Predicative Adjectives: An Unsupervised Criterion to Extract Subjective Adjectives

Michael Wiegand
Spoken Language Systems
Saarland University
michael.wiegand@lsv.uni-saarland.de

Josef Ruppenhofer
Dept. of Information Science
and Language Technology
Hildesheim University
ruppenho@uni-hildesheim.de

Dietrich Klakow
Spoken Language Systems
Saarland University
dietrich.klakow@lsv.uni-saarland.de

Abstract

We examine predicative adjectives as an unsupervised criterion to extract subjective adjectives. We do not only compare this criterion with a weakly supervised extraction method but also with gradable adjectives, i.e. another highly subjective subset of adjectives that can be extracted in an unsupervised fashion. In order to prove the robustness of this extraction method, we will evaluate the extraction with the help of two different state-of-the-art sentiment lexicons (as a gold standard).

1 Introduction

Since the early work on sentiment analysis, it has been established that the part of speech with the highest proportion of subjective words are adjectives (Wiebe et al., 2004) (see Sentence (1)). However, not all adjectives are subjective (2).

(1) A *grumpy* guest made some *impolite* remarks to the *insecure* and *inexperienced* waitress.
(2) The *old* man wearing a *yellow* pullover sat on a *plastic* chair.

This justifies the exploration of criteria to automatically separate the subjective adjectives from the non-subjective adjectives.

In this work, we are interested in an out-of-context assessment of adjectives and therefore evaluate them with the help of sentiment lexicons. We examine the property of being a *predicative* adjective as an extraction criterion. Predicative adjectives are adjectives that do not modify the head of a noun phrase, but which predicate a property of the referent of a noun phrase to which they are linked via a copula or a control predicate (3).

We show that adjectives that frequently occur as predicative adjectives are more likely to convey subjectivity (in general) than adjectives that occur non-predicatively, such as the pre-nominal (attributive) adjectives (4). A subjective adjective may occur both as a predicative (3) and a non-predicative (5) adjective and also convey subjectivity in both contexts. However, a large fraction of non-subjective adjectives do not occur as predicative adjectives (6).

(3) Her idea was *brilliant*.
(4) This is a *financial* problem.
(5) She came up with a *brilliant* idea.
(6) ?The problem is *financial*.

2 Related Work

The extraction of subjective adjectives has already attracted some considerable attention in previous research. Hatzivassiloglou and McKeown (1997) extract polar adjectives by a weakly supervised method in which subjective adjectives are found by searching for adjectives that are conjuncts of a pre-defined set of polar seed adjectives. Wiebe (2000) induces subjective adjectives with the help of distributional similarity. Hatzivassiloglou and Wiebe (2000) examine the properties of dynamic, gradable and polar adjectives as a means to detect subjectivity. Vegnaduzzo (2004) presents another bootstrapping method of extracting subjective adjectives with the help of head nouns of the subjective candidates and distributional similarity. Baroni and Vegnaduzzo

(2004) employ Web-based Mutual information for this task and largely outperform the results produced by Vegnaduzzo (2004).

3 Method

In the following, we present different features with the help of which subjective adjectives can be extracted. For all resulting lists, the adjectives will be ranked according to their frequency of co-occurring with a particular feature.

3.1 Extracting Predicative Adjectives (PRD)

For the extraction of predicative adjectives, we exclusively rely on the output of a dependency parser. Predicative adjectives are usually connected to the subject of the sentence via the dependency label `nsubj` (Example (7) would correspond to Sentence (3)).

(7) `nsubj(brilliant, idea)`

3.2 Extracting Gradable Adjectives (GRD)

As an alternative extraction method, we consider morpho-syntactically *gradable adjectives*. Gradable adjectives, such as *nice* or *small*, are adjectives "that can be inflected to specify the degree or grade of something" (Wiktionary[1]). It has been stated in previous work that if some adjective can build a comparative *nicer* or a superlative *nicest*, then this adjective tends to be subjective (Hatzivassiloglou and Wiebe, 2000).

