SICS at NTCIR-7 MOAT: Constructions represented in parallel with lexical items

Jussi Karlgren Gunnar Eriksson Oscar Täckström
Swedish Institute of Computer Science
Box 1263
SE-164 29 Kista
Sweden
{jussi,gunnar,oscar}@sics.se

Abstract

This paper describes experiments to find attitudinal expressions in written English text. The experiments are based on an analysis of text with respect to not only the vocabulary of content terms present in it (which most other approaches use as a basis for analysis) but also on structural features of the text as represented by presence of function words (in other approaches often removed by stop lists) and by presence of constructional features (typically disregarded by most other analyses). In our analysis, following a constructional grammatical framework, structural features are treated similarly to vocabulary features.

Our results give us reason to conclude – provisionally, until more empirical verification experiments can be performed – that:

- *Linguistic structural information* does *help in establishing whether a sentence is opinionated or not; whereas*
- *Linguistic information of this specific type* does not *help in distinguishing sentences of differing polarity.*

Keywords: *NTCIR, Constructional features, linguistic opinion analysis*

1 Representing attitude in text

Our approach takes as its starting point the observation that lexical resources always are noisy, out of date, and most often suffer simultaneously from being both too specific and too general. Not only are lexical resources inherently somewhat unreliable or costly to maintain, but they do not cover all the possibilites of expression afforded by human linguistic behaviour: we believe that attitudinal expression in text is not solely a lexical issue. We have previously used a resource-thrifty approach for valence or polarity annotation of news headlines reported in the Semeval proceedings [5]. In that experiment we used a minimal set of hand-picked target terms to identify opinionated expressions, expanding the coverage through a distributional word space model trained on general newsprint. For our present experiments reported here no attitudinal lexical resources were used — only general purpose linguistic analysis was employed to establish the constructions used in the further processes.

It has previously been suggested that attitude in text is carried by dependencies among words, rather than by keywords or high-frequency words[1]. While in the cited work the authors hoped to learn such patterns implicitly, we take a different route and explicitly incorporate constructions in our representation, by making use of external general linguistic knowledge resources.

1.1 Data and linguistic processing

We use the NTCIR-6 topics and assessments as training data. The 5854 assessed sentences, of which 1391 were judged to be opinionated (NEU 651, NEG 535, POS 204), are used as a target set. All texts are preprocessed by a linguistic analysis toolkit[1], resulting in a lexical categorisation of each word and a full dependency parse for each sentence. From that analysis, three types of features are extracted to represent sentences: *content words* (I), *function words* (F) and *construction markers* (K).

1.2 Content and function words

All words that are assigned a content part-of-speech category[2] by the lexical analysis are considered mem-

[1] The Conexor Functional Dependency (FDG) parser for English [6]

[2] In this experiment NOUNS, ADJECTIVES, VERBS (including verbal uses of PARTICIPLES), ADVERBS, ABBREVIATIONS, NUMERALS, INTERJECTIONS, and NEGATION are considered *content words*.

bers of the *content word* (I) class and the base form of such words are used as I features when occurring in a sentence.

All other words[3] in a sentence are judged *function words* and their base forms are used as F features in the sentence representation.

1.3 Construction markers

Besides word-based feature classes we introduce a further class intended to capture aspects of the constructions in employ in the sentence. Some of these *constructional features* (K) concern clause semantics and sentence or clause structure – such as the transitivity of the clauses in the sentence, the occurrence of *that*-clauses or relative clauses, the occurrence of predicate constructions, the occurrence of manner, spatial, and temporal adverbials, etc. Other construction markers concern morphological features such as tense forms of verbs present in the sentence or the degree of comparison of occurring adjectives.

As in the case of the word-bases features, these features are extracted from the linguistic analysis. Most of them are based directly from the available information about the morphological or dependency status of a certain word in the sentence, while some other features need the aggregation of information from several words or different analysis levels.

In this experiment all constructional K features are treated as sentence features, exactly as the lexical I and F features are treated, i.e., no coupling between the features and the words carrying them is performed.

