DepecheMood:
a Lexicon for Emotion Analysis from Crowd-Annotated News

Jacopo Staiano
University of Trento
Trento - Italy
staiano@disi.unitn.it

Marco Guerini
Trento RISE
Trento - Italy
marco.guerini@trentorise.eu

Abstract

While many lexica annotated with words polarity are available for sentiment analysis, very few tackle the harder task of emotion analysis and are usually quite limited in coverage. In this paper, we present a novel approach for extracting – in a totally automated way – a high-coverage and high-precision lexicon of roughly 37 thousand terms annotated with emotion scores, called DepecheMood. Our approach exploits in an original way 'crowd-sourced' affective annotation implicitly provided by readers of news articles from rappler.com. By providing new state-of-the-art performances in unsupervised settings for regression and classification tasks, even using a naïve approach, our experiments show the beneficial impact of harvesting social media data for affective lexicon building.

1 Introduction

Sentiment analysis has proved useful in several application scenarios, for instance in buzz monitoring – the marketing technique for keeping track of consumer responses to services and products – where identifying positive and negative customer experiences helps to assess product and service demand, tackle crisis management, etc.

On the other hand, the use of finer-grained models, accounting for the role of individual emotions, is still in its infancy. The simple division in 'positive' vs. 'negative' comments may not suffice, as in these examples: '*I'm so miserable, I dropped my IPhone in the water and now it's not working anymore*' (SADNESS) vs. '*I am very upset, my new IPhone keeps not working!*' (ANGER). While both texts express a negative sentiment, the latter, connected to anger, is more relevant for buzz monitoring. Thus, emotion analysis represents a natural evolution of sentiment analysis.

Many approaches to sentiment analysis make use of lexical resources – i.e. lists of positive and negative words – often deployed as baselines or as features for other methods, usually machine learning based (Liu and Zhang, 2012). In these lexica, words are associated with their prior polarity, i.e. whether such word out of context evokes something positive or something negative. For example, *wonderful* has a positive connotation – prior polarity – while *horrible* has a negative one.

The quest for a high precision and high coverage lexicon, where words are associated with either sentiment or emotion scores, has several reasons. First, it is fundamental for tasks such as affective modification of existing texts, where words' polarity together with their score are necessary for creating multiple *graded* variations of the original text (Inkpen et al., 2006; Guerini et al., 2008; Whitehead and Cavedon, 2010).

Second, considering word order makes a difference in sentiment analysis. This calls for a role of compositionality, where the score of a sentence is computed by composing the scores of the words up in the syntactic tree. Works worth mentioning in this connection are: Socher et al. (2013), which uses recursive neural networks to learn compositional rules for sentiment analysis, and (Neviarouskaya et al., 2009; Neviarouskaya et al., 2011) which exploit hand-coded rules to compose the emotions expressed by words in a sentence. In this respect, compositional approaches represent a new promising trend, since all other approaches, either using semantic similarity or Bag-of-Words (BOW) based machine-learning, cannot handle, for example, cases of texts with same wording but different words order: "*The dangerous killer escaped one month ago, but recently he was arrested*" (RELIEF, HAPPYNESS) vs. "*The dangerous killer was arrested one month ago, but re-*

cently he escaped" (FEAR). The work in (Wang and Manning, 2012) partially accounts for this problem and argues that using word bigram features allows improving over BOW based methods, where words are taken as features in isolation. This way it is possible to capture simple compositional phenomena like polarity reversing in "*killing cancer*".

Finally, tasks such as copywriting, where evocative names are a key element to a successful product (Ozbal and Strapparava, 2012; Ozbal et al., 2012) require exhaustive lists of emotion related words. In such cases no context is given and the brand name alone, with its perceived prior polarity, is responsible for stating the area of competition and evoking semantic associations. For example *Mitsubishi* changed the name of one of its SUVs for the Spanish market, since the original name *Pajero* had a very negative prior polarity, as it means 'wanker' in Spanish (Piller, 2003). Evoking emotions is also fundamental for a successful name: consider names of a perfume like *Obsession*, or technological products like MacBook *air*.

