focus defines context more accurately than the scope for our topic of interest, we just identify the scope since the focus is hard to identify with precision provided that it is deeply rooted with semantic and pragmatic issues. In this work, the scope always corresponds to a syntactic component, that is a phrase, a clause or a sentence. We have implemented a quite naive procedure to decide when a relation or event $fact$ is under the context of a modal verb. Following the algorithm depicted below we have found the distribution of modal verbs over belief types depicted in tables 4 and 5.

Step 1: Extract the sentence S where $fact$ happens
Step 2: Obtain the dependency tree DT of S
Step 3: Starting in the node where the trigger word of $fact$ happens go upward DT until a modal verb or the root of DT is reached
Step 4: If a modal verb is reached in the previous step, then $fact$ is considered under the scope of such a modal verb

In order to extract the sentence where the relation or event happens, it is defined as the minimal text segment that meets the following three requisites: (i) it includes the arguments of the fact, (ii) the starting point is a dot (not initial) or the beginning of the paragraph and (iii) the end point is a dot (not initial) or the end of the paragraph. The dependency tree is obtained by parsing S whith FreeLing[6].

Table 4. beliefs distribution under the scope of modal verbs (English,discussion forums training dataset)

	CB	**NCB**	**ROB**	**NA**	TOTAL
might	7	1	0	7	15
could	47	1	0	23	61
need to	4	0	0	5	9
couldn't	2	0	0	1	3
have to	19	1	0	6	20
ought to	0	0	0	0	0
may	12	3	0	9	24
can	39	2	0	32	73
wouldn't	9	0	0	1	10
would	103	7	0	71	181
should	23	0	0	19	42
can't	12	1	0	4	17
shouldn't	4	0	0	0	4
TOTAL	281	16	0	178	459
no modal	4486	101	42	1129	5758

If we compare the distribution of the relations and events over the 4 types of beliefs (including NA as a kind of belief) we observe that it is more likely to find other beliefs different from CB when the modal verb happens. For example, by comparing Table 2 and Table 4 we see that the probability of NA for English, discussion forums, training corpus is $1296/6208 = 0.208$. If we consider the cases where the modal verb is close to the relation or event, then this probability raises up to $178/459 = 0.38$. Consequently, we think that modal verbs should be

Table 5. beliefs distribution under the scope of modal verbs (English,news, training dataset)

	CB	**NCB**	**ROB**	**NA**
might	0	0	0	0
could	1	0	0	1
need to	0	0	0	0
couldn't	0	0	0	0
have to	0	0	0	0
ought to	0	0	0	0
may	0	0	0	0
can	5	0	0	0
wouldn't	0	0	0	3
would	36	0	0	0
should	1	1	0	0
can't	0	0	0	0
shouldn't	0	0	0	0
TOTAL	43	1	0	4
no modal	1812	20	16	144

accordingly managed, at least when it is a subjective text since this verbs are quite scarce when the source of the text are news (see Table 5).

3 Description of the system and results

In spite of empathizing the classification of the belief in this work, since it is necessary the identification of the source for every belief, we follow the same naive approach as last year for source identification: the source of the belief is the same that the source of the document or post where such a belief happens. Consequently, if the source of the document is unknown or anonymous, then every belief will be equally unknown or anonymous.

Table 6. Gold ere files, English collection (DF=forums, NW=news). Issues={k1,k2,k3}

	Micro average		Macro average	
Classifier	DF	NW	DF	NW
HBN	P: .67 R: .64 F: .65	P: .73 R: .50 F: .59	P: .67 R: .64 F: .65	P: .72 R: .53 F: .61
KNeighbors	P: .62 R: .72 F: .66	P: .73 R: .51 F: .60	P: .62 R: .73 F: .67	P: .71 R: .54 F: .61
MultinomialNB	P: .53 R: .63 F: .58	P: .66 R: .48 F: .56	P: .53 R: .64 F: .58	P: .65 R: .51 F: .58
RandomForest	P: .62 R: .72 F: .66	P: .73 R: .50 F: .59	P: .62 R: .73 F: .67	P: .72 R: .53 F: .61
Perceptron	P: .63 R: .72 F: .67	P: .72 R: .50 F: .59	P: .63 R: .74 F: .68	P: .71 R: .53 F: .61
MLP	P: .64 R: .73 F: .69	P: .73 R: .50 F: .60	P: .65 R: .75 F: .69	P: .72 R: .54 F: .61
SVM	P: .64 R: .73 F: .69	P: .73 R: .50 F: .60	P: .65 R: .75 F: .69	P: .72 R: .54 F: .61
LinearSVC	P: .63 R: .73 F: .68	P: .72 R: .50 F: .59	P: .63 R: .75 F: .68	P: .71 R: .53 F: .61

