A Multi- versus a Single-classifier Approach for the Identification of Modality in the Portuguese Language

João Sequeira[1], **Teresa Gonçalves**[1], **Paulo Quaresma**[1], **Amália Mendes**[2], **Iris Hendrickx**[23]
[1]Department of Informatics, University of Évora, Portugal
[2]Center of Linguistics, University of Lisbon, Portugal
[3]Centre for Language Studies, Radboud University Nijmegen, The Netherlands
d11594@alunos.uevora.pt, tcg@uevora.pt, pq@uevora.pt,
amaliamendes@letras.ulisboa.pt, iris@i-hx.nl

Abstract

This work presents a comparative study between two different approaches to build an automatic classification system for Modality values in the Portuguese language. One approach uses a single multi-class classifier with the full dataset that includes eleven modal verbs; the other builds different classifiers, one for each verb. The performance is measured using precision, recall and F_1. Due to the unbalanced nature of the dataset a weighted average approach was calculated for each metric. We use support vector machines as our classifier and experimented with various SVM kernels to find the optimal classifier for the task at hand. We experimented with several different types of feature attributes representing parse tree information and compare these complex feature representation against a simple bag-of-words feature representation as baseline. The best obtained F_1 values are above 0.60 and from the results it is possible to conclude that there is no significant difference between both approaches.

Keywords: Natural Language Processing, Modality, Feature Selection, Support Vector Machines

1. Introduction

In the last years there was a great development in fields related to Machine Learning in the pursuit of forms of Artificial Intelligence (AI), which has become a major research trend of both academic and companies, like Google and Microsoft. The fields are diverse, ranging from financial fraud identification, image recognition and even systems that can rewrite their own code or write other programs (Caughill, 2017; Gershgorn, 2017; Galeon, 2017). Natural Language Processing (NLP) is a related research field that includes tasks aiming at understanding texts, information extraction and text classification.

Presently, one of the most active sub-field focuses on sentiment analysis and opinion mining (Pang and Lee, 2008). This includes tasks such as the (automatic) distinction between the factual and non-factual nature of events and the detection of the subjective perspective underlying texts. Modality is one indicator of subjectivity and factuality and it is usually defined as the expression of the speaker's opinion and attitude towards the proposition (Palmer, 1986). Traditionally, it covers epistemic modality, which is related to the degree of commitment of the speaker to the truth of the proposition (whether the event is perceived as possible, probable or certain), but also deontic modality (obligation or permission), capacity and volition (Sequeira et al., 2016). Information about the modality of a text is crucial for the above mentioned trends on automatic fact finding and information extraction.

This work extends the experiments done previously (Sequeira et al., 2016) in the pursuit of creating a semi-automatic modality tagging system for the Portuguese language from a manually annotated corpus that uses the modality scheme described by Hendrickx et al. (2012) and Mendes et al. (2016).

In this study we focus on machine learning optimization and feature selection for modality detection and labeling. We compare two different system architectures, namely one classifier trained on all modal verbs and one architecture where we train a classifier for each modal verb separately. Such 'word expert' approach is known to work well in word sense disambiguation (a closely related task) (Hoste et al., 2002). We also investigate whether the complex feature representation based on parse information as applied in our previous work (Sequeira et al., 2016) is indeed more informative than a simple bag-of-word feature representation.

The paper is structured as follows: Section 2 introduces related work done in the field of modality, Section 3 presents the developed system describing the experimental setup, corpus and dataset information, attributes extracted and results obtained and Section 4 discusses some conclusions and future work aiming at improving the system.

2. Related work

As Portuguese is one of the 10 most spoken languages in the world, with more than 260 millions of speakers (da Língua Portuguesa, 2015), the development of natural language processing tools and linguistically annotated resources for Portuguese are crucial to keep up with the current information society (Branco et al., 2012).

However, most studies related to modality still focus on the English language, and besides our own work, not much tools have been developed for Portuguese. Baker et al. (2010), Matsuyoshi et al. (2010), Nirenburg and McShane (2008) and Sauri et al. (2006) present modality annotation schemes for the English language; for Portuguese we can identify the work from Hendrickx et al. (2012) for written European Portuguese, Ávila and Melo (2013) for spoken Brazilian Portuguese, and the updated proposal of both teams in Ávila et al. (2015).