In this work, we aim at automatically producing a high coverage and high precision emotion lexicon using distributional semantics, with numerical scores associated with each emotion, like it has already been done for sentiment analysis. To this end, we take advantage in an original way of massive crowd-sourced affective annotations associated with news articles, obtained by crawling the `rappler.com` social news network. We also evaluate our lexicon by integrating it in unsupervised classification and regression settings for emotion recognition. Results indicate that the use of our resource, even if automatically acquired, is highly beneficial in affective text recognition.

2 Related Work

Within the broad field of sentiment analysis, we hereby provide a short review of research efforts put towards building sentiment and emotion lexica, regardless of the approach in which such lists are then used (machine learning, rule based or deep learning). A general overview can be found in (Pang and Lee, 2008; Liu and Zhang, 2012; Wilson et al., 2004; Paltoglou et al., 2010).

Sentiment Lexica. In recent years there has been an increasing focus on producing lists of words (lexica) with prior polarities, to be used in sentiment analysis. When building such lists, a trade-off between coverage of the resource and its precision is to be found.

One of the most well-known resources is *SentiWordNet* (SWN) (Esuli and Sebastiani, 2006; Baccianella et al., 2010), in which each entry is associated with the numerical scores `Pos(s)` and `Neg(s)`, ranging from 0 to 1. These scores – automatically assigned starting from a bunch of seed terms – represent the positive and negative valence (or posterior polarity) of each entry, that takes the form `lemma#pos#sense-number`. Starting from SWN, several prior polarities for words (*SWN-prior*), in the form `lemma#PoS`, can be computed (e.g. considering only the first-sense, averaging on all the senses, etc.). These approaches, detailed in (Guerini et al., 2013), produce a list of 155k words, where the lower precision given by the automatic scoring of SWN is compensated by the high coverage.

Another widely used resource is *ANEW* (Bradley and Lang, 1999), providing valence scores for 1k words, which were manually assigned by several annotators. This resource has a low coverage, but the precision is maximized. Similarly, the *SO-CAL* entries (Taboada et al., 2011) were manually tagged by a small number of annotators with a multi-class label (from `very_negative` to `very_positive`). These ratings were further validated through crowd-sourcing, ending up with a list of roughly 4k words. More recently, a resource that replicated ANEW annotation approach using crowd-sourcing, was released (Warriner et al., 2013), providing sentiment scores for 14k words. Interestingly, this resource annotates the most frequent words in English, so, even if lexicon coverage is still far lower than SWN-prior, it grants a high coverage, with human precision, of language use.

Finally, the *General Inquirer* lexicon (Stone et al., 1966) provides a binary classification (`positive`/`negative`) of 4k sentiment-bearing words, while the resource in (Wilson et al., 2005) expands the General Inquirer to 6k words.

Emotion Lexica. Compared to sentiment lexica, far less emotion lexica have been produced, and all have lower coverage. One of the most used resources is *WordNetAffect* (Strapparava and Valitutti, 2004) which contains manually assigned affective labels to WordNet synsets (ANGER, JOY, FEAR, etc.). It currently provides 900 annotated synsets and 1.6k words in the form

	AFRAID	AMUSED	ANGRY	ANNOYED	DONT_CARE	HAPPY	INSPIRED	SAD
doc_10002	0.75	0.00	0.00	0.00	0.00	0.00	0.25	0.00
doc_10003	0.00	0.50	0.00	0.16	0.17	0.17	0.00	0.00
doc_10004	0.52	0.02	0.03	0.02	0.02	0.06	0.02	0.31
doc_10011	0.40	0.00	0.00	0.20	0.00	0.20	0.20	0.00
doc_10028	0.00	0.30	0.08	0.00	0.00	0.23	0.31	0.08

Table 1: An excerpt of the Document-by-Emotion Matrix - M_{DE}

`lemma#PoS#sense`, corresponding to roughly 1 thousand `lemma#PoS`.

AffectNet, part of the SenticNet project (Cambria and Hussain, 2012), contains 10k words (out of 23k entries) taken from ConceptNet and aligned with WordNetAffect. This resource extends WordNetAffect labels to concepts like 'have breakfast'. *Fuzzy Affect Lexicon* (Subasic and Huettner, 2001) contains roughly 4k `lemma#PoS` manually annotated by one linguist using 80 emotion labels. *EmoLex* (Mohammad and Turney, 2013) contains almost 10k lemmas annotated with an intensity label for each emotion using Mechanical Turk. Finally *Affect database* is an extension of SentiFul (Neviarouskaya et al., 2007) and contains 2.5K words in the form `lemma#PoS`. The latter is the only lexicon providing words annotated also with emotion scores rather than only with labels.

