A CCG-based System for Valence Shifting for Sentiment Analysis

František Simančík and Mark Lee

School of Computer Science, University of Birmingham, Birmingham, UK
frantisek.simancik@worc.ox.ac.uk, m.g.lee@cs.bham.ac.uk

Abstract. The automatic classification of sentiment in text is becoming an important area of research. In this work, we present a linguistic system for sentence-level valence annotation. Our system uses the formalism of Combinatory Categorial Grammar to represent words as functions acting on their syntactic arguments, which provides a unified way of implementing various classes of valence shifters. We propose two simple semi-automatic methods for estimating the valence of individual terms based on the lexical relations of WordNet. We evaluate the system on the data generated for the Affective Text task of SemEval 2007 and show that it compares favourably with the systems participating in the task.

Keywords: sentiment analysis, valence annotation, valence shifters, headlines, combinatory categorial grammar.

1 Introduction

The number of opinion-rich resources such as discussions, blogs and review sites has been growing rapidly in recent years. As a result of this, there is a demand for tools capable of classifying texts not only by the topic but also the attitude and opinion they convey; giving rise to new areas in Natural Language Processing called Opinion Mining and Sentiment Analysis.

One of the most prominent tasks in the field is the classification of valence (positive/negative orientation). Researchers (Pang et al. [7], Kennedy and Inkpen [6] and others) have successfully applied supervised machine learning methods[1] to determine the valence of longer texts. These approaches rely on the availability of a large amount of human-tagged training data and, compared to linguistic methods, reveal very little about the nature of the connection between a text and the opinion it expresses.

[1] Naive Bayes Classifiers, Support Vector Machines, etc.

© A. Gelbukh (Ed.)
Advances in Computational Linguistics.
Research in Computing Science 41, 2009, pp. 99-108

Received 10/11/08
Accepted 15/12/08
Final version 04/02/09

The Affective Text task (Strapparava and Mihalcea [11]), conducted as a part of SemEval 2007, focuses on unsupervised sentence-level classification of emotions and valence in newspaper headlines. The reason behind this, the authors said, was to emphasise the study of emotion lexical semantics, and avoid biasing participants toward simple "text categorization" approaches ([11]). Indeed, the average length of a headline in the SemEval data is only seven words, which is too short to be susceptible to statistical analysis without adequate training data.

The task consisted of two independent parts: emotion labelling (using a fixed set of predefined labels) and ternary valence annotation. We present a system for the second subtask. The reasons for choosing the setting of SemEval were twofold - it sets a well-defined problem and gives us direct means of comparing our results with other existing systems.

1.1 The SemEval Data Sets and Evaluation

The data sets gathered for the Affective Text task were formed of newspaper headlines, which are believed to have a high load of emotional content and are therefore suitable for sentence-level sentiment analysis ([11]). The headlines were collected from major online newspaper portals such as New York Times, CNN and BBC News.

The participants were presented with a smaller development data set consisting of 250 headlines, while the final submissions were evaluated on a larger test set with 1000 headlines. The valence of each headline was labelled independently by six human annotators in the interval $[-100, 100]$. For the coarse-grained evaluation it was subsequently mapped to three classes: negative $[-100, -50]$, neutral $(-50, 50)$ and positive $[50, 100]$.

1.2 Brief Outline of Our Method

Similarly to other existing systems for this task (Andreevskaia and Bergler [1], Chaumartin [2]), we use a pre-built dictionary of sentiment-bearing unigrams, which provides a mapping from terms to their valence. To construct the dictionary, we manually compiled a list of seed words with strong valence and then extended it through WordNet's lexical links.

Using the valence dictionary alone is, however, not enough. Consider, for example, the following sentence from the development data set: *"Nigeria hostage feared dead is freed."* which has a positive meaning even though it contains three negative (*hostage, fear, dead*) and only one positive word (*free*).

In order to improve the performance, we enhanced the simple bag-of-words approach by employing sentence-level valence shifters: words which influence the sentiment expressed by other words in the sentence (Polanyi and Zaenen [8]). In the example above, it is the role of the phrase *"is freed"* to shift the valence of its subject *"Nigeria hostage feared dead"* from negative to positive.