Thompson et al. (2008) addressed the identification of ex-

pressions linked to modality in biomedical texts using three dimensions: the kind of knowledge, level of certainty and point of view. Their approach uses a list of words and phrases with modal characteristics specific for the biomedical domain. Baker et al. (2010) tested two rule-based modality taggers that identify both the modal trigger (word or word list where modality is expressed, usually by the use of modal verbs) and its target (the event, state, or relation over which the modality has scope) and achieved results of 86% precision for a standard LDC data set.
Ruppenhofer and Rehbein (2012) developed a modal verb annotation scheme for news articles written in English. The system uses a classifier of maximum entropy (Ratnaparkhi, 1996) to identify the verbs *can*, *may/might*, *must* and *should*. The attributes used are divided into three categories: (i) target/verb; (ii) context; (iii) path. They used different combinations of attributes with different context sizes and the results were compared to those of a baseline system always assigning the most common value to each verb. The best result was achieved for the verb *must* with an accuracy of value 93,50%, followed by the verb *shall/should* with 91,61%, *may/might* with 85,71% and finally *can* with 68,70%.
In what concerns the Portuguese language, Sequeira et al. (2016) is a earlier and less developed version of the work presented here. The goal was to select the best set of attributes for creating automatic taggers and compare the results with a bag-of-words (bow) approach. The paper covers the creation of the corpus (composed by eleven verbs), the use of a parser to extract syntactic and semantic information from the sentences and a machine learning approach to identify modality values.

3. Experiments

Like in the previous experiments, eleven Portuguese modal verbs (that we call triggers) are studied. They are:"arriscar" (chance/risk/dare), "aspirar" (aspire), "conseguir" (manage to/succeed in/be able to), "considerar" (consider/regard), "dever" (shall/might),"esperar" (wait/expect), "necessitar" (need/require), "permitir" (allow/permit), "poder" (may/can), "precisar" (need) and "saber" (know).
These verbs are polysemous and are deliberately chosen as our focus verbs because they can express more than one type of modality. For example, the verb "poder" can be *Epistemic* stating that something is possible, *Deontic* denoting permission or may express an *Internal Capacity* when expressing the fact that someone is able to do something.
In Sequeira et al. (2016) several combinations of classes of attributes, namely trigger, path and context were tested and the best one was selected (path+context). An attribute ranker (using information gain) singled out the following attributes as the most informative for *path*:

- presence of an Accusative node between the root and the verb node
- no explicit subject in the left brother node
- the left brother node receives the semantic role Theme
- presence of an infinitive clause in the path: either from the root to the verb or as the right brother node
- the left brother node is a Dative object with the function Beneficiary

For *context*, some of the most important attributes occur in the left tree:

- the lemma *lei* 'law' occurs in the left context
- the dative clitic *lhe* 'to him/her' occurs in the left context

These attributes point to certain properties of the trigger and the context that lead to one modal interpretation. They may be somehow unexpected, as the case of the attribute "Dative brother node in left context". The combination of attributes for path listed above, namely the presence of an Accusative node which is of the type infinitival clause, favours an epistemic reading of the verb *permitir*, as illustrated in (1). Moreover, many of the examples of epistemic possibility reading with *permitir* are associated to constructions where the left brother node is a Dative object, another attribute listed for path (example (2)).

(1) Mas estes primeiros dias já *permitem tirar conclusões*.
'But these first days already make it possible to draw conclusions.'

(2) Agora, embora não seja capaz de pintar porque não tenho técnica para o fazer, descobri que o computador *me permite* transformar as minhas imagens de tal maneira que ficam a parecer autênticas pinturas.
'Now, although I'm not capable of painting because I don't have the technique to do so, I discovered that the computer allows me to transform my images in such a way that they end up looking like authentic paintings.'

In what concern the class of attributes for context, a deontic reading of the verb *permitir* is strongly related to the presence of the lemma *lei* 'law' in subject position (4 contexts out of 5), as illustrated in (3).

(3) E acrescenta que não existe nenhuma lei que permita à Portugal Telecom cortar o serviço telefónico por os utentes não pagarem, por exemplo, as chamadas de valor acrescentado, tipo telefonemas eróticos, etc.
'And [he/she] adds that there is no law that allows Portugal Telecom to cut the phone service when users don't pay, for instance, value added calls, such as erotic phone calls.'

