

The Good, the Bad, and the Unknown: Morphosyllabic Sentiment Tagging of Unseen Words

Karo Moilanen and Stephen Pulman

Oxford University Computing Laboratory

Wolfson Building, Parks Road, Oxford, OX1 3QD, England

{ Karo.Moilanen | Stephen.Pulman }@comlab.ox.ac.uk

Abstract

The omnipresence of unknown words is a problem that any NLP component needs to address in some form. While there exist many established techniques for dealing with unknown words in the realm of POS-tagging, for example, guessing unknown words' semantic properties is a less-explored area with greater challenges. In this paper, we study the semantic field of sentiment and propose five methods for assigning prior sentiment polarities to unknown words based on known sentiment carriers. Tested on 2000 cases, the methods mirror human judgements closely in three- and two-way polarity classification tasks, and reach accuracies above 63% and 81%, respectively.

1 Introduction

One of the first challenges in sentiment analysis is the vast lexical diversity of subjective language. Gaps in lexical coverage will be a problem for any sentiment classification algorithm that does not have some way of intelligently guessing the polarity of unknown words. The problem is exacerbated further by misspellings of known words and POS-tagging errors which are often difficult to distinguish from genuinely unknown words. This study explores the extent to which it is possible to categorise words which present themselves as unknown, but which may contain known components using morphological, syllabic, and shallow parsing devices.

2 Morphosyllabic Modelling

Our core sentiment lexicon contains 41109 entries tagged with positive (+), neutral (N), or nega-

tive (-) prior polarities (e.g. *lovely*⁽⁺⁾, *vast*^(N), *murder*⁽⁻⁾) across all word classes. Polarity reversal lexemes are tagged as [-] (e.g. *never*^{(N)[-]}). We furthermore maintain an auxiliary lexicon of 314967 known neutral words such as names of people, organisations, and geographical locations.

Each unknown word is run through a series of sentiment indicator tests that aim at identifying in it at least one possible **sentiment stem** - the longest subpart of the word with a known (+), (N), or (-) prior polarity. An unknown word such as *healthcare-related*^(?) can be traced back to the stems *health*^{(N)(+)}, *care*⁽⁺⁾, *healthcare*⁽⁺⁾, or *relate*^{(d)(N)} which are all more likely to be found in the lexica, for example. Note that the term 'stem' here does not have its usual linguistic meaning but rather means 'known labelled form', whether complex or not.

We employ a classifier society of five rule-driven classifiers that require no training data. Each classifier adopts a specific analytical strategy within a specific window inside the unknown word, and outputs three separate polarity scores based on the number of stems found (S_{pos} , S_{ntr} , S_{neg}) (initially 1). The score for polarity p for unknown word w is calculated as follows:

$$(1) \quad scr(p) = \Phi_p \frac{L_s}{L_w} \frac{1}{S_w} \frac{S_p}{S_{pos} + S_{ntr} + S_{neg}}$$

where Φ_p = polarity coefficient (default 1)
 L_s = # of characters in the stem
 L_w = # of characters in w
 S_w = # of punctuation splits in w

Polarity coefficients balance the stem counts: in particular, (N) polarity is suppressed by a Φ_{ntr} of < 1

because (N) stem counts dominate in the vast majority of cases. L_s reflects differing degrees of reliability between short and long stems in order to favour the latter. S_w targets the increased ambiguity potential in longer punctuated constructs. The highest-scoring polarity across the three polarity scores from each of the five classifiers is assigned to w .

Conversion [A]. It is generally beneficial to impose word class polarity constraints in the lexicon (e.g. $[smart]^{(+)}_{ADJ}$ vs. $[smart]^{(-)}_{V}$). Due to creative lexical conversion across word classes, hard constraints can however become counterproductive. The first classifier estimates zero-derived paronyms by retagging the unknown word with different POS tags and requerying the lexica.