We believe that most words (and verbs in particular) may exhibit valence-shifting behaviour. In our model, each term potentially affects the valence of all its syntactic arguments (subjects/objects). This effect is always in the form of multiplication by an appropriate factor, which may be different for different arguments. Thus, for example, a transitive word (e.g. *reject*) may flip the valence of its object while preserving its subject unchanged.

We extend this analysis one step further. The effect of a term on its arguments is not only applied to their valence but also to the effects they have themselves. Thus, for example, *not* flips the effect of *very* in *"not very"* from intensifying to diminishing. Similarly, under this model, the negating effect of *reject* on its object will be inverted in *"don't reject"*, producing a phrase which is neutral to both its subject and object.

We applied Combinatory Categorial Grammar (Steedman [10]) to determine the structural dependencies between individual terms in a sentence. The main reason for this decision was the fact that the syntax of CCG gives rise to a semantic interpretation whose structure (e.g. treating adjectives as functions from nouns to nouns) maps easily to the functionality of valence shifters. We used Clark and Curran's CCG parser ([3]) which seems to work reasonably well even on fragmented sentences.

2 Resources

In this section we will introduce the resources, theories and formalisms which form the basis of our system.

- WordNet 3.0
- Contextual Valence Shifters
- Combinatory Categorial Grammar

2.1 WordNet 3.0

WordNet ([4]), one of the best-known NLP resources, is a lexical database of English developed and maintained at Princeton University. It organises nouns, verbs, adjectives and adverbs into groups of synonyms (synsets), with each synset representing a distinct meaning (word sense). The synsets are interconnected by various semantic links, of which the following are the most relevant to our purpose: *hyponym* (links to a more specific concept), *hypernym* (links to a more general concept), *similar to* and *see also.* The last two are, however, only present amongst adjectives.

2.2 Contextual Valence Shifters

The presence of certain words and phrases in a sentence can modify (intensify, diminish or even flip) the valence expressed by other terms. For instance, in the sentence *"He is not bright."* the valence of *bright* is shifted by *not* from positive to negative. In the following subsections we describe the valence shifters we used in our system. Polanyi and Zaenen ([8]) investigate this phenomenon to a great depth; we implement only a fraction of their suggestions.

Negatives, Intensifiers and Diminishers

Negatives (*not, none, never...*), intensifiers (*very, rather...*) and diminishers (*slightly, a bit...*) are the most obvious valence shifters. In our model, the effect of such a term is to multiply the valence of its argument, which can be a single word or a longer constituent, by a predefined factor (-1 in case of negatives) and also modify its effect appropriately (e.g. negating an intensifier results into a diminisher, intensifying a diminisher produces a stronger diminisher, affecting a neutral term leaves it unchanged).

It has to be recognised, however, that this is an oversimplification. There are occasions on which the above approach is insufficient, often when two or more of these terms compose. For example, consider the phrase *"not very good"*, whose meaning depends strongly on the context and may range from negative to slightly positive. Under our model it always evaluates to the same as *"quite good"*.

Connectors

Certain conjunctions (*but, while, although, however...*) are often used to set up a deliberate contrast in the discussion by firstly introducing a new piece of information and contradicting it immediately. In such cases, it is only the main clause of the sentence which expresses the attitude of the speaker, the effect of the first clause is neutralised by the connector. For example:

> *The plot sounds promising but the audience is likely to leave unimpressed.*

Verbs

Even though not directly mentioned in [8], we believe that verbs have the strongest impact on the overall sentiment of a sentence. For very short sentences, as in the case of headlines, these are often the only valence shifters present at all and their role must not be overlooked. Consider these examples:

> *EU criticises the war in Georgia.*
> *Threat against airlines has been eliminated.*

Although both the sentences are composed of negative and neutral words only, the verbs *criticise* and *eliminate* flip the valence of their objects from negative to positive, and the resulting messages are thus positive. There are many other verbs with this functionality: *attack, stop, forbid, prevent, dislike, reject...*

Other less radical verbs may act on their objects by weakening or intensifying their valence (*emphasise, support, increase...*).