We employ the property of gradability, since, firstly, it is very predictive towards subjectivity and, secondly, it is the only other unsupervised criterion currently known to extract subjective adjectives. For the extraction of gradable adjectives, we rely, on the one hand, on the part-of-speech labels `JJR` (comparative) and `JJS` (superlative). On the other hand, we also consider adjectives being modified by either *more* or *most*. For the former case, we need to normalize the comparative (*nicer*) or superlative (*nicest*) word form to the canonical positive word form (*nice*) that is commonly used in sentiment lexicons.

[1] `http://en.wiktionary.org/wiki/gradable`

3.3 Weakly-Supervised Extraction (WKS)

We also consider a weakly supervised extraction method in this paper, even though it is not strictly fair to compare such a method with our two previous extraction methods which are completely unsupervised. WKS considers an adjective subjective, if it co-occurs as a conjunct of a previously defined highly subjective (seed) adjective (8). In order to detect such conjunctions, we employ the dependency relation `conj`. By just relying on surface patterns, we would not be able to exclude spurious conjunctions in which other constituents than the two adjectives are coordinated, such as Sentence (10).

(8) This approach is *ill-conceived* and *ineffective*.

(9) `conj(ill-conceived, ineffective)`

(10) [Evil witches are stereotypically dressed in *black*] and [*good* fairies in white].

We also experimented with other related weakly-supervised extraction methods, such as *mutual information* of two adjectives at the sentence level (or even smaller window sizes). However, using conjunctions largely outperformed these alternative approaches so we only pursue conjunctions here.

4 Experiments

As a large unlabeled (training) corpus, we chose the North American News Text Corpus (LDC95T21) comprising approximately 350 million words of news text. For syntactic analysis we use the Stanford Parser (Finkel et al., 2005). In order to decide whether an extracted adjective is subjective or not, we employ two sentiment lexicons, namely the *Subjectivity Lexicon (**SUB**)* (Wilson et al., 2005) and *SO-CAL (**SOC**)* (Taboada et al., 2011). According to the recent in-depth evaluation presented in Taboada et al. (2011), these two sentiment lexicons are the most effective resources for English sentiment analysis. By taking into account two different lexicons, which have also been built independently of each other, we want to provide evidence that our proposed criterion to extract subjective adjectives is not sensitive towards a particular gold standard (which would challenge the general validity of the proposed method).

ALL	other new last many first such next political federal own several few good* former same economic public major recent American second big* foreign high small local military financial little* national
PRD	able* likely available clear* difficult* important* ready* willing* hard* good* due possible* sure* interested unlikely necessary* high responsible* easy* strong* unable* different enough open aware happy impossible* right* wrong* confident*

Table 2: The 30 most frequent adjectives (ALL) and predicative adjectives (PRD); * marks matches with both sentiment lexicons SUB and SOC.

In order to produce the subjective seed adjectives for the weakly supervised extraction, we collect from the sentiment lexicon that we evaluate the n most frequent subjective adjectives according to our corpus. In order to further improve the quality of the seed set, we only consider *strong subjective* expressions from SUB and expressions with the intensity strength ± 5 from SOC.

Table 1 lists the size of the different sentiment lexicons and the rankings produced by the different extraction methods. Of course, the list of all adjectives from the corpus **(ALL)** is the largest list[2] while PRD is the second largest and GRD the third largest. The rankings produced by WKS are fairly sparse, in particular the ones induced with the help of SOC; apparently there are more frequently occurring strong subjective adjectives in SUB than there are high intensity adjectives in SOC.

