Q-WordNet: Extracting Polarity from WordNet Senses

Rodrigo Agerri
GSI Group
Universidad Politécnica de Madrid (UPM)
r.agerri@upm.es

Ana García-Serrano
NLP & IR Group
UNED, Madrid
agarcia@lsi.uned.es

1 Introduction

This paper presents Q-WordNet,[1] a lexical resource consisting of WordNet senses automatically annotated by positive and negative polarity. Polarity classification amounts to decide whether a text (sense, sentence, etc.) is associated to a *positive* or *negative* connotations. This task is becoming important for determining opinions about commercial products, on companies reputation management, brand monitoring, or to track attitudes by mining online forums, blogs, etc. Inspired by work on classification of word senses by polarity in SentiWordNet (Esuli and Sebastiani, 2006), and taking WordNet as a starting point, we create Q-WordNet, but instead of applying supervised classifiers, we decided to effectively maximize the linguistic information contained in WordNet (by human annotators). The quantitative evaluation of Q-WordNet as a binary classification task shows important improvements on SentiWordNet. However, we would like to stress that Q-WordNet is not a finished resource. Although it can be used in its current form for lexical sentiment analysis tasks (like Affective Task in SemEval-07), it can also be used as a training set for supervised classifiers that would subsequently be applied for the improvement of Q-WordNet.

2 Previous Related Work

There is a huge amount of work on opinion mining at document level focusing on the automatic analysis of commercial and cultural products (Pang and Lee, 2004; Danescu-Niculescu-Mizil et al., 2009). Other previous approaches aim at determining the subjectivity of sentences by means of *terms* that are markers of opinionated content (Kim and Hovy, 2004; Takamura et al., 2005). These approaches also classify subjective words by their polarity. Note that these approaches classify *words* instead of senses, which means that they are not able to capture the fact that a word may contain various senses of which some of them could have different polarities.

There seems to be an assumption in these works that polarity classification (to determine whether the opinion expressed by a word sense, sentence or text) actually depends on subjectivity detection. In other words, that prior to the task of assigning polarity we need to determine whether it is objective (factual) or subjective (opinion). From this point of view, only those expressions deemed to be subjective are classified by polarity (positive or negative) (Kim and Hovy, 2004; Takamura et al., 2005). This leaves out all those expressions that are suitable to be classified as *objectively* positive or negative (Su and Markert, 2008). Clear cases are those denoting illnesses, such as 'cancer', 'tuberculosis', etc., which stereotypically carry negative associations. Note that these senses may not be negative in a medical domain. We therefore take the "general domain" to be a stereotypical domain, but our classification may need to be refined for its use in other domains.

There are some approaches that work at sense level: Su and Markert (2009) and Wiebe and Milhacea (2006) annotate subjectivity and objectivity of word senses without assuming that subjectivity is a previous step of polarity classification. Moreover, (Esuli and Sebastiani, 2006) annotate WordNet 2.0 senses by using a ternary polarity classification (Positive, Negative and Objective) in which each polarity value is assign a numerical score in such a way that the sum of the three scores is 1.0. They proceed by manually annotating a seed of senses which carry strong polarity associations and then follow a number of WordNet relations between the senses to expand the original seeds. This set is used to train 7 supervised classifiers to do the annotation. An evaluation of their classification is provided as an estimation of its Mean Squared Error (Esuli, 2008). They seem to be assuming that those senses that do not get classified as either Positive or Negative are in fact Objective, namely, they associate opin-

[1] http://nlp.uned.es/semantics/qwordnet.tar.gz

ionated content with positive or negative associations.

The work presented in this paper focuses on the classification of WordNet senses by their polarity regardless of whether they express subjective opinions or factual information. Unlike SentiWordNet (Esuli and Sebastiani, 2006), which is built using supervised classifiers to annotate WordNet senses, we exploit the linguistic information (provided by human annotators) contained in WordNet itself.