2.3 Combinatory Categorial Grammar

Combinatory Categorial Grammar (Steedman [10]) is a grammatical theory based on categorial calculus and combinatory logic. It provides a completely straightforward mapping from the syntactic properties of its terms to their semantic functionalities, yet it is still efficiently parseable.

In a categorial grammar, all constituents (including individual words) are assigned specific categories describing their syntactic behaviour. Combinatory rules[2] (functional application, composition, etc.) then specify how phrases can combine into larger constituents according to their categories.

The class of syntactic categories can be defined recursively as the set including the atomic categories N (noun), NP (noun phrase), PP (prepositional phrase), S (sentence), and others, and complex categories (compound of the atomic categories) of the form X/Y and $X\backslash Y$, where X and Y are categories.

Using Steedman's notation, complex categories X/Y and $X\backslash Y$ are functors taking argument of category Y and returning a result of category X. The type of the slash specifies the directionality of the argument: / indicates that the argument appears to the right of the functor, whereas \ means that the argument comes to the left.

For example, the category of an English adjective may be written as N/N, indicating that it is a function from nouns (which it takes to its right) to nouns. Similarly, a typical transitive verb has the category $(S\backslash NP)/NP$, making the verb a curried function of two noun phrases (its object and subject) producing a sentence.

To demonstrate the correspondence between syntax and semantics, consider the parse of the following sentence:

$$
\begin{array}{ccccc}
\text{John} & \text{has} & \text{very} & \text{little} & \text{money} \\
\hline
N & (S\backslash NP)/NP & (N/N)/(N/N) & (N/N) & N \\
\cline{1-1}\cline{3-4}
NP & & \multicolumn{2}{c}{N/N} & \\
\cline{3-5}
 & & \multicolumn{3}{c}{N} \\
\cline{3-5}
 & & \multicolumn{3}{c}{NP} \\
\cline{2-5}
 & \multicolumn{4}{c}{S\backslash NP} \\
\hline
\multicolumn{5}{c}{\mathrm{S}}
\end{array}
$$

Giving rise (by means of functional application) to the following semantic structure:

$$
\begin{array}{ccccc}
\text{John} & \text{has} & \text{very} & \text{little} & \text{money} \\
\hline
John & \lambda x.\lambda y.has(x)(y) & \lambda f.\lambda x.very(f)(x) & \lambda x.little(x) & money \\
\cline{3-4}
 & & \multicolumn{2}{c}{\lambda x.very(little)(x)} & \\
\cline{3-5}
 & & \multicolumn{3}{c}{very(little)(money)} \\
\cline{2-5}
 & \multicolumn{4}{c}{\lambda y.has(very(little)(money))(y)} \\
\hline
\multicolumn{5}{c}{has(very(little)(money))(John)}
\end{array}
$$

[2]See Steedman [10] or Hockenmaier [5] for full treatment of combinatory rules.

3 The System

Our system consists of four basic components: a dictionary of sentiment-bearing and valence-shifting words, a simple preprocessor, the C&C CCG parser (Clark and Curran [3]), and a classifier, which links the other components together.

3.1 Construction of the Dictionary

Starting with a small set of manually selected seed words, we follow the links in WordNet to derive a larger collection of words with similar meaning. Because the structure of WordNet (in terms of which links are present) differs significantly across the word classes, we propose two different methods for this task: one treats adjectives and adverbs, the other one nouns and verbs.

It is necessary to address the problem of homonymy at this point. The context provided by newspaper headlines is too short to perform sense disambiguation, so we decided to ignore this issue completely and to each word form we simply assign the valence of its most common synset (based on WordNet's relative frequency counts).

The dictionary also contains a short list of common negatives, intensifiers, diminishers and conjunctions (as mentioned in 2.2).

Adjectives and Adverbs

We compiled small sets of positively (*good, beautiful, happy, pleasant, clean, friendly, healthy, correct, lucky, alive, clever*) and negatively (*bad, hideous, sad, unpleasant, dirty, hostile, sick, wrong, unfortunate, dead, stupid*) oriented adjectives, capturing a variety of distinct concepts from the class of sentiment-bearing terms.