4.1 Frequent Adjectives vs. Frequent Predicative Adjectives

Table 2 compares the 30 most frequent adjectives (ALL) and predicative adjectives (PRD). Not only does this table show that the proportion of subjective adjectives is much larger among the predicative adjectives but we may also gain some insight into what non-subjective adjectives are excluded. Among the high frequent adjectives are many quantifiers (*many*, *few* and *several*) and ordinal expressions (*first*, *next* and *last*). In principle, most of these expressions are not subjective. One may argue that these adjectives behave like function words. Since they occur very frequently, one might exclude some of them by just ignoring the most frequent adjectives. However, there are also other types of adjectives, especially pertainyms (*political*, *federal*, *economic*, *public*, *American*, *foreign*, *local*, *military*, *financial* and *national*) that appear on this list which could not be excluded by that heuristic. We found that these non-subjective content adjectives are present throughout the entire ranking and they are fairly frequent (on the ranking). On the list of predicative adjectives all these previous types of adjectives are much less frequent. Many of them only occur on lower ranks (and we assume that several of them only got on the list due to parsing errors).

[2]It will also contain many words erroneously tagged as adjectives, however, this is unlikely to affect our experiments since we only focus on the highly ranked (i.e. most frequent) words. The misclassifications rather concern infrequent words.

4.2 Comparison of the Different Extraction Methods

Table 3 compares the precision of the different extraction methods at different cut-off values. It is interesting to see that for ALL in particular the higher ranks are worse than the lower ranks (e.g. rank 1000). We assume that this is due to the high-frequency adjectives which are similar to function words (see Section 4.1). At all cut-off values, however, this baseline is beaten by every other method, including our proposed method PRD. The two unsupervised methods PRD and GRD perform on a par with each other. On SUB, PRD even mostly outperforms GRD. The precision achieved by WKS is quite good. However, the coverage of this method is low. It would require more seed expressions to increase it, however, this would also mean considerably more manual guidance.

Table 3 also shows that the precision of all extraction methods largely drops on the lower ranks. However, one should not conclude from that the extraction methods proposed only work well for highly frequent words. The drop can be mostly explained by the fact that the two sentiment lexicons we use for evaluation are finite (i.e. SUB: 4396 words/SOC: 2827 words (Table 1)), and that neither of these lexicons (nor their union) represents the complete set of all English subjective adjectives. Both lexicons will have a bias towards frequently occurring subjective expressions.

Inspecting the ranks 3001-3020 as displayed in Table 4, for example, actually reveals that there are still many more subjective adjectives than the

Lexicons		Extraction Methods											
					WKS-5		WKS-10		WKS-25		WKS-50	
SUB	SOC	ALL	PRD	GRD	SUB	SOC	SUB	SOC	SUB	SOC	SUB	SOC
4396	2827	212287	20793	7942	292	81	440	131	772	319	1035	385

Table 1: Statistics regarding the size (i.e. number of adjectives) of the different sentiment lexicons and rankings.

artistic* appealable airtight adjustable* activist* accommodating acclimated well-meaning weakest upsetting* unsurpassed unsatisfying* unopposed unobtrusive* unobjectionable unemployable understanding* uncharacteristic submerged speechless

Table 4: A set of entries PRD on lower ranks (ranks 3001-3020); * marks matches with either of the sentiment lexicons SUB or SOC.

matches with our sentiment lexicons suggest. In other words, these are less frequent words; many of them are actually subjective even though they are not listed in the sentiment lexicons. Moreover, irrespective of the drop in precision on the lower ranks, PRD and GRD still outperform ALL on both sentiment lexicons (Table 3). Despite the sparseness of our two gold standards on the lower ranks, we thus have some indication that PRD/GRD are more effective than ALL.

The problem of the evaluation of less-frequent words could not be solved by an extrinsic evaluation, either, e.g. by using the extracted lists for some text classification task (at the sentence/document level). The evaluation on contextual classification on corpora would also be biased towards high-frequency words (as the word distribution is typically Zipfian). For instance, on the MPQA-corpus (Wiebe et al., 2005), i.e. the standard dataset for (fine-grained) sentiment analysis, there is not a single mention of the subjective words *appealable*, *accommodating*, *unsurpassed*, *unopposed*, *unobtrusive* or *speechless*.