3 Extracting Polarity from WordNet senses

Our approach to classifying WordNet senses by polarity is based on the view of polarity as an *association of a positive or a negative quality to something or to someone*. The idea is to (1) link a sense to a quality (e.g., positive or negative) and then (2) devise, if needed, a procedure to quantify such association, e.g., using similarity measures (Agirre et al., 2009), or confidence scores in our classification by establishing a ranking (Esuli and Sebastiani, 2007). We focus here on (1), building a polarity annotated lexical resource. Assigning polarity is therefore seen in this work as a binary classification task. This differs from approaches such as SentiWordNet in which positive or negative associations are assigned to word senses at a certain level. We believe that it is not very intuitive to say that some sense is positive or negative at the 0.125 level. For example the *too-bad#a#1* synset is considered to be positive at the 0.125 level, and 0.75 at a negative level. It does seem very difficult to effectively use such a fine-grained classification.

In consonance with this qualitative aspect, our objective is also to maximize the human effort employed in building WordNet, and see how far we could go by *walking* WordNet collecting positive and negative senses as we pass by. Of course, the issue here is where to start *walking*. Whereas the common approach has been to start from a manually collected seed terms, we will not do any manual collection but turn to the quality synsets in WordNet itself. There are 5 noun quality synsets in WordNet 2.0 and 3.0, two of which contain attribute relations (to adjectives). For the synset *quality#n#01* the attribute relation takes us to *positive#a#01*, *negative#a#01*, *good#a#01* and *bad#a#01*; *quality#n#02* leads to *superior#a#01* and *inferior#a#02*. We take these six synsets expressing positive and negative qualities to be the departure, and then algorithmically walk through some WordNet relations collecting (i.e., annotating) those synsets that are accessible from the seeds. The result is Q-WordNet (Q from Quality), the set of WordNet synsets we classify by positive or negative polarity (quality).

Several other approaches have used WordNet relations to extract affective-related words or senses from WordNet. WordNet-Affect 1.0 (Strapparava and Valitutti, 2004) consists of annotating WordNet 1.6 senses with labels describing an emotional state by expanding on an initial list of 1903 manually collected terms. In total the resource contains 2874 synsets annotated with affective information. The WordNet relations they use are antonym, similarity, pertains-to, derived-from, also-see and attribute, because they consider that they preserve the affective meaning.

As we mentioned earlier, the corpus used to train the supervised classifiers used for SentiWordNet 1.0 was obtained by expanding from an initial list of manually collected terms, using the same set of relations used for WordNet-Affect. The general idea is also employed by Su and Markert (2008) although they use more WordNet relations: antonym, similar-to, derived-from, attribute, also-see, direct hyponym, direct-hypernym, extended antonym.

It is quite clear that if we want Q-WordNet to be proportionate in terms of POS, then we may not only need to walk through WordNet senses, but also to jump from one part-of-speech (POS) to the other. Otherwise, as our seeds are adjectives, we risk Q-WordNet consisting mainly of adjectives plus some testimonial nouns, verbs and adverbs. We therefore need to use every WordNet relation, not just those that preserve the affective content, as in WordNet-Affect and SentiWordNet (Strapparava and Valitutti, 2004; Esuli and Sebastiani, 2006) that would allow us to jump from adjectives to other POS. We complement this with the information contained in the glosses. Once we have gone through all WordNet relations, we parse WordNet's glosses looking for those synsets we have already collected. The parsing in done by a pipeline consisting of the free available toos, C&C tokenizer (Clark and Curran, httpsvnaskitusydeduautraccandc), CRFTagger (Phan, 2006) and Stanford Dependency parser (de Marneffe et al., 2006). If a positive synset is matched in a gloss, then the synset whose gloss we have analyzed is also annotated as positive.