1.4 Training target features

The assessments – opinionated or not, and one of the three categories NEU, NEG, POS – are added as yet another lexical-style feature A, along with the I, F, and K features outlined above.

	I believe we have found the appropriate balance, he says.
I	have find appropriate balance say
F	i that we the he
K	<CLSthat> <CMPRabs> <CLStrans> <Tpres> <Tpast> <Tshift>
A	<Opinionated> POS

Figure 1. Example analysis.

2 Computation

Given the above representations, we aggregate them for each sentence using two different methods which we contrast in our submission.

[3] PREPOSITIONS, DETERMINERS, CONJUNCTIONS, PRONOUNS, ...

2.1 First order features

In line with most previous supervised approaches, we use a first order representation in which a sentence is represented as a bag of feature counts. The novelty of our approach lies in that we take constructions into account in addition to surface features, and that we evaluate the effect of using structural terms in contrast to content terms. Each sentence is represented as a concatenation of feature vectors of the different feature types, and the resulting feature vector is then normalized so that its Euclidian length equals 1.

2.2 Second order features

As an alternative we use an implementation of a distributionally determined Word Space Model [4] where we accord each feature a position in a vector space based on which other features it cooccurs with in the training sentences. Intially, each feature – as in the first order representation – is given a representation vector in the form of a vector with a single 1 in a unique position. Each feature is also given an initially empty context vector. This context vector is trained by scanning through each sentence in turn: for each feature present adding in the representation vector for each other feature also present in the sentence. Thus, each feature will in its context vector carry information about every other feature it has cooccurred with.

2.3 Classification algorithms

To classify test sentences based on the features we have determined we use two different supervised learning algorithms. For the first order features, we used a support vector machine (SVM) with a linear kernel.[4] For the word space model we used a simple centroid based classification, targeting the position of the given opinion features NEU, NEG, and POS and selecting sentences with a centroid position at some given threshold angle to the poles.

For the word space model we used a simple centroid based classification, targeting the position of the sentence using the centroid of its component features in word space and measuring its angle to the `<Opinionated>` feature position. This angle, using a suitable threshold is used to assess whether the sentence is opinionated or not. To assess the polarity we perform a similar computation, using the opinion features NEU, NEG, and POS, again selecting sentences with a centroid position at some given threshold angle to the poles. From our previous experiments [5] we expect to find this procedure delivers high recall of opinionated sentences. Distinguishing between polarity is likely to be less precise task for this approach than for a purely lexical approach.

[4] We used the open source LIBLINEAR library [2].

2.4 Feature selection

In the current experiments, we use all possible combinations of the feature types, i.e. I, F, K, IF, IK, FK, IFK.

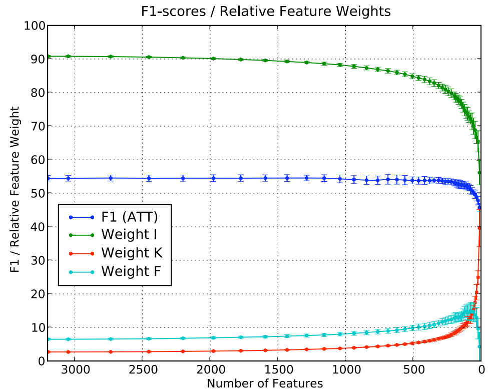
In order to improve generalization ability and to gain some insight into which features are most useful for attitude identification and polarity classification, we use a feature set selection technique, SVM-RFE (Support Vector Machine - Recursive Feature Elimination) [3]. SVM-RFE exploits the duality between the feature space and the instance space in linear discriminant models. The feature selection is conducted in a backward elimination procedure, at each iteration removing the feature with the least influence on the decision boundary. The advantage of this feature selection algorithm compared to traditional algorithms based on e.g. mutual information, is that inter-dependences between features are taken into account. This is important, since it might well be the case that a certain construction or function word is only informative when combined with lexical information. Due to its greedy nature SVM-RFE only finds a locally optimal feature set. Still, it can give valuable information on general characteristics of the problem at hand.