Taking one of the seed words at a time, we perform a breadth-first search starting from its three most frequent synsets and proceeding along the *similar to* and *see also* links. These restrict the search space to the class of adjective synsets and, in our experience, reliably preserve valence while still provide satisfactory expansion. To avoid exploring the vast space of irrelevant terms, we terminate the search at depth 10 and treat all the unexplored synsets as being at depth 11.

To calculate the valence of an adjective synset, we sum its distances from the negative seeds and subtract its distances from the positive seeds, employing the intuition that positive synsets are closer to the positive seeds than to the negative ones. Finally, we scale the valence to the interval $[-100, 100]$.

We applied the same method to derive a list of valence-shifting adjectives, using seed sets with increasing (*extreme, large, huge, enormous, immense*) and decreasing (*mild, small, minute, micro, slight*) semantics. This time we restricted our search to the *similar to* links only, which preserve the meaning more accurately than the *see also* ones. The resulting value was mapped exponentially to the range $[\frac{1}{2}, 2]$.

This way we obtained about 4200 sentiment-bearing and 100 quantity-affecting adjectives.

Our treatment of adverbs is morphological. Those derived from adjectives (by appending a morpheme such as *-y, -ly, -ily*) are given the same properties as the corresponding adjectives. This is also applied to common noun-generating morphemes (*-ity, -ness*).

Nouns and Verbs

There are no similarity links amongst nouns and verbs in WordNet, so the above approach (and its adaptations) cannot be used. Instead, we turn our attention to hyponymy, which is often the only semantic link present at all. We compiled a list of general concepts whose valence and effect (as described in 2.2) were estimated manually. All their hyponyms were then assigned the same values as the original concepts.

Our selection was based on the trial data set, but a lot remained open to our intuition only. We tried to include concepts which are likely to appear in a newspaper setting, such as *catastrophe, misfortune, disrespect, immorality, pain, fear, mistreat* and *celebrate, pleasure, protect, cure, wonder, success.*

The entire list contains a hundred concepts. These were inflated through hyponymy into about 5700 synsets giving us 3900 different word forms.

3.2 Parsing and Preprocessing

The C&C parser expects all tokens (even quotation and punctuation marks) to be separated by spaces; we use a simple preprocessor to achieve this. While inserting an extra space is enough in most cases, certain grammatical constructions require special care. For example, we detect negated auxiliaries and expand them to their long forms (e.g. *don't* to *do not*).

We then present the processed text to the parser, which annotates each token with its grammatical category and produces a structure of combinatory rules to be used at each level. Conveniently, the parser also labels words with their morphological base forms, which simplifies their look-up in the dictionary.

3.3 Classification

The final component combines information from the dictionary and the parser. Firstly, all the words are converted into functions of zero or more parameters according to their syntactic categories. The atomic categories are mapped to a basic type, whose instances are completely described by their valence alone. The complex categories give rise to functional types and we treat them as functions modifying the valence and effect of their arguments. Their action is represented by their own valence and by one scaling factor for each of their parameters, which describes their effect (intensification, diminution, negation or just neutral propagation) on that particular argument. Scaling a function by a constant results into multiplying its valence by that constant and modifying its own effects appropriately (e.g. negating an intensifier yields a diminisher). All these numbers are looked up in the dictionary and neutral values are used (0 for valence and 1 for scaling factor) if no matching is found.

A collection of handwritten rules (one for each type) defines how a function processes its arguments. Most rules fall under this scheme: a function takes several arguments of the same type X, scales each of them by the corresponding factor, sums their valence and combines them into one object, and adds its own valence to the result, which is again of type X.

The above scheme applies to, among others, the following type classes:

X/X , the category of adjectives ($X = N$), certain adverbs ($X = N/N$) and common negatives.

(X\X)/X , the category of conjunctions.

(S\NP)[/NP... /NP] , the category of common verbs. The innermost NP refers to the subject, the others to the objects of the verb.

Once all the functions are defined, they are combined according to the combinatory rules suggested by the parser. In an ideal case, the final category of a headline would be S, giving us directly a result of the basic type. Quite often, however, a headline is only a fragment of a sentence and its category is complex. In this case, we return the valence of the resulting function, which corresponds to evaluating it on neutral arguments.