Using every available WordNet relation and their glosses in the way described above is bound to cause a lot of noise in the classified data, namely, a lot of false positives and false negatives. Two measures were taken to prevent this. First, at every step in our algorithm, e.g. at the application of every WordNet relation or in the glosses analysis, we filter those synsets that appear in both positive and negative categories. Second, the syntactic parse of the glosses allow us to detect those matched synsets that fall under the scope of a *negation*. For example, if 'not pleasant' is found in a synset's gloss, and we have a 'pleasant' synset among our positives, then the gloss's synset is classified as negative (not positive).

Summarizing, our algorithm to built Q-WordNet starts at the attributes of *quality#n#01* and *quality#n#02*, which

are adjectives. We perform 10 iterations over the *also_see* relation (our experiments showed that more than 10 iterations creates too much noise). We then go to *similar-to*. From all the collected adjectives, we get their attributes, which allows us to move to nouns. We obtain nouns and verbs from adjectives through the lexical relation *derived-from*. We then use *hyperonymy*, *hyponymy*, *pertainyms*, *derived-from*, *verb-group* and *cause* (plus the *antonym* relation at every step to filter false positives and negatives) to nouns and verbs. Most of the adverbs in Q-WordNet are extracted from a reverse application of the *pertainym* relation to adjectives (adverbs are pertainyms of adjectives).

We obtain 7402 positive and 8108 negative synsets from the application of our method to WordNet 3.0. We are able to automatically annotate 2884 as positive and 2100 as negative in WordNet 2.0. As a comparison, WordNet-Affect consists of 2874 synsets (admittedly, affect annotation is harder than polarity). Although we cannot straitforwardly compare to SentiWordNet (built from WordNet 2.0) because they their classification is graded at a specific level, table 1 shows when the polarity scores are lower, then the number of polarity-classified synsets grows quite large (35049 synsets) but at the cost of a huge increase false positives and negatives. We discuss this and other issues in the next section, which describes the evaluation and comparison of Q-WordNet to SentiWordNet 1.0.

4 Evaluation

In order to evaluate Q-WordNet and compare it to previous related approaches such as SentiWordNet, we use MicroWnOp as testset (Esuli and Sebastiani, 2006), a manually created corpus of senses annotated by polarity. However, as MicroWnOp is annotated using SentiWordNet's graded three score method, some modifications to are required. First, all Objective synsets are removed, as we want to evaluate positive and negative cases in a binary classification task. This results in a final testset of 737 synsets at the lowest polarity score, namely, 0.125. As we increase the score, the number of synsets decreases. For example, at 0.375 level, the number of synsets classified by polarity is 527. Second, some parts of MicroWnOp were annotated by more than one annotator. For these cases, the polarity scores were averaged. The binary classification evaluation is performed in terms of precision, recall and F measure for positive classification, and negative precision, true negative rate (TNR) and F Measure for negatives. We also measure accuracy. This evaluation method means that systems will get lower results as the number of false positives and negatives increases.

Table 1 shows that Q-WordNet clearly outperforms SentiWordNet at every level except at 0.875 and 1.0 where SentiWordNet gets perfect scores. At these levels, the testset consisted of 238 and 195 synsets respectively, and SentiWordNet of 466 and 14. As Q-WordNet does not at this state quantify or rank the polarity annotation, it gets evaluated on all 4984 synsets at every level, which in principle would make it more vulnerable to false positive and negatives. We say in principle because the differences between SentiWordNet and Q-WordNet at 0.875 and 1.0 are not significant.

Another issue worth noticing is the fact that as the level of confidence decreases, Q-WordNet results do not degrade as much as those of SentiWordNet. This may be due to the fact that our strict filtering of synsets during the annotation process has left us with very few false positives and negatives. As Q-WordNet is by no means a finished resource, we believe that these results show excellent potential to carry on enriching it with a better linguistic processing of glosses (perhaps using disambiguated glosses) and its use as training data to build classifiers which would be later be deployed on WordNet for a larger and richer polarity annotated resource. Our procedure is also suitable to be combined to similarity and ranking algorithms to offer graded polarity is required by any particular applications.