We keep the same set of attributes in this experiment. Besides a baseline using a bag-of-words approach, this work uses that attribute setting to compare different classification experiments, namely:

- `exp.A`: to build a specific classifier for each verb, aiming at detecting the specific modality types (setting used in (Sequeira et al., 2016))
- `exp.B`: to build a single classifier with the full corpus (all verbs and all types of modality)

- `exp.C`: the same as `exp.B` but with an extra attribute – the lemma of the trigger

The advantage of `exp.A` is that such a classifier has to learn a smaller set of possible modal values (2-4 instead of 11) but has less examples to train on. `exp.B` and `exp.C` on the other hand has to learn to distinguish 11 different modal values but has more training examples to learn from. Note that `exp.B` is only included to measure the effect of knowledge about the trigger. While the modal values are shared among different verbs, we expect knowledge about the trigger still to be crucial in obtaining good results.

The Sequential Minimal Optimization (SMO) (Platt, 1999) algorithm, an improved version of the Support Vector Machine (SVM) (Vapnik, 1998), is used to build the automatic classification system and the performance is measured using precision, recall and F_1 measures.

3.1. Corpus and dataset

The dataset is composed by 936 sentences (examples) containing these modal verbs. In total eleven different modality values are expressed by these modal verbs, but each verb itself has between 2 -4 possible modal meanings. Table 1 characterizes each verb. As we can see, the number of sentences for each verb varies from 51 for "necessitar" to 254 for "poder". The number of sentences for each modal value also varies (from 11 examples for *evaluation* to 299 examples for *epistemic possibility*).

3.2. Feature extraction

The extracted attributes are the same as the ones reported in (Sequeira et al., 2016): a set related with the (i) *trigger* (modal verb information), another related with (ii) *context* (with a window of size five), and one related with the (iii) *path* (syntactic and morphological information extracted from the parse tree trigger). The selection of trigger, context and path is inspired by the work of (Ruppenhofer and Rehbein, 2012) and our goal was to be able to compare our results with their findings. The attributes are based on the syntactic and morphological analysis trees generated by the PALAVRAS parser (Bick, 1999; Bick, 2000). Table 2 summarizes the attributes extracted: for the *trigger* set, information from the trigger itself and from the ancestors; for the *path* set, information about the trigger's siblings and the path from the trigger to root; for the *context* set, information about the words to the left and right of the trigger.

3.3. Experimental setup

Using a SVM model and a 5-fold stratified cross-validation procedure, precision, recall and F_1 weighted averages were calculated for the three different classification experiments and compared with a bag-of-words approach as baseline. Different kernels with default parameters were tested, namely the polynomial kernel with degrees 1 (linear kernel) , 2 and 3 and the radial basis function.

Appropriate statistical tests with 95% of significance were applied to analyse the differences between results. These machine learning experiments were conducted using Weka framework (Hall et al., 2009).

verb	modality type	# example	
arriscar	effort	20	46
	epistemic belief	1	
	epistemic possibility	25	
aspirar	epistemic belief	18	52
	volition	34	
conseguir	participant-internal capacity	42	87
	epistemic possibility	4	
	success	41	
considerar	epistemic belief	15	26
	evaluation	11	
dever	epistemic belief	2	124
	deontic permission	3	
	deontic obligation	78	
	epistemic possibility	41	
esperar	epistemic belief	30	57
	volition	26	
	epistemic possibility	1	
necessitar	deontic obligation	8	51
	participant-internal necessity	41	
	participant-internal capacity	2	
permitir	deontic permission	19	80
	epistemic possibility	61	
poder	deontic permission	46	254
	deontic obligation	1	
	participant-internal capacity	40	
	epistemic possibility	167	
precisar	deontic obligation	10	56
	participant-internal necessity	46	
saber	participant-internal capacity	10	103
	epistemic knowledge	93	

Table 1: Corpus characterization: number of sentences per modal value for each verb.

3.4. Results

Table 3 present the weighted average precision for the described experiments.

The best weighted precision value (0.691) was obtained using the bag-of-words approach with the verb lemma as additional attributes using a single linear classifier (`exp.C`) but there's no significant difference with the experiment using *path+context* attributes with 11 classifiers, one for each verb (0.689) with a polykernel of degree 2 (`exp.A`). The worst result (0.102) was obtained using a bag-of-words representation with a RBF kernel and a single classifier without verb lemma information.

As expected knowledge about the trigger is very informative for the classifier and the results in `exp.B` are the lowest of the three options.