3 Dataset Collection

To build our emotion lexicon we harvested all the news articles from `rappler.com`, as of June 3rd 2013: the final dataset consists of 13.5 M words over 25.3 K documents, with an average of 530 words per document. For each document, along with the text we also harvested the information displayed by Rappler's *Mood Meter*, a small interface offering the readers the opportunity to click on the emotion that a given Rappler story made them feel. The idea behind the Mood Meter is actually "getting people to *crowdsource* the mood for the day"[1], and returning the percentage of votes for each emotion label for a given story. This way, hundreds of thousands votes have been collected since the launch of the service. In our novel approach to 'crowdsourcing', as compared to other NLP tasks that rely on tools like Amazon's Mechanical Turk (Snow et al., 2008), the subjects are aware of the 'implicit annotation task' but they are not paid. From this data, we built a document-by-emotion matrix M_{DE}, providing the voting percentages for each document in the eight affective dimensions available in Rappler. An excerpt is provided in Table 1.

The idea of using documents annotated with emotions is not new (Strapparava and Mihalcea, 2008; Mishne, 2005; Bellegarda, 2010), but these works had the limitation of providing a single emotion label per document, rather than a score for each emotion, and, moreover, the annotation was performed by the author of the document alone.

Table 2 reports the average percentage of votes for each emotion on the whole corpus: HAPPINESS has a far higher percentage of votes (at least three times). There are several possible explanations, out of the scope of the present paper, for this bias: (i) it is due to cultural characteristics of the audience (ii) the bias is in the dataset itself, being formed mainly by 'positive' news; (iii) it is a psychological phenomenon due to the fact that people tend to express more positive moods on social networks (Quercia et al., 2011; Vittengl and Holt, 1998; De Choudhury et al., 2012). In any case, the predominance of happy mood has been found in other datasets, for instance `LiveJournal.com` posts (Strapparava and Mihalcea, 2008). In the following section we will discuss how we handled this problem.

EMOTION	Votes$_\mu$	EMOTION	Votes$_\mu$
AFRAID	0.04	DONT_CARE	0.05
AMUSED	0.10	HAPPY	0.32
ANGRY	0.10	INSPIRED	0.10
ANNOYED	0.06	SAD	0.11

Table 2: Average percentages of votes.

4 Emotion Lexicon Creation

As a next step we built a word-by-emotion matrix starting from M_{DE} using an approach based on compositional semantics. To do so, we first lemmatized and PoS tagged all the documents (where PoS can be adj., nouns, verbs, adv.) and kept only those `lemma#PoS` present also in WordNet, similar to SWN-prior and WordNetAffect resources, to which we want to align. We then computed the term-by-document matrices using raw

[1] http://nie.mn/QuD17Z

Word	AFRAID	AMUSED	ANGRY	ANNOYED	DONT_CARE	HAPPY	INSPIRED	SAD
awe#n	0.08	0.12	0.04	0.11	0.07	0.15	**0.38**	0.05
comical#a	0.02	**0.51**	0.04	0.05	0.12	0.17	0.03	0.06
crime#n	0.11	0.10	**0.23**	0.15	0.07	0.09	0.09	0.15
criminal#a	0.12	0.10	**0.25**	0.14	0.10	0.11	0.07	0.11
dead#a	0.17	0.07	0.17	0.07	0.07	0.05	0.05	**0.35**
funny#a	0.04	**0.29**	0.04	0.11	0.16	0.13	0.15	0.08
future#n	0.09	0.12	0.09	0.12	0.13	0.13	**0.21**	0.10
game#n	0.06	0.15	0.06	0.08	0.15	**0.23**	0.15	0.12
kill#v	**0.23**	0.06	**0.21**	0.07	0.05	0.06	0.05	**0.27**
rapist#n	0.02	0.07	**0.46**	0.07	0.08	0.16	0.03	0.12
sad#a	0.06	0.12	0.09	0.14	0.13	0.07	0.15	**0.24**
warning#n	**0.44**	0.06	0.09	0.09	0.06	0.06	0.04	0.16

