

NRC Emotion Lexicon

SAIF M. MOHAMMAD AND PETER D. TURNEY
National Research Council Canada
Ottawa, Ontario, Canada, K1A 0R6
{saif.mohammad,peter.turney}@nrc-cnrc.gc.ca

November 15, 2013

1 Introduction

Words are associated with emotions. For example, *delightful* and *yummy* indicate the emotion of joy, *gloomy* and *cry* are indicative of sadness, *shout* and *boiling* are indicative of anger, and so on. In this paper, we present a large word–emotion association lexicon we created through a massive online annotation project.

A number of theories have been proposed about which emotions are basic (Ekman, 1992; Plutchik, 1962; Parrot, 2001; James, 1884). See Ortony and Turner (1990) for a detailed review of many of these models. Ekman (1992) argues that there are six basic emotions: joy, sadness, anger, fear, disgust, and surprise. Plutchik (1962, 1980, 1994) proposes eight basic emotions. These include Ekman’s six as well as trust and anticipation. Since annotating words with hundreds of emotions is expensive for us and difficult for annotators, we decided to annotate words with Plutchik’s eight basic emotions.

We created the word–emotion association lexicon through Amazon’s Mechanical Turk.¹ Mechanical Turk is an online crowdsourcing platform that is especially suited for tasks that can be done over the Internet through a computer or a mobile device. It is already being used to obtain human annotation on various linguistic tasks (Snow et al., 2008; Callison-Burch, 2009). However, one must define the task carefully to obtain annotations of high quality. Several checks must be included to ensure that random and erroneous annotations are discouraged, rejected, and re-annotated.

The details of our method are outlined in Mohammad and Turney (2010). We summarize them briefly below. We use *Roget’s Thesaurus* as the source for terms.² Since the 1911 US edition of *Roget’s