Inducing Domain-specific Noun Polarity Guided by Domain-independent Polarity Preferences of Adjectives

Klenner, Manfred ; Amsler, Michael ; Hollenstein, Nora

Inducing Domain-specific Noun Polarity Guided by Domain-independent Polarity Preferences of Adjectives

Manfred Klenner
Computational Linguistics
University of Zurich
Switzerland
klenner@cl.uzh.ch

Michael Amsler
Computational Linguistics
University of Zurich
Switzerland
mamsler@ifi.uzh.ch

Nora Hollenstein
Computational Linguistics
University of Zurich
Switzerland
hollenstein@ifi.uzh.ch

Abstract

In this paper, we discuss how domain-specific noun polarity lexicons can be induced. We focus on the generation of good candidates and compare two machine learning scenarios in order to establish an approach that produces high precision. Candidates are generated on the basis of polarity preferences of adjectives derived from a large domain-independent corpus. The *polarity preference* of a word, here an adjective, reflects the distribution of positive, negative and neutral arguments the word takes (here: its nominal head). Given a noun modified by some adjectives, a vote among the polarity preferences of these adjectives establishes a good indicator of the polarity of the noun. In our experiments with five domains, we achieved f-measure of 59% up to 88% on the basis of two machine learning approaches carried out on top of the preference votes.

1 Introduction

Polarity lexicons are crucial for fine-grained sentiment analysis. For instance, in approaches carrying out sentiment composition (Moilanen and Pulman, 2007), where phrase-level polarity is composed out of word level polarity (e.g. disappointed$^-$ hope$^+$ $\rightarrow$ NP$^-$). However, often freely available lexicons are domain-independent, which is a problem with domain-specific texts, since lexical gaps reduce composition anchors. But how many domain-specific words do we have to expect? Is it a real or rather a marginal problem? In our experiments, we found that domain-specific nouns do occur quite often - so they do matter. In one of our domains, we identified about 1000 negative nouns, 409 were domain-specific. In that domain, the finance sector, more than 13'000 noun

types exist that do not occur at all in the DeWac corpus - a large Web corpus (in German) with over 90 Million sentences. Thus, most of them must be regarded as domain-specific. It would be quite time-consuming to go through all of them in order to identify and annotate the polar ones. Could we, rather, predict good candidates? We would need polarity predictors - words that take other, polar words e.g. as their heads. If they, moreover, had a clear-cut preference, i.e. they mostly took one kind of polar words, say negative, then they were perfect predictors of the polarity of nouns. We found that adjectives (e.g. *acute*) can be used as such polarity predictors (e.g. *acute* mostly takes negative nouns, denoted n$^-$, e.g. *acute pain*)).

Our hypothesis is that the polarity preferences of adjectives are (more or less) domain-independent. We can learn the preferences from domain-independent texts and apply it to domain-specific texts and get good candidates of domain-specific polar nouns. Clearly, if the polarity preferences of an adjective are balanced (0.33 for each polarity), than the predictions could not help at all. But if one polarity clearly prevails, we might even get a good performance by just classifying the polarity of unknown nouns in a domain according to the dominant polarity preference of the adjectives they co-occur with.

In this paper, we show how to generate such a preference model on the basis of a large, domain-independent German corpus and a domain-independent German polarity lexicon. We use this model to generate candidate nouns from five domain-specific text collections - ranging from 3'200 up to 37'000 texts per domain. In order to see how far an automatic induction of a domain-specific noun lexicon could go, we also experimented with machine learning scenarios on the output of the baseline system. We experimented with a distributional feature setting on the basis of unigrams and used the Maximum En-

tropy learner, Megam (Daumé III, 2004), to learn a classifier. We also worked with Weka (Frank et al., 2010) and features derived from the German polarity lexicon. Both approaches yield significant gains in terms of precision - so they realize a high-precision scenario.