Table 3: An excerpt of the Word-by-Emotion Matrix (M_{WE}) using normalized frequencies (nf). Emotions weighting more than 20% in a word are highlighted for readability purposes.

frequencies, normalized frequencies, and tf-idf ($M_{WD,f}$, $M_{WD,nf}$ and $M_{WD,tfidf}$ respectively), so to test which of the three weights is better. After that, we applied matrix multiplication between the document-by-emotion and word-by-document matrices ($M_{DE} \cdot M_{WD}$) to obtain a (raw) word-by-emotion matrix M_{WE}. This method allows us to 'merge' words with emotions by summing the products of the weight of a word with the weight of the emotions in each document.

Finally, we transformed M_{WE} by first applying normalization column-wise (so to eliminate the over representation for happiness as discussed in Section 3) and then scaling the data row-wise so to sum up to one. An excerpt of the final Matrix M_{WE} is presented in Table 3, and it can be interpreted as a list of words with scores that represent how much weight a given word has in the affective dimensions we consider. So, for example, `awe#n` has a predominant weight in INSPIRED (0.38), `comical#a` has a predominant weight in AMUSED (0.51), while `kill#v` has a predominant weight in AFRAID, ANGRY and SAD (0.23, 0.21 and 0.27 respectively). This matrix, that we call `DepecheMood`[2], represents our emotion lexicon, it contains 37k entries and is freely available for research purposes at http://git.io/MqyoIg.

5 Experiments

To evaluate the performance we can obtain with our lexicon, we use the public dataset provided for the SemEval 2007 task on 'Affective Text' (Strapparava and Mihalcea, 2007). The task was focused on emotion recognition in one thousand news headlines, both in regression and classification settings. Headlines typically consist of a few words and are often written with the intention to 'provoke' emotions so to attract the readers' attention. An example of headline from the dataset is the following: "*Iraq car bombings kill 22 People, wound more than 60*". For the regression task the values provided are: `<anger (0.32), disgust (0.27), fear (0.84), joy (0.0), sadness (0.95), surprise (0.20)>` while for the classification task the labels provided are {`FEAR, SADNESS`}.

This dataset is of interest to us since the 'compositional' problem is less prominent given the simplified syntax of news headlines, containing, for example, fewer adverbs (like negations or intensifiers) than normal sentences (Turchi et al., 2012). Furthermore, this is to our knowledge the only dataset available providing numerical scores for emotions. Finally, this dataset was meant for unsupervised approaches (just a small trial sample was provided), so to avoid simple text categorization approaches.

As the affective dimensions present in the test set – based on the six basic emotions model (Ekman and Friesen, 1971) – do not exactly match with the ones provided by Rappler's Mood Meter, we first define a mapping between the two when possible, see Table 4. Then, we proceed to transform the test headlines to the `lemma#PoS` format.

SemEval	Rappler	SemEval	Rappler
FEAR	AFRAID	**SURPRISE**	**INSPIRED**
ANGER	ANGRY	-	ANNOYED
JOY	HAPPY	-	AMUSED
SADNESS	SAD	-	DON'T CARE

Table 4: Mapping of Rappler labels on Semeval2007. In bold, cases of suboptimal mapping.

Only one test headline contained exclusively words not present in `DepecheMood`, further indi-

[2] In French, 'depeche' means dispatch/news.

cating the high-coverage nature of our resource. In Table 5 we report the coverage of some Sentiment and Emotion Lexica of different sizes on the same dataset. Similar to Warriner et al. (2013), we observe that even if the number of entries of our lexicon is far lower than SWN-prior approaches, the fact that we extracted and annotated words from documents grants a high coverage of language use.

Sentiment Lexica	ANEW	1k entries	0.10
	Warriner et. al	13k entries	0.51
	SWN-prior	155k entries	**0.67**
Emotion Lexica	WNAffect	1k entries	0.12
	DepecheMood	37k entries	**0.64**

Table 5: Statistics on words coverage per headline.

Since our primary goal is to assess the quality of DepecheMood we first focus on the regression task. We do so by using a very naïve approach, similar to "WordNetAffect presence" discussed in (Strapparava and Mihalcea, 2008): for each headline, we simply compute a value, for any affective dimension, by averaging the corresponding affective scores –obtained from DepecheMood- of all lemma#PoS present in the headline.