Since attitude identification and polarity classification might benefit from different feature types and training parameter settings, we perform these steps separately for all feature type combinations. In learning the identification step, all available training data is used with attitudinal sentences assigned the class ATT and all other sentences assigned the class NOATT. In learning the classification step, only the attitudinal sentences are used for training data, with sentences assigned one of the classes POS, NEG and NEU.

The results of the feature selection gives us an indication of the utility of the different feature types for the different steps. As illustrated by figure 2, the identification step benefits from constructions and function words in addition to content words, while the classification step only benefits from content words. The top part of figure 2 shows that for identification, it is possible to replace content words by function words and constructions, which we hypothesize leads to a classifier with better generalization ability. The bottom part of figure 2 shows that for classification, removing constructions and function words actually gives an improvement in performance. This indicates that classification is a much more lexicalized process than identification.

3 Submissions

In conclusion, after initial experiments on the training data, we elected to submit two English-language runs based on the word space model, and one based on the SVM scheme.



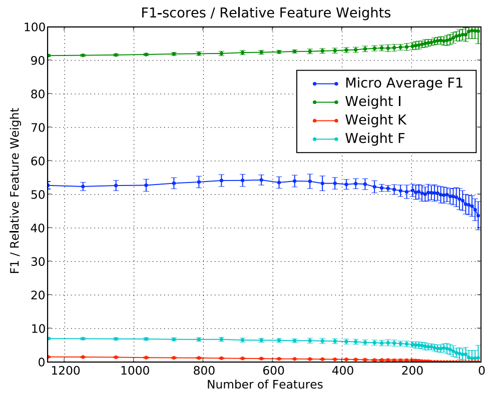


Figure 2. Top: F1-score and relative weights of the feature types I,F and K in the identification step. I features are removed in favor of F and K features. Bottom: Micro Averaged F1-score and relative weights of the feature types I,F and K in the classification step. F and K features are removed in favor of I features. The order of the lines in each graph from top to bottom is: I,$F1$,K,F.

The SVM scheme submission ($sics-1$) is based on the IFK feature set for identification and on the I set for classification. The SVM cost parameter and the number of features to use, were selected in order to optimize F_1-score for identification and in order to optimize accuracy for classification, based on 10-fold cross validation. The word space model runs are based on feature sets IFK ($sics-2$) and IF ($sics-3$) to investigate the effect of the K features on unseen data.

There are still unresolved issues regarding scaling which need further investigation, but on the training data, using 10-fold cross validation, the results from our training runs on NTCIR-6 data are quite encouraging – showing positive effects of using the K features for identifying attitudinal sentences as opposed to non-attitudinal. In distinguishing between the various polarity categories, however, K features do not seem to improve results.

These predictions were borne out in the results. The SVM scheme submission which gave excellent results for identification of opinionated sentences, did not perform better than fair for polarity classification; the word space scheme submissions gave fair results on identification, but decidedly underwhelming results for polarity classification.

This gives us reason to conclude – provisionally, until more empirical verification experiments can be performed – that:

- Linguistic information (in the shape of our K features) *does* help in establishing whether a sentence is opinionated or not; whereas

- Linguistic information of this specific type *does not* help in distinguishing sentences of differing polarity.

The K features in play are chosen to potentially signify opinion but are not specific to negative or positive expressions. That information is encoded in the terms currently included in the I feature set. We will need to refine that set for future experiments.

References

[1] X. Bai, R. Padman, and E. Airoldi. On learning parsimonious models for extracting consumer opinions. In *Proceedings of HICSS-05,the 38th Annual Hawaii International Conference on System Sciences*, page 75b, Washington, DC, USA, 2005. IEEE Computer Society.

[2] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.

[3] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1-3):389–422, 2002.

[4] M. Sahlgren. *The Word-Space Model: using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. PhD thesis, Stockholm University, 2006.

[5] M. Sahlgren, J. Karlgren, and G. Eriksson. Sics: Valence annotation based on seeds in word space. In *Proceedings of Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, page 4, Prague, Czech Republic, 2007.

[6] P. Tapanainen and T. Järvinen. A non-projective dependency parser. In *In Proceedings of the 5th Conference on Applied Natural Language Processing*, pages 64–71, 1997.