Table 7. Gold ere files, English collection (DF=forums, NW=news). Issues={k1,k2,k3,k4} (K4: True if the relation or event is under the scope of a modal verb)

	Micro average		Macro average	
Classifier	DF	NW	DF	NW
HBN	P: .69 R: .65 F: .67	P: .79 R: .54 F: .67	P: .70 R: .65 F: .68	P: .77 R: .57 F: ..67
KNeighbors	P: .62 R: .72 F: .66	P: .73 R: .51 F: .60	P: .62 R: .73 F: .67	P: .71 R: .54 F: .61
MultinomialNB	P: .53 R: .63 F: .58	P: .67 R: .48 F: .56	P: .53 R: .64 F: .58	P: .65 R: .51 F: .57
RandomForest	P: .62 R: .72 F: .66	P: .73 R: .50 F: .60	P: .62 R: .73 F: .67	P: .72 R: .54 F: .62
Perceptron	P: .63 R: .73 F: .68	P: .73 R: .50 F: .60	P: .63 R: .74 F: .68	P: .72 R: .54 F: .61
MLP	P: .65 R: .74 F: .69	P: .74 R: .51 F: .60	P: .65 R: .76 F: .70	P: .72 R: .54 F: .62
SVM	P: .65 R: .74 F: .70	P: .74 R: .51 F: .60	P: .65 R: .76 F: .70	P: .72 R: .54 F: .62
LinearSVC	P: .62 R: .72 F: .67	P: .73 R: .50 F: .60	P: .63 R: .74 F: .68	P: .72 R: .53 F: .61

Regarding the classification of the belief, we have evaluated the impact of modal verbs and we have tested a number of different classifiers together the hierarchical Bayesian model. More concisely, the issues are the event/relation sub-type (K1), the type of the entities(K2), the arguments of this relations and events(K3) and if the relation or event is under the scope of a modal verb such as it is depicted in 2.2 (K4). The tested classifiers are Hierarchical Bayesian Model (HBM) such is depicted in [2], Naive Bayes, Random Forest, multi-layer and lineal Perceptron, lineal Support Vector Machine (SVM) and Support Vector Machine (SVM). We find a discrete improvement (5%) when HBN is used in spite of using a quite flat topology such as is depicted in [2]. On the other hand, we have not found significant differences when modal verbs are in place.

4 Conclusions

In this paper we describe the impact of modal verbs in the Belief and Sentiment Evaluation Task, more concisely in the classification of a given relation or event according to the type of belief that it represents. Thus, we mark those relations or events that are under the scope of a modal verb by using the dependency tree of the sentence where the fact happens. Even though, there is a clear correlation between the scope of a modal verb and belief type, it does not improve the performance of some of the most popular classifiers. We think that a more precise distinction should be done by differentiating epistemic and deontic usages of the modal verb, since according to the annotation guidelines of the BeST corpus both senses frequently determine the type of belief in a different way. In addition, we want to check if the results would be improved if we implement a more sophisticated algorithm to enable the detection of the focus of the verb instead of the scope.

References

1. Bessiere, P., Mazer, E., Ahuactzin, J.M., Mekhnacha, K.: Bayesian programming. CRC Press (2013)
2. Gutiérrez, Y., Jiménez-Zafra, S.M., Moreno, I., Martín-Valdivia, M.T., Tomás, D., Montejo-Ráez, A., Montoyo, A., Ureña-López, L.A., Martínez-Barco, P., Martínez-Santiago, F., Blanco, E.: Redes at tac knowledge base population 2016: Edl and best tracks
3. Huddleston, R.D., Pullum, G.K.: The Cambridge Grammar of the English Language. Cambridge University Press (April 2002)
4. Jaynes, E.T.: Probability theory: The logic of science. Cambridge university press (2003)
5. Jiménez-Zafra, S.M., Martin, M., Lopez, L.A.U., Marti, T., Taulé, M.: Problematic cases in the annotation of negation in spanish. In: Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM). pp. 42–48 (2016)
6. Padró, L., Stanilovsky, E.: Freeling 3.0: Towards wider multilinguality. In: LREC2012 (2012)
7. Prabhakaran, V., By, T., Hirschberg, J., Rambow, O., Shaikh, S., Strzalkowski, T., Tracey, J., Arrigo, M., Basu, R., Clark, M., et al.: A new dataset and evaluation for belief/factuality. In: Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics. pp. 82–91 (2015)
8. Rambow, O., Yu, T., Radeva, A., Hamidian, S., Fabbri, A., Ghosh, D., Peng, T., Diab, M., McKeown, K., Muresan, S.: The columbia-gwu system at the 2016 tac kbp best evaluation (2016)