2 Inducing the Preference Model

We seek to identify adjectives which impose a clear-cut polar preference on their head nouns. The *polarity preference* of an adjective reflects the distribution of positive, negative and neutral nouns the adjective modifies given to some text corpus. We used the domain-independent DeWac corpus (Baroni M., 2009) comprising about 90 million German sentences. We selected those adjectives that frequently co-occurred with polar nouns from PoLex, a freely available German polarity lexicon (Clematide and Klenner, 2010). Since the original polarity lexicon contained no neutral nouns, we first identified 2100 neutral nouns and expanded the lexicon[1]. Altogether 5'500 nouns were available, 2100 neutral, 2100 negative and 1250 positive. For each adjective, we counted how often it took (i.e. modified) positive, negative or neutral nouns in the DeWac corpus and determined their polarity preferences for each class (positive, negative and neutral). This way, 28'500 adjectives got a probability distribution, most of them, however, with a dominating neutral polarity preference. Two lexicons were derived from it: a positive and a negative polarity preference lexicon.

An adjective obeys a polar polarity preference if the sum of its positive and negative polarity preferences is higher than its neutral preference. If the positive preference is higher than the negative, the adjective is a positive polarity predictor, otherwise it is a negative polarity predictor. This procedure leaves us with 506 adjectives, 401 negative polarity predictors and 105 positive polarity predictors. Figure 1 shows some examples of negative polarity predictors. It reveals that, for instance, the adjective *akut* (acute) is mostly coupled with negative nouns (61.50%). Nouns not in PoLex that co-occur with an adjective are not considered. We assume that these unknown nouns of an adjective follow the same distribution that we are sampling from the known co-occurring nouns. Note that po-

larity predictors not necessarily must have a prior polarity itself. Actually, only 3 of the 12 adjectives from Figure 1 do have a prior polarity (indicated as n^-). For instance, the adjective *plötzlich* (immediate) is not polar but has a negative polarity preference. The polarity preference of a word is not useful in composition, it just reveals the empirical (polar) context of the word. If, however, the polarity of the context word is unknown, the preference might license an informed polarity guess.

adjective	English	POS	NEG	#n$^-$
arg$^-$	bad/very	02.65	55.14	301
heftig	intensive	07.73	48.77	814
völlig	total	25.79	42.43	787
akut	acute	06.27	61.50	478
latent	latent	07.96	47.76	402
ziemlich	rather	14.16	52.36	233
drohend$^-$	threatening	35.1	52.54	824
plötzlich	immediate	17.78	41.82	703
gravierend	grave	04.5	48.5	400
chronisch	chronic	03.26	72.11	398
schleichend	subtle	03.76	52.97	319
hemmunglos$^-$	unscrupulous	15.49	43.19	213

Figure 1: Negative Polarity Predictors

Here is the formula for the estimation of the negative polarity preference as given in Figure 1 (n^- denotes a negative noun from PoLex, a_j an adjective modifying an instance of n^-)[2]:

$$prefn^-(a_j) = \frac{\#a_j \, n^-}{\#a_j^{+,-,=}}$$

Note that we count the number of adj-noun types ($\#a_j \, n^-$), not tokens. $\#a_j^{+,-,=}$ is the number of adj-noun types of the adjective a_j for all classes: positive (+), negative(-) and neutral (=).

Figure 2 gives examples of positive polarity predictors with some of their nouns.

German	English	POS
ungetrübt	unclouded	joy
unbeirrbar	unerring	hope
überströmend	overwhelming	happiness
bewunderswert	mirable	competence
falschverstanden	falsely-understood	tolerance
wiedergewonnen	regained	freedom

Figure 2: Positive Polarity Predictors

3 Applying the Preference Model

We applied the preference model to texts from five domains: banks (37'346 texts), transport (3221),

---

[1] We searched for nouns that frequently co-occurred with the same adjectives the polar nouns from the polarity lexicon did and stopped annotating when we reached 2'100 neutral nouns.

[2] This could be interpreted as the conditional probability of a negative noun given the adjective.

insurance (4768), politics (3208) and pharma (4790). These texts have been manually classified over the last 15 years by an institute carrying out media monitoring[3], not only wrt. their domain, but also wrt. target-specific polarity (we just use the domain annotation, currently).

The polarity of a noun is predicted by the vote of the adjectives it occurred with. The following formula shows the polarity prediction $pol^{+,-,=}$ for the class *negative* (pol^-):

$$pol^-(n_i) = A_i * \prod_{a_j \in PM^- \wedge \exists(a_j, n_i)} prefn^-(a_j)$$

A_i is the number of adjectives that modify the noun n_i in the domain-specific texts. PM^- is the set of adjectives from the polarity model (PM) with a negative polarity preference and (a_j, n_i) is true, if the adjective a_j modifies the noun n_i according to the domain-specific documents.