In Table 6 we report the results obtained using the three versions of our resource (Pearson correlation), along with the best performance on each emotion of other systems[3] ($best_{se}$); the last column contains the upper bound of inter-annotator agreement. For all the 5 emotions we improve over the best performing systems (DISGUST has no alignment with our labels and was discarded).

Interestingly, even using a sub-optimal alignment for SURPRISE we still manage to outperform other systems. Considering the naïve approach we used, we can reasonably conclude that the quality and coverage of our resource are the reason of such results, and that adopting more complex approaches (i.e. compositionality) can possibly further improve performances in text-based emotion recognition.

As a final test, we evaluate our resource in the classification task. The naïve approach used in this case consists in mapping the average of the scores of all words in the headline to a binary decision with fixed threshold at 0.5 for each emotion (after min-max normalization on all test headlines scores). In Table 7 we report the results (F1 measure) of our approach along with the best performance of other systems on each emotion ($best_{se}$), as in the previous case. For 3 emotions out of 5 we improve over the best performing systems, for one emotion we obtain the same results, and for one emotion we do not outperform other systems. In this case the difference in performances among the various ways of representing the word-by-document matrix is more prominent: normalized frequencies (nf) provide the best results.

[3] Systems participating in the 'Affective Text' task plus the approaches in (Strapparava and Mihalcea, 2008). Other supervised approaches in the classification task (Mohammad, 2012; Bellegarda, 2010; Chaffar and Inkpen, 2011), reporting only overall performances, are not considered.

	$DepecheMood$			$best_{se}$	upper
	f	*nf*	*tfidf*		
FEAR	**0.56**	0.54	0.53	0.45	0.64
ANGER	0.36	**0.38**	0.36	0.32	0.50
SURPRISE*	**0.25**	0.21	0.24	0.16	0.36
JOY	0.39	**0.40**	0.39	0.26	0.60
SADNESS	**0.48**	0.47	0.46	0.41	0.68

Table 6: Regression results – Pearson's correlation

	$DepecheMood$			$best_{se}$
	f	*nf*	*tfidf*	
FEAR	0.25	**0.32**	0.31	0.23
ANGER	0.00	0.00	0.00	**0.17**
SURPRISE*	0.13	**0.16**	0.09	0.15
JOY	0.22	0.30	**0.32**	**0.32**
SADNESS	0.36	**0.40**	0.38	0.30

Table 7: Classification results – F1 measures

6 Conclusions

We presented DepecheMood, an emotion lexicon built in a novel and totally automated way by harvesting crowd-sourced affective annotation from a social news network. Our experimental results indicate high-coverage and high-precision of the lexicon, showing significant improvements over state-of-the-art unsupervised approaches even when using the resource with very naïve classification and regression strategies. We believe that the wealth of information provided by social media can be harnessed to build models and resources for emotion recognition from text, going a step beyond sentiment analysis. Our future work will include testing Singular Value Decomposition on the word-by-document matrices, allowing to propagate emotions values for a document to similar words non present in the document itself, and the study of perceived mood effects on virality indices and readers engagement by exploiting tweets, likes, reshares and comments.

This work has been partially supported by the Trento RISE PerTe project.

References

S. Baccianella, A. Esuli, and F. Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Conference on International Language Resources and Evaluation (LREC)*, pages 2200–2204, Valletta, Malta.

J. R. Bellegarda. 2010. Emotion analysis using latent affective folding and embedding. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pages 1–9. Association for Computational Linguistics.

M. Bradley and P. Lang. 1999. Affective norms for english words (ANEW): Instruction manual and affective ratings. *Technical Report C-1, University of Florida*.

E. Cambria and A. Hussain. 2012. *Sentic computing*. Springer.

S. Chaffar and D. Inkpen. 2011. Using a heterogeneous dataset for emotion analysis in text. In *Advances in Artificial Intelligence*, pages 62–67. Springer.

M. De Choudhury, S. Counts, and M. Gamon. 2012. Not all moods are created equal! exploring human emotional states in social media. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*.

P. Ekman and W. V. Friesen. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17:124–129.

A. Esuli and F. Sebastiani. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the Conference on International Language Resources and Evaluation (LREC)*, pages 417–422, Genova, IT.