Acknowledgments

This work has been supported by Madrid R+D Regional Plan, MAVIR Project, S-0505/TIC/000267. (http://www.mavir.net)

References

Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Pasca, and Aitor Soroa. 2009. A study on similarity and relatedness using distributional and wordnet-based approaches. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 19–27, Boulder, Colorado, June. Association for Computational Linguistics.

Stephen Clark and James Curran. http://svn.ask.it.usyd.edu.au/trac/candc. C&C tools.

Cristian Danescu-Niculescu-Mizil, Gueorgi Kossinets, Jon Kleinberg, and Lillian Lee. 2009. How opinions are received by online communities: A case study on Amazon.com helpfulness votes. In *Proceedings of the WWW*, pages 141–150.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of Language Resources and Evaluation Conference (LREC)*.

pol@	SentiWordNet 1.0								Q-WordNet (4984 synsets)						
	Positive			Negative					Positive			Negative			
	Prec.	Rec.	F	PNeg	TNR	F	Acc.	Syns	Prec.	Rec.	F	PNeg.	TNR	F	Acc.
0.125	0.66	0.68	0.67	0.63	0.60	0.62	0.65	35049	0.84	0.95	0.89	0.89	0.67	0.76	0.85
0.25	0.77	0.76	0.76	0.70	0.71	0.70	0.74	22855	0.89	0.95	0.92	0.88	0.77	0.83	0.89
0.375	0.84	0.87	0.85	0.83	0.78	0.80	0.83	14611	0.94	0.97	0.95	0.93	0.86	0.89	0.93
0.5	0.88	0.93	0.90	0.88	0.80	0.84	0.88	9445	0.94	0.97	0.96	0.92	0.88	0.90	0.94
0.625	0.88	0.90	0.89	0.84	0.80	0.82	0.86	4923	0.96	0.96	0.96	0.91	0.91	0.91	0.95
0.75	0.92	1.0	0.96	1.0	0.84	0.91	0.94	2027	0.96	0.97	0.97	0.94	0.91	0.92	0.95
0.875	1.0	1.0	1.0	1.0	1.0	1.0	1.0	466	0.96	1.0	0.98	1.0	0.93	0.97	0.98
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	14	0.98	1.0	0.99	1.0	0.96	0.98	0.99

Table 1: Comparing SentiWN and Q-WN.

A. Esuli and F. Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)*, pages 417–422, Genoa, Italy, May.

Andrea Esuli and Fabrizio Sebastiani. 2007. Pageranking wordnet synsets: An application to opinion mining. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 424–431, Prague, Czech Republic, June. Association for Computational Linguistics.

Andrea Esuli. 2008. *Automatic Generation of Lexical Resources for Opinion Mining: Models, Algorithms and Applications*. Ph.D. thesis, Università di Pisa.

Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of Coling 2004*, pages 1367–1373, Geneva, Switzerland, Aug 23–Aug 27. COLING.

Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pages 271–278, Barcelona, Spain, July.

Xuan-Hieu Phan. 2006. CRFTagger: CRF English POS Tagger, http://sourceforge.net/projects/crftagger.

Carlo Strapparava and Alessandro Valitutti. 2004. Wordnet-affect: an affective extension of wordnet. In *Proceedings of the 4th International Conference on Languages Resources and Evaluation (LREC 2004)*, pages 1083–1086, Lisbon, May.

Fangzhong Su and Katja Markert. 2008. From words to senses: A case study of subjectivity recognition. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 825–832, Manchester, UK, August.

Fangzhong Su and Katja Markert. 2009. Subjectivity recognition on word senses via semi-supervised mincuts. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1–9, Boulder, Colorado, June. Association for Computational Linguistics.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 133–140, Ann Arbor, Michigan, June. Association for Computational Linguistics.

Janyce Wiebe and Rada Mihalcea. 2006. Word sense and subjectivity. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 1065–1072, Sydney, Australia, July. Association for Computational Linguistics.