4 Improving the Predictions

The preference model serves two purposes: it generates a list of candidates for polar nouns and it establishes a baseline. We experimented with two feature settings in order to find out whether we could improve on these results.

In the first setting, the WK setting, we wanted to exploit the fact that for some adjectives that modify a noun, we know their prior polarity (from the polarity lexicon). These adjectives do not necessarily have a clear positive or negative polarity preference. If not, then they are not used in the prediction of the noun polarity.

But could the co-occurrence of a noun with adjectives bearing a prior polarity also be indicative of the noun polarity? For instance, if a noun is coupled frequently and exclusively with negative adjectives. Does this indicate something? Ones intuition might mislead, but a machine learning approach could reveal correlations. We used Simple Logistic Regression (SRL) from Weka and the following features:

1. the number of positive adjectives with a prior polarity that modify the noun

2. the number of negative adjectives with a prior polarity that modify the noun

3. the difference between 1) and 2): absolute and ratio

4. the ratio of positive and negative adjectives

5. two binary features indicating the majority class

6. three features for the output of the preference model: the positive, negative and neutral scores: $pol^-, pol^+, pol^=$, respectively.

In the second setting, the MG setting, we trained Megam, a Maximum Entropy learner, among the following lines: we took all polar nouns from PoLex and extracted from the DeWac corpus all sentences containing these nouns. For each noun, all (context) words (nouns, adjectives, verbs) co-occurring with it in these sentences are used as bag of words training vectors. In other words, we learned a tri-partite classifier to predict the polarity class (positive, negative or neutral) given a target noun and its context, i.e. those nouns co-occurring with it in a text collection.

5 Experiments

The goal of our experiments were the prediction of positive and negative domain-specific nouns in five domains. We used our preference model to generate candidates. Then we manually annotated the results in order to obtain a domain-specific gold standard. We evaluated the output of the preference model relative to the new gold standards and we run our experiments with Megam and Weka's Simple Logistic Regression (SRL). Megam and Weka's SLR were trained on the basis of the positive, negative and neutral nouns from PoLex and the DeWac corpus.

Figure 3 shows the results. #PM gives the number of nouns predicted by the preference model to be negative (e.g. 220 in the politics domain). These are the nouns we annotated for polarity and that formed our gold standard afterwards (e.g. 75.90 out of 110 predicted are true negative nouns and are kept as the gold standard). Since the generation of the gold standard is based on the preference model's output, its recall is 1. We cannot fix the real recall since this would require to manually classify all nouns occurring in those texts (e.g. 13'000 in the banks domain). However, since we wanted to compare the machine learning performance with the preference model, we had to mea-

ID	domain	texts	#PM	prec	f	#WK	prec	rec	f	#MG	prec	rec	f
D1	politics	3208	220	75.90	**86.29**	195	78.97	92.22	83.26	130	*81.54*	63.48	69.13
D2	transport	3221	141	71.63	**83.47**	127	73.22	92.07	80.57	64	*78.12*	49.50	58.54
D3	insurance	4768	255	76.86	**86.91**	238	78.57	95.40	85.13	155	*79.35*	62.75	69.09
D4	pharma	4790	257	71.59	**83.44**	228	76.75	95.11	81.69	137	*87.83*	65.40	68.35
D5	banks	37346	1013	70.38	**88.02**	825	77.84	90.07	79.02	437	*81.23*	49.78	58.32

Figure 3: Prediction of Negative Nouns

sure recall, otherwise we could not determine the overall performance.

From Figure 3 we can see that the preference model (PM) performs best in terms of f-measure (in bold). Of course, recall (i.e. 1, not shown) is idealized, since we took the output of the preference model to generate the gold standard. Note however that this was our premise, that we needed an approach that delivers good candidates, otherwise we were lost given the vast amount of candidate nouns (e.g. remember the 13'000 nouns in the finance sector).