4.3 How Different Are Gradable and Predicative Adjectives?

Since in the previous experiments the proportion of subjective adjectives was similar among the gradable adjectives and the predicative adjectives, we may wonder whether these two extraction methods produce the same adjectives. In principle, the set of gradable adjectives extracted is much smaller than the list of extracted predicative adjectives (see Table 1). We found that the gradable adjectives are a proper subset of predicative adjectives, which is in line with the observation by (Bolinger, 1972, 21) that gradable adjectives (which he calls degree words) readily occur predicatively whereas non-gradable ones tend not to.

However, while gradability implies compatibility with predicative use, the reverse is not true. Accordingly, we found adjectives that are definitely not gradable among the predicative adjectives that are subjective, for instance *endless*, *insolvent*, *nonexistent*, *stagnant*, *unavailable* or *untrue*. This means that with the criterion of predicative adjectives one is able to extract relevant subjective adjectives that cannot be caught by the gradability criterion alone, namely *complementary* adjectives that refer to a simple binary opposition (Cruse, 1986, 198-99).

4.4 Intersecting the Different Unsupervised Criteria

In this section, we want to find out whether we can increase the precision by considering intersections of the two different unsupervised extraction criteria. (Due to the sparsity of WKS, it does not make sense to include that method in this experiment.) In our previous experiments it turned out that as far as precision is concerned, our new proposed extraction criterion was similar to the gradability criterion. If, however, the intersection of these two criterions produces better results, then we have provided some further proof of the effectiveness of our proposed criterion (even though we may sacrifice some exclusive subjective adjectives in PRD as pointed out in Section 4.3). It would mean that this criterion is also beneficial in the presence of the gradability criterion.

Figure 1 shows the corresponding results. We computed the intersection of PRD and GRD at using various cut-off values n. The resulting intersection comprises m ranks with $m < n$. The precision of the intersection was consequently compared against the precision of PRD and GRD at rank m. The figure shows that with the exception of the higher ranks on

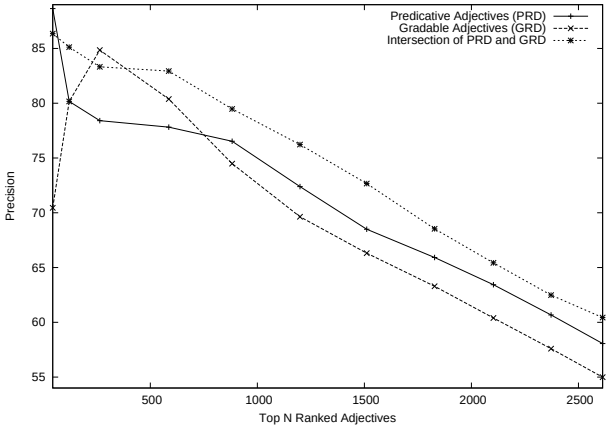
	ALL		PRD		GRD		WKS-5		WKS-10		WKS-25		WKS-50	
Rank n	**SUB**	**SOC**	**SUB**	**SOC**	**SUB**	**SOC**	**SUB**	**SOC**	**SUB**	**SOC**	**SUB**	**SOC**	**SUB**	**SOC**
10	10.00	30.00	90.00	90.00	80.00	60.00	80.00	90.00	80.00	90.00	90.00	70.00	90.00	70.00
25	20.00	32.00	88.00	60.00	64.00	60.00	92.00	80.00	91.00	80.00	92.00	80.00	92.00	84.00
50	30.00	34.00	88.00	64.00	70.00	68.00	82.00	78.00	92.00	78.00	92.00	84.00	90.00	86.00
100	37.00	38.00	81.00	68.00	79.00	75.00	80.00	*N/A*	82.00	72.00	89.00	78.00	92.00	77.00
250	45.60	43.20	79.60	75.60	84.80	76.00	70.80	*N/A*	74.40	*N/A*	80.40	67.50	82.04	67.20
500	48.00	49.20	77.20	70.00	82.20	74.00	*N/A*	*N/A*	*N/A*	*N/A*	72.60	*N/A*	75.20	*N/A*
1000	48.70	48.10	75.50	65.60	72.60	65.00	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	64.30	*N/A*
1500	49.07	46.53	68.60	59.07	66.27	58.60	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*
2000	48.00	43.85	64.55	55.40	61.55	54.25	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*
2500	46.08	40.96	59.52	51.28	56.36	50.00	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*
3000	44.20	39.17	54.63	47.13	51.47	46.03	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*