Comparing the kernel functions one can conclude that RBF kernel has the worst precision values; on the other hand the linear kernel seems to be the best when using a bag-of-words approach (baseline), while the polynomial kernel with degree 2 is better when using *path+context* attributes. Looking at the different classification settings, it seems that using a single classifier for the 11 modal values does not improve precision when comparing to a setting using a dif-

attribute set	source	attributes
trigger	trigger	POS function role morphological semantic
	ancestors	POS function
path	siblings	POS function role morphological semantic
	trigger to root	POS function
context	left/right trigger	POS word lemma

Table 2: Attributes extracted from trigger, path and context

ferent classifier for each verb (but also does not seem to hurt the classification system).

attributes	kernel	`exp.A`	`exp.B`	`exp.C`
path+ context	poly, d1	.678	.420	.659
	poly, d2	**.689**	.447	.627
	poly, d3	.678	.447	.582
	rbf	.615	.285	.544
baseline	poly, d1	.681	.355	**.691**
	poly, d2	.652	.261	.553
	poly, d3	.612	.266	.320
	rbf	.605	.102	.424

Table 3: Weighted average precision values.

Table 4 present the weighted average recall for the described experiments. The best value (0.708) was also obtained using the bag-of-words approach with the verb lemma as additional attributes using a single linear classifier (`exp.C`) but, once again, there's no significant difference with the experiment using *path+context* attributes with 11 classifiers (0.698) with a polykernel of degree 2 (`exp.A`). The worst result (0.121) was obtained using a bag-of-words representation with a polynomial kernel of degree 3 and a single classifier without verb lemma information.
For the recall measure it seems that the kernel function does not influence the performance of the classifier as it was the case for precision. Nonetheless, it seems that for the bag-of-words the linear kernel is the best while the polynomial kernel with degree 2 is better when using *path+context* attributes. Looking at the different classification settings, it seems that using a single classifier for the 11 modal values (`exp.C`) can hurt recall when comparing to a setting using a different classifier for each verb (`exp.A`), as values of 0.698 vs .0652 were obtained respectively.
Finally, Table 5 present the weighted average F_1 values

attributes	kernel	`exp.A`	`exp.B`	`exp.C`
path + context	poly, d1	.678	.436	.675
	poly, d2	**.698**	.475	.652
	poly, d3	.693	.453	.584
	rbf	.673	.408	.530
baseline	poly, d1	.689	.385	**.708**
	poly, d2	.667	.314	.578
	poly, d3	.640	.121	.353
	rbf	.668	.319	.456

Table 4: Weighted average recall values.

for the described experiments. As expected, the best value (0.683) was obtained using the bag-of-words approach with the verb lemma as additional attribute using a single linear classifier (`exp.C`) even if there's no significant difference with the experiment using *path+context* attributes with 11 classifiers (0.678) with a polykernel of degree 2 (`exp.A`). The worst result (0.129) was obtained using a bag-of-words representation with a polynomial kernel of degree 3 and a single classifier without verb lemma information.
Using one classifier for all modal values with additional verb lemma attribute (`exp.C`) or a specific classifier for each verb similar F_1 performance values are achieved when using a linear kernel (both for the baseline and the *path+context* set of attributes). While the value maintains stable for kernels of higher degrees for the `exp.A` setting , it consistently decreases for `exp.C` one.

attributes	kernel	`exp.A`	`exp.B`	`exp.C`
path + context	poly, d1	.673	.426	.664
	poly, d2	**.678**	.433	.620
	poly, d3	.664	.386	536
	rbf	.628	.279	.425
baseline	poly, d1	.658	.327	**.683**
	poly, d2	.627	.223	.536
	poly, d3	.616	.129	.275
	rbf	.627	.155	.333

Table 5: Weighted average F_1 values of the outcomes with different settings of the polykernel with degree 1 (d1),2 (d2) and 3(d3), and the RBF kernel.

4. Conclusions and Future work

This work extends previous experiments that try to identify the best automatic approach to tag modality in the Portuguese language.
Eleven modal verbs were used and morphological, syntactic and some semantic attributes were extracted from the 936 sentences using the PALAVRAS parser. Several experiments were conducted using two different sets of attributes: a bag-of-words representation was used as baseline and the second set includes several attributes taken from the syntax parse tree path and modal verb context. Two different kernel functions were tested (polynomial with degree ranging from 1 to 3 and RBF kernels). Three different classification approaches were set up: (a) a set of

11 classifiers, one for each verb (using the corresponding subset of sentences); (b) a single multi-class classifier (11 classes, one for each modal value); (c) the same as (b) but adding a valuable attribute: the trigger verb lemma.

Comparing the performance of the different systems, one can conclude that adding lemma information improves the performance when using a single multi-class classifier, but there is no significant difference to the multi-classifier approach. With the individual classifiers all values (for all three measures) were above 0.60, independently of the set of attributes and kernels used. This is not true when using a single classifier for all the existing modal values.