German	English
Wertverminderung	impairment of assets
Stagflation	stagflation
Geldschwemme	money glut
Überhitzungssymptom	overheating symptom
Hyperinflation	hyperinflation
Euroschwäche	weakness of the euro
Werterosion	erosion in value
Nachfrageüberhang	surplus in demand
Margendruck	pressure on margins
Klumpenrisiko	cluster risk
Virus	virus
Handekzem	hand eczema
Schweinegrippe	swine flu
Gebärmutterriss	ruptured uterus
Alzheimer	Alzheimer
Sehstörung	defective eye sight
Tinnitus	tinnitus

Figure 4: Domain-specific Negative Nouns

Figure 4 shows examples of negative nouns from two domains: banks and pharma. But: are all found nouns *domain-specific* negative nouns? In the bank domain, we have manually annotated for domain specificity: out of 1013 nouns predicted to be negative by the model, 409 actually were domain-specific (40.3 %)[4]. The other nouns could also be in a domain-independent polarity lexicon.

Now, we turn to the prediction of positive domain-specific nouns. It is not really surprising that the preference model is unbalanced - that there are far more negative than positive polarity predictors: 401 compared to 105. PoLex, the pool

of nouns used for learning of the polarity preferences already is unbalanced (2100 negative compared to 1250 positive nouns). Also, the majority of the texts in our five domains are negative (all texts are annotated for document-level polarity). It is obvious then that our model is better in the prediction of negative than positive polarity. Actually, our base model comprising 105 positive polarity predictors does not trigger often within the whole corpus. For instance, only 10 predictions were made in the banks domain, despite the 37'346 texts. Clearly, newspaper texts often are critical and thus more negative than positive vocabulary is used. This explains the very low recall.

However, what if we relaxed our model? If we, for example, keep those adjectives in our model that have a positive polarity preference > 0.35, at least 35 out of 100 nouns co-occurring with those adjectives should be positive.

ID	#1	prec	#2	prec	#3	prec	#4	prec
D1	18	66.6	25	60.0	25	60.0	8	50
D2	14	85.7	16	75.0	0	0	3	33.3
D3	13	69.2	15	60.0	5	100	1	100
D4	13	84.6	15	80.0	9	55.5	2	100
D5	135	76.2	174	71.2	58	87.9	40	82.5

Figure 5: Prediction of Positive Nouns

We report the results of two runs. The first one, labelled #1, where adjectives are used to predict a positive noun polarity if they have a positive polarity preference > 0.35 and where the negative polarity preference is < 0.1. In the second run, labelled #2, we only require the positive preference to be > 0.35. Table 5 shows the results. We also show the results of Weka (label #3) and Megam (label #4) for the candidates generated by #2.

Compared to the negative settings, the number of found positive nouns is rather low. For instance, in the banks domain, 174 nouns were suggested compared to 1013 negative ones. However, precision has not dropped and it is especially higher than the threshold value of 0.35 seemed to indicate (as discussed previously). Weka (#3) and Megam (#4) again show better precision, however

---

[4]51 of the 131 (38.93%) as positive classified nouns actually were domain-specific.

the number of found nouns is too low (in a setting that suffers already from low numbers). Figure 6 shows a couple of found positive nouns.

German	English
Versammlungsfreiheit	freedom of assembly
Ausländerintegration	integration of foreigners
Einlagesicherung	deposit protection
Lohntransparenz	wage transparency
Haushaltsdisziplin	budgetary discipline
Vertriebsstärke	marketing strength
Anlegervertrauen	confidence of investors
Kritikfähigkeit	ability for criticism
Führungskompetenz	leadership competencies

Figure 6: Predicted Positive Nouns

So far, we have discussed a binary approach where each class (positive, negative) was predicted and classified independently and where especially no adjectives with a neutral preference where considered. What happens if we include these adjectives? The results are given in Figure 7.

domain	#neg	prec	#pos	prec
banks	288	80.16	3	66.66
pharma	141	70.92	32	68.75
transport	78	67.94	0	0
politics	115	76.52	0	0
insurance	132	66.66	0	0

Figure 7: Unrestricted Prediction of Noun Polarity

Although precision is good, the results are very conservative, e.g. in the banks domain, only 288 nouns were found compared to 1013 nouns given the binary mode. Recall and f-measure are lower compared to the binary setting. The huge amount of neutral preference adjectives (about 28'000) seems to neutralize polar tendencies. But even then, some predictions survive - so these contexts seem to be strong.