Table 3: Precision at rank n of the different extraction methods; *WKS-m* denotes that for the extraction the m most frequent subjective adjectives from the respective sentiment lexicon were considered as seed expressions.

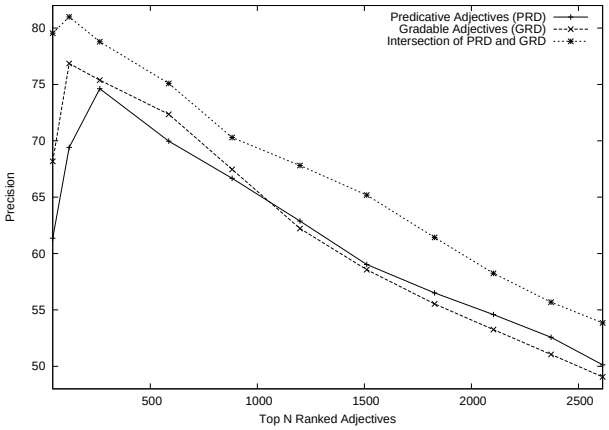
SUB (< 200) there is indeed a systematic increase in precision when the intersection of PRD and GRD is considered.

5 Conclusion

We examined predicative adjectives as a criterion to extract subjective adjectives. As this extraction method is completely unsupervised, it is preferable to weakly supervised extraction methods since we are not dependent on a manually designed high quality seed set and we obtain a much larger set of adjectives. This extraction method is competitive if not slightly better than gradable adjectives. In addition, combining these two unsupervised methods by assessing their intersection results mostly in an increase in precision.



(a) Evaluation on SUB lexicon



(b) Evaluation on SOC lexicon

Figure 1: Comparison of the individual rankings of GRD and PRD with their intersection.

Acknowledgements

This work was performed in the context of the SoftwareCluster project EMERGENT. Michael Wiegand was funded by the German Federal Ministry of Education and Research (BMBF) under grant no. "01IC10S01". Josef Ruppenhofer was supported by the German Research Foundation DFG under grant PI 154/9-3. The authors would like to thank Maite Taboada for providing her sentiment lexicon (SO-CAL) to be used for the experiments presented in this paper.

References

Marco Baroni and Stefano Vegnaduzzo. 2004. Identifying Subjective Adjectives through Web-based Mutual Information. In *Proceedings of KONVENS*, pages 17–24, Vienna, Austria.

Dwight Bolinger. 1972. *Degree words*. Mouton, The Hague.

David Alan Cruse. 1986. *Lexical Semantics*. Cambridge University Press, Cambridge, UK.

Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 363–370, Ann Arbor, MI, USA.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the Semantic Orientation of Adjectives. In *Proceedings of the Conference on European Chapter of the Association for Computational Linguistics (EACL)*, pages 174–181, Madrid, Spain.

Vasileios Hatzivassiloglou and Janyce Wiebe. 2000. Effects of Adjective Orientation and Gradability on Sentence Subjectivity. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pages 299–305, Saarbrücken, Germany.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2):267 – 307.

Stefano Vegnaduzzo. 2004. Acquisition of Subjective Adjectives with Limited Resources. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, Stanford, CA, USA.

Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning Subjective Language. *Computational Linguistics*, 30(3).

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, 39(2/3):164–210.

Janyce M. Wiebe. 2000. Learning Subjective Adjectives from Corpora. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 735–740, Austin, TX, USA.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pages 347–354, Vancouver, BC, Canada.