The corpus is relatively small (specially if we take into account the number of possible different classes) and is not balanced. This certainly influences the performance of the system.

As future work, we intend to expand the corpus trying to get a more balanced version of examples. Next steps for building a complete automatic modality tagging system are to identify the source of the modality and the target linked to the modality value.

5. Bibliographical References

Baker, K., Bloodgood, M., Dorr, B., Filardo, N. W., Levin, L., and Piatko, C. (2010). A modality lexicon and its use in automatic tagging. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

Bick, E. (1999). *The parsing system PALAVRAS*. Aarhus University Press.

Bick, E. (2000). *The Parsing System "Palavras". Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. University of Arhus, Århus.

Branco, A., Mendes, A., Pereira, S., Henriques, P., Pellegrini, T., Meinedo, H., Trancoso, I., Quaresma, P., Strube de Lima, V. L., and Bacelar, F. (2012). *A língua portuguesa na era digital – The Portuguese Language in the Digital Age*. META-NET White Paper Series. Georg Rehm and Hans Uszkoreit (Series Editors). Springer. Available online at `http://www.meta-net.eu/whitepapers`.

Caughill, P. (2017). Google's AI is Learning to Make Other AI.

da Língua Portuguesa, O. (2015). As 10 línguas mais faladas no mundo. Technical report.

Galeon, D. (2017). New AI Can Write and Rewrite Its Own Code to Increase Its Intelligence.

Gershgorn, D. (2017). Microsoft's AI is learning to write code by itself, not steal it.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November.

Hendrickx, I., Mendes, A., and Mencarelli, S. (2012). Modality in text: a proposal for corpus annotation. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).

Hoste, V., Hendrickx, I., Daelemans, W., and van den Bosch, A. (2002). Parameter optimization for machine-learning of word sense disambiguation. *Natural Language Engineering*, 8(4):311–325.

Matsuyoshi, S., Eguchi, M., Sao, C., Murakami, K., Inui, K., and Matsumoto, Y. (2010). Annotating event mentions in text with modality, focus, and source information. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

Mendes, A., Hendrickx, I., Gonçalves, T., Quaresma, P., and Sequeira, J. (2016). Modality annotation for portuguese: from manual annotation to automatic labeling. *LiLT - Linguistic Issues in Language Technology*, volume 14.

Nirenburg, S. and McShane, M. (2008). Annotating modality. Technical report, University of Maryland, Baltimore County, USA, March.

Palmer, F. R. (1986). *Mood and Modality*. Cambridge textbooks in linguistics. Cambridge University Press.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.

Platt, J. C. (1999). Advances in kernel methods. chapter Fast Training of Support Vector Machines Using Sequential Minimal Optimization, pages 185–208. MIT Press, Cambridge, MA, USA.

Ratnaparkhi, A. (1996). A maximum entropy model part-of-speech tagger. In *EMNLP'96 – Empirical Methods in Natural Language Processing Conference*, pages 133–141.

Ruppenhofer, J. and Rehbein, I. (2012). Yes we can!? annotating english modal verbs. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).

Sauri, R., Verhagen, M., and Pustejovsky, J. (2006). Annotating and recognizing event modality in text. In *FLAIRS Conference*, pages 333–339.

Sequeira, J., Gonçalves, T., Quaresma, P., Mendes, A., and Hendrickx, I. (2016). Using syntactic and semantic features for classifying modal values in the portuguese language. In *CICLing 2016, 17th International Conference on Intelligent Text Processing and Computational Linguistics*.

Thompson, P., Venturi, G., McNaught, J., Montemagni, S., and Ananiadou, S. (2008). Categorising modality in biomedical texts. In *Proceedings of the LREC 2008 Workshop on Building and Evaluating Resources for Biomedical Text Mining*, pages 27–34, Marrakech, Marrocos.

Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley-Interscience.

Ávila, L. and Melo, H. (2013). Challenges in modality annotation in a brazilian portuguese spontaneous speech corpus. In *Proceedings of IWCS 2013 WAMM Workshop on the Annotation of Modal Meaning in Natural Language*, Postam, Germany. Association for Computational Linguistics.

Ávila, L., Mendes, A., and Hendrickx, I. (2015). Towards a unified approach to modality annotation in portuguese. In *Proceedings of the IWCS Workshop on Models for Modality Annotation, MOMA 2015*, pages 1–8.