6 Related Work

The expansion or creation of sentiment lexicons has been investigated in many variations from different perspectives and for various goals. Liu and Zhang (2012) subdivide the work in this field into three groups: manual approaches, dictionary-based approaches and corpus-based approaches. While the manual approach is time-consuming, it is still often used to create core lexicons which are not domain-specific, e.g. (Taboada et al., 2011).

The dictionary-based approaches which are also called thesaurus-based approaches (Huang et al., 2014) try to make use of existing dictionaries or thesauri like WordNet (e.g. (Esuli and Sebastiani, 2006; Baccianella et al., 2010; Neviarouskaya et al., 2011)) while the corpus-based approaches rely on statistical measures based on different concepts, for example, sentiment consistency (Hatzivassiloglou and McKeown, 1997), pointwise mutual information (Turney, 2002), context coherency (Kanayama and Nasukawa, 2006), double propagation (Qiu et al., 2011) or label propagation (Huang et al., 2014). Our approach is based on the use of an existing dictionary and of an domain-independent corpus. But rather than using the corpus to directly detect new entries for the lexicon, we use it to derive the polarity preference of adjectives which in turn is used to generate candidates from the domain-specific corpus.

The model most similar to our approach is (Klenner and Petrakis, 2014), where the contextual and prior polarity of nouns is learned from the polarity preference of verbs for the verb's direct object. However, no attempt is made to induce domain-specific polarity as we do. We also focus on the polarity preference of adjectives and we also try to improve precision by machine learning.

7 Conclusions

We have introduced a plain model for the induction of domain-specific noun lexicons. First, the polarity preferences of adjectives are learned from domain-independent text and from a general polarity lexicon. A voting approach then predicts noun polarity from adjective noun pairings sampled from domain-specific texts. The predictions based only on adjectives acting as positive or negative polarity predictors perform astonishingly well. Machine Learning can be used to improve precision at the cost of recall. Our approach thus even might be useful for fully automatic generation of a high precision, domain-specific prior noun polarity lexicons.

In future work, we will apply our approach to other languages than German. We then will also have to cope with multiword expressions as well, since compounds not longer - as in German - come as single words. We also would like to carry out an extrinsic evaluation in order to see how big the impact of an induced domain-specific lexicon on polarity text classification actually is.

References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proc. of LREC 2010*, volume 10, pages 2200–2204.

Ferraresi A. Zanchetta E. Baroni M., Bernardini S. 2009. The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226.

Simon Clematide and Manfred Klenner. 2010. Evaluation and extension of a polarity lexicon for German. In *Proceedings of the First Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 7–13.

Hal Daumé III. 2004. Notes on CG and LM-BFGS optimization of logistic regression. *Paper available at http://pub.hal3.name#daume04cg-bfgs, implementation available at http://hal3.name/megam.*

Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proc. of LREC 2006*, volume 6, pages 417–422.

Eibe Frank, Mark Hall, Geoffrey Holmes, Richard Kirkby, Bernhard Pfahringer, Ian H. Witten, and Len Trigg. 2010. Weka-A Machine Learning Workbench for Data Mining. In Oded Maimon and Lior Rokach, editors, *Data Mining and Knowledge Discovery Handbook*, chapter 66, pages 1269–1277. Springer US, Boston, MA.

Vasileios Hatzivassiloglou and Kathleen R McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proc. of the ACL 1997*, pages 174–181. Association for Computational Linguistics.

Sheng Huang, Zhendong Niu, and Chongyang Shi. 2014. Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. *Knowledge-Based Systems*, 56:191–200.

Hiroshi Kanayama and Tetsuya Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proc. of EMNLP 2006*, pages 355–363. Association for Computational Linguistics.

Manfred Klenner and Stefanos Petrakis. 2014. Inducing the contextual and prior polarity of nouns from the induced polarity preference of verbs. *Data & Knowledge Engineering*, 90:13–21.

Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In *Mining Text Data*, pages 415–463. Springer.

Karo Moilanen and Stephen Pulman. 2007. Sentiment composition. In *Proc. of RANLP 2007*, pages 378–382, Borovets, Bulgaria, September 27-29.

Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. 2011. Sentiful: A lexicon for sentiment analysis. *Affective Computing, IEEE Transactions on*, 2(1):22–36.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.

Peter D Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proc. of the ACL 2002*, pages 417–424. Association for Computational Linguistics.