Morphological Derivation [B]. The second classifier relies on regular derivational (e.g. *-ism*, *-ify*, *-esque*) and inflectional (e.g. *-est*, *-s*) morphology. The unknown word is transformed incrementally into shorter paronymic aliases using pure affixes and (pseudo and neo-classical) combining forms. A recursive derivation table of find/replace pairs is used to model individual affixes and their regular spelling alternations (e.g. *-pping*▷*p*; *-ation*▷*e*; *-iness*▷*y*; *-some*▷ \emptyset ; *re-*▷ \emptyset). Polarity reversal affixes such as *-less*^{(N)[¬]} and *not-so-*^{(N)[¬]} are supported. The table is traversed until a non-neutral sentiment (NB. not morphological) stem is found. Prefixes are matched first. Note that the prefix-driven configuration we have adopted is an approximation to a (theoretically) full morphemic parse. The derivation for *antirationalistic*^(?), for example, first matches the prefix *anti-*^{(N)[¬]}, and then truncates the immediate constituent *rationalistic*^(?) incrementally until a sentiment stem (e.g. *rational*^{(N)(+)}) is encountered. The polarity reversal prefix *anti-*^{(N)[¬]} then reverses the polarity of the stem: hence, *antirationalistic*^(?)▷*rationalistic*^(?)▷*rationalist*^(?)▷*rational*⁽⁺⁾▷*antirationalistic*⁽⁻⁾. 322 (N) and 67 [¬] prefixes, and 174 (N) and 28 [¬] suffixes were used.

Affix-like Polarity Markers [C]. Beyond the realm of pure morphemes, many non-neutral sentiment markers exist. Examples include prefix-like elements in *well-built*⁽⁺⁾, *badly-behaving*⁽⁻⁾, and *strange-looking*⁽⁻⁾; and suffix-like ones in *rat-infested*⁽⁻⁾, *burglarproof*⁽⁺⁾, and *fruit-loving*⁽⁺⁾. Because the polarity of a

non-neutral marker commonly dominates over its host, the marker propagates its sentiment across the entire word. Hence, a full-blown derivation is not required (e.g. *easy-to-install*^(?)▷*easy-to-install*⁽⁺⁾; *necrophobia*^(?)▷*necrophobia*⁽⁻⁾). We experimented with 756 productive prefixes and 640 suffixes derived from hyphenated tokens with a frequency of ≥ 20 amongst 406253 words mined from the WAC 2006 corpus¹. Sentiment markers are captured through simple regular expression-based longest-first matching.

Syllables [D]. We next split unknown words into individual syllables based on syllabic onset, nucleus, and coda boundaries obtained from our own rule-based syllable chunker. Starting with the longest, the resultant monosyllabic and permutative order-preserving polysyllabic words are used as aliases to search the lexica. Aliases not found in our lexica are treated as (N). Consider the unknown word *freedomfortibet*^(?). In the syllabified set of singular syllables {*free*, *dom*, *for*, *ti*, *bet*} and combinatory permutations such as {*free.dom*, *dom.ti*, *for.ti.bet*, ...}, *free* or *free.dom* are identified as (+) while all others become (N). Depending on the Φ_{ntr} value, *free.dom.for.ti.bet*^(?) can then be tagged as (+) due to the (+) stem(s). Note that cruder substring-based methods can always be used instead. However, a syllabic approach shrinks the search space and ensures the phonotactic well-formedness of the aliases.

Shallow Parsing [E]. At a deepest level, we approximate the internal quasi-syntactic structure of unknown words that can be split based on various punctuation characters. Both exotic phrasal nonce forms (e.g. *kill-the-monster-if-it's-green-and-ugly*⁽⁻⁾) and simpler punctuated compounds (e.g. *butt-ugly*⁽⁻⁾, *girl-friend*⁽⁺⁾) follow observable syntactic hierarchies amongst their subconstituents. Similar rankings can be postulated for sentiment. Since not all constituents are of equal importance, the sentiment salience of each subconstituent is estimated using a subset of the grammatical polarity rankings and compositional processes proposed in Moilanen and Pulman (2007). The unknown word is split into a virtual sentence and POS-tagged². The rightmost subconstituent

¹Fletcher, W. H. (2007). English Web Corpus 2006. www.webascorpus.org/searchwc.html

²Connexor Machine Syntax 3.8. www.connexor.com

Table 1: Average (A)ccuracy, kappa, and error distribution against ANN-2 and ANN-3

Classifier	Φ_{ntr}	ALL POL		NON-NTR		¬LAZY	ERROR DISTRIBUTION		
		A	k	A	k	A	FATAL	GREEDY	LAZY
[A] CONVERSION	.2	76.70	.03	96.88	.94	99.53	0.08	2.47	97.44
[B] DERIVATION	.8	74.15	.11	80.05	.59	93.90	2.81	22.86	74.33
[C] AFFIX MARKERS	.2	72.33	.21	77.93	.55	88.05	6.10	39.07	54.83
[D] SYLLABLES	.8	69.55	.23	71.88	.45	82.75	9.37	48.62	42.01
[E] PARSING	.7	64.33	.25	79.09	.59	73.50	9.03	65.40	25.57
ALL		63.20	.28	80.61	.61	70.20	9.49	71.41	19.10
ALL ¬UNSURE		64.60	.28	82.19	.64	69.71	7.43	77.95	14.62

in the word is expanded incrementally leftwards by combining it with its left neighbour until the whole word has been analysed. At each step, the sentiment grammar in *idem.* controls (i) non-neutral sentiment propagation and (ii) polarity conflict resolution to calculate a global polarity for the current composite construct. The unknown word *help-children-in-distress*^(?) follows the sequence $N:[\underline{distress}^{(-)}]^{(-)} \triangleright PP:[\underline{in}^{(N)} \underline{distress}^{(-)}]^{(-)} \triangleright NP:[\underline{children}^{(N)}]^{(N)} [\underline{in} \underline{distress}]^{(-)}]^{(-)} \triangleright VP:[\underline{help}^{(+)}]^{(+)} [\underline{children} \underline{in} \underline{distress}]^{(-)}]^{(+)}$, and is thus tagged as (+).

3 Evaluation

We compiled a dataset of 2000 infrequent words containing hapax legomena from the BNC³ and “junk” entries from the WAC 2006 corpus (Footnote 1). The dataset contains simple, medium-complexity, and extreme complex cases covering single words, (non-)hyphenated compounds, nonce forms, and spelling anomalies (e.g. *anti-neo-nazi-initiatives*, *funny-because-its-true*, and *s’gonnacostyaguvna*). Three human annotators classified the entries as (+), (-), or (N) (with an optional UNSURE tag) with the following distribution:

(2)

Human	(+)	(N)	(-)	UNSURE
ANN-1	24.55	53.45	22	11.75
ANN-2	12.60	68.60	18.80	10.85
ANN-3	5.25	84.55	10.20	0.65

We report results using all polarities (ALL-POL) and non-neutral polarities (NON-NTR) resulting in average pairwise inter-annotator Kappa scores of

³Kilgarriff, A. (1995). BNC database and word frequency lists. www.kilgarriff.co.uk/bnc-readme.html

.40 (ALL-POL) and .74 (NON-NTR), or .48 (ALL-POL) and .83 (NON-NTR) without UNSURE cases. We used ANN-1’s data to adjust the Φ_{ntr} coefficients of individual classifiers, and evaluated the system against both ANN-2 and ANN-3. The average scores between ANN-2 and ANN-3 are given in Table 1.

Since even human polarity judgements become fuzzier near the neutral/non-neutral boundary due to differing personal degrees of sensitivity towards neutrality (cf. low (N) agreement in Ex. 2; Andreevskaia and Bergler (2006)), not all classification errors are equal for classifying a (+) case as (N) is more tolerable than classifying it as (-), for example. We therefore found it useful to characterise three distinct disagreement classes between human *H* and machine *M* encompassing FATAL ($H^{(+)}M^{(-)}$ or $H^{(-)}M^{(+)}$), GREEDY ($H^{(N)}M^{(-)}$ or $H^{(N)}M^{(+)}$), and LAZY ($H^{(+)}M^{(N)}$ or $H^{(-)}M^{(N)}$) cases.

The classifiers generally mimic human judgements in that accuracy is much lower in the three-way classification task - a pattern concurring with past observations (cf. Esuli and Sebastiani (2006); Andreevskaia and Bergler (2006)). Crucially, FATAL errors remain below 10% throughout. Further advances can be made by fine-tuning the Φ_{ntr} coefficients, and by learning weights for individual classifiers which can currently mask each other and suppress the correct analysis when run collectively.

4 Related Work

Past research in Sentiment Tagging (cf. Opinion Mining, Sentiment Extraction) has targeted classification along the subjectivity, sentiment polarity, and strength/degree dimensions towards a common goal

of (semi-)automatic compilation of sentiment lexica. The utility of word-internal sentiment clues has not yet been explored in the area, to our knowledge.

Lexicographic Methods. Static dictionary-/thesaurus-based methods rely on the lexical-semantic knowledge and glosses in existing lexicographic resources alongside known non-neutral seed words. The semi-supervised learning method in Esuli and Sebastiani (2005) involves constructing a training set of non-neutral words using WordNet synsets, glosses and examples by iteratively adding syn- and antonyms to it and learning a term classifier on the glosses of the terms in the training set. Esuli and Sebastiani (2006) used the method to cover objective (N) cases. Kamps et al. (2004) developed a graph-theoretic model of WordNet’s synonymy relations to determine the polarity of adjectives based on their distance to words indicative of subjective evaluation, potency, and activity dimensions. Takamura et al. (2005) apply to words’ polarities a physical spin model inspired by the behaviour of electrons with a (+) or (-) direction, and an iterative term-neighbourhood matrix which models magnetisation. Non-neutral adjectives were extracted from WordNet and assigned fuzzy sentiment category membership/centrality scores and tags in Andreevskaia and Bergler (2006).

Corpus-based Methods. Lexicographic methods are necessarily confined within the underlying resources. Much greater coverage can be had with syntactic or co-occurrence patterns across large corpora. Hatzivassiloglou and McKeown (1997) clustered adjectives into (+) and (-) sets based on conjunction constructions, weighted similarity graphs, minimum-cuts, supervised learning, and clustering. A popular, more general unsupervised method was introduced in Turney and Littman (2003) which induces the polarity of a word from its Pointwise Mutual Information (PMI) or Latent Semantic Analysis (LSA) scores obtained from a web search engine against a few paradigmatic (+) and (-) seeds. Kaji and Kitsuregawa (2007) describe a method for harvesting sentiment words from non-neutral sentences extracted from Japanese web documents based on structural layout clues. Strong adjectival subjectivity clues were mined in Wiebe (2000) with a distributional similarity-based word clustering method seeded by hand-labelled annotation.

Riloff et al. (2003) mined subjective nouns from unannotated texts with two bootstrapping algorithms that exploit lexico-syntactic extraction patterns and manually-selected subjective seeds.

5 Conclusion

In this study of unknown words in the domain of sentiment analysis, we presented five methods for guessing the prior polarities of unknown words based on known sentiment carriers. The evaluation results, which mirror human sentiment judgements, indicate that the methods can account for many unknown words, and that over- and insensitivity towards neutral polarity is the main source of errors.

References

- Alina Andreevskaia and Sabine Bergler. 2006. Mining WordNet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses. In *Proceedings of EACL 2006*.
- Andrea Esuli and Fabrizio Sebastiani. 2005. Determining the Semantic Orientation of Terms through Gloss Classification. In *Proceedings of CIKM 2005*.
- Andrea Esuli and Fabrizio Sebastiani. 2006. Determining Term Subjectivity and Term Orientation for Opinion Mining. In *Proceedings of EACL 2006*.
- Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the Semantic Orientation of Adjectives. In *Proceedings of ACL 1997*.
- Jaap Kamps, Maarten Marx, Robert J. Mokken and Maarten de Rijke. 2004. Using WordNet to Measure Semantic Orientations of Adjectives. In *Proceedings of LREC 2004*.
- Nabuhiko Kaji and Masaru Kitsuregawa. 2007. Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents. In *Proceedings of EMNLP-CoNLL 2007*.
- Karo Moilanen and Stephen Pulman. 2007. Sentiment Composition. In *Proceedings of RANLP 2007*.
- Ellen Riloff, Janyce Wiebe and Theresa Wilson. 2003. Learning Subjective Nouns using Extraction Pattern Bootstrapping. In *Proceedings of CoNLL 2003*.
- Hiroya Takamura, Takashi Inui and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *Proceedings of ACL 2005*.
- Peter Turney and Michael Littman. 2003. Measuring Praise and Criticism: Inference of Semantic Orientation from Association. *ACM Transactions on Information Systems*, October, 21(4): 315–46.
- Janyce Wiebe. 2000. Learning Subjective Adjectives from Corpora. In *Proceedings of AAAI 2000*.