4 Results

Table 1 shows the results of our system on the test data. The full system achieves accuracy of 63.20% and F-score (Rijsbergen [9]) of 51.81 and compares favourably with the systems participating in the SemeEval task[3], where the best results were 55.10% for accuracy and 42.43 for F-measure and, as shown in Table 2, even these were obtained by two different systems.

Table 1 also shows how the performance changes when we restrict the dictionary to certain word classes. It transpires that the effect of adjectives and adverbs is only marginal and the system draws its strength from its treatment of nouns and verbs. We attribute this to the fact that newspaper headlines are often too short to contain any sentiment-bearing adjectives, in which case their valence has to be determined from nouns and verbs only.

5 Conclusions

The results of the Affective Text task indicate that valence annotation is not easy. Our system performs relatively well (for a ternary classifier) in both

[3] See [11] for the full table of results.

Table 1: System results on the test data.

Dictionary in use	Accuracy	Precision	Recall	F_1
full	63.20	53.21	50.48	51.81
adjectives and adverbs only	58.80	43.48	2.44	4.62
nouns and verbs only	62.30	52.00	50.73	51.36

Table 2: The best systems (achieving highest accuracy and F-measure) participating in the SemEval task.

System	Accuracy	Precision	Recall	F_1
CLaC	55.10	61.42	9.20	16.00
CLaC-NB	31.20	31.18	66.38	42.43

precision and recall and improves upon the results obtained by the participating programs.

We adopted the formalism of Combinatory Categorial Grammar to represent words as functions acting on their arguments, which provides a unified and transparent way of implementing some common classes of valence shifters. Our work also emphasises the role of nouns and verbs in short sentence sentiment tagging. We argue that if WordNet is to be used to estimate their valence, the absence of the similarity-like links forces us to abandon the methods commonly used for adjectives. Instead, we proposed a crude semi-automatic approach based on hyponymy.

Acknowledgments. This project was carried out under the Summer Research Scholarship scheme at the School of Computer Science, the University of Birmingham. František Simančík was supported by a grant from the SPP Foundation.

References

1. Andreevskaia, A & Bergler, S 2007, 'CLaC and CLaC-NB: Knowledge-based and corpus-based approaches to sentiment tagging', *Proceedings of SemEval-2007: 4th International Workshop on Semantic Evaluations at ACL 2007*, Prague.
2. Chaumartin, FR 2007, 'UPAR7: A knowledge-based system for headline sentiment tagging', *Proceedings of SemEval-2007: 4th International Workshop on Semantic Evaluations at ACL 2007*, Prague.
3. Clark, S & Curran, JR 2007, 'Wide-Coverage Efficient Statistical Parsing with CCG and Log-Linear Models', *Computational Linguistics*, vol. 33, no. 4, pp. 493-552.

4. Fellbaum, C 1998 *WordNet, An Electronic Lexical Database*, MIT Press, Cambridge, MA (1998).
5. Hockenmaier, J 2003, 'Data and Models for Statistical Parsing with Combinatory Categorial Grammar', PhD thesis, University of Edinburgh.
6. Kennedy, A & Inkpen, D 2005. 'Sentiment Classification of Movie and Product Reviews Using Contextual Valence Shifters', *Proceedings of FINEXIN-05, Workshop on the Analysis of Informal and Formal Information Exchange during Negotiations*.
7. Pang, B, Lee, L & Vaithyanathan, S 2002, 'Thumbs up? Sentiment Classification using Machine Learning Techniques', *Proceedings of EMNLP-02, the Conference on Empirical Methods in Natural Language Processing*, Philadelphia, US, pp. 79-86.
8. Polanyi, L & Zaenen, A 2005 'Contextual valence shifters', *Computing Attitude and Affect in Text: Theory and Applications*, pp. 1-10.
9. Rijsbergen, CJV 1979, *Information Retrieval*, Butterworth-Heinemann.
10. Steedman, M 2000, *The Syntactic Process*, Spech and Communication Series, MIT Press Language.
11. Strapparava, C & Mihalcea, R 2007, 'SemEval-2007 Taks 14: Affective Text', *Proceedings of SemEval-2007: 4th International Workshop on Semantic Evaluations at ACL 2007*, Prague.