M. Guerini, O. Stock, and C. Strapparava. 2008. Valentino: A tool for valence shifting of natural language texts. In *Proceedings of the Conference on International Language Resources and Evaluation (LREC)*, Marrakech, Morocco.

M. Guerini, L. Gatti, and M. Turchi. 2013. Sentiment analysis: How to derive prior polarities from sentiwordnet. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1259–1269.

D. Z. Inkpen, O. Feiguina, and G. Hirst. 2006. Generating more-positive and more-negative text. In *Computing Attitude and Affect in Text: Theory and Applications*, pages 187–198. Springer.

B. Liu and L. Zhang. 2012. A survey of opinion mining and sentiment analysis. *Mining Text Data*, pages 415–463.

G. Mishne. 2005. Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access*, volume 19.

S. M. Mohammad and P. D. Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

S. M. Mohammad. 2012. # Emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*Sem)*, pages 246–255. Association for Computational Linguistics.

A. Neviarouskaya, H. Prendinger, and M. Ishizuka. 2007. Textual affect sensing for sociable and expressive online communication. In A. Paiva, R. Prada, and R. Picard, editors, *Affective Computing and Intelligent Interaction*, volume 4738 of *Lecture Notes in Computer Science*, pages 218–229. Springer Berlin Heidelberg.

A. Neviarouskaya, H. Prendinger, and M. Ishizuka. 2009. Compositionality principle in recognition of fine-grained emotions from text. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*.

A. Neviarouskaya, H. Prendinger, and M. Ishizuka. 2011. Affect analysis model: novel rule-based approach to affect sensing from text. *Natural Language Engineering*, 17(1):95.

G. Ozbal and C. Strapparava. 2012. A computational approach to the automation of creative naming. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*.

G. Ozbal, C. Strapparava, and M. Guerini. 2012. Brand pitt: A corpus to explore the art of naming. In *Proceedings of the Conference on International Language Resources and Evaluation (LREC)*.

G. Paltoglou, M. Thelwall, and K. Buckley. 2010. Online textual communications annotated with grades of emotion strength. In *Proceedings of the 3rd International Workshop of Emotion: Corpora for research on Emotion and Affect*, pages 25–31.

B. Pang and L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

I. Piller. 2003. 10. advertising as a site of language contact. *Annual Review of Applied Linguistics*, 23:170–183.

D. Quercia, J. Ellis, L. Capra, and J. Crowcroft. 2011. In the mood for being influential on twitter. *Proceedings of IEEE SocialCom'11*.

R. Snow, B. O'Connor, D. Jurafsky, and A. Ng. 2008. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 254–263.

R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1631–1642.

P. Stone, D. Dunphy, and M. Smith. 1966. *The General Inquirer: A Computer Approach to Content Analysis.* MIT press.

C. Strapparava and R. Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 70–74. Association for Computational Linguistics.

C. Strapparava and R. Mihalcea. 2008. Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1556–1560. ACM.

C. Strapparava and A. Valitutti. 2004. WordNet-Affect: an affective extension of WordNet. In *Proceedings of the Conference on International Language Resources and Evaluation (LREC)*, pages 1083 – 1086, Lisbon, May.

P. Subasic and A. Huettner. 2001. Affect analysis of text using fuzzy semantic typing. *Fuzzy Systems, IEEE Transactions on*, 9(4):483–496.

M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.

M. Turchi, M. Atkinson, A. Wilcox, B. Crawley, S. Bucci, R. Steinberger, and E. Van der Goot. 2012. Onts: optima news translation system. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 25–30. Association for Computational Linguistics.

J. R. Vittengl and C. S. Holt. 1998. A time-series diary study of mood and social interaction. *Motivation and Emotion*, 22(3):255–275.

S. Wang and C. Manning. 2012. Baselines and bigrams: Simple, good sentiment and topic classification. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL).*

A. B. Warriner, V. Kuperman, and M. Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207.

S. Whitehead and L. Cavedon. 2010. Generating shifting sentiment for a conversational agent. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 89–97, Los Angeles, CA, June. Association for Computational Linguistics.

T. Wilson, J. Wiebe, and R. Hwa. 2004. Just how mad are you? finding strong and weak opinion clauses. In *Proceedings of AAAI*, pages 761–769.

T. Wilson, J. Wiebe, and P. 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*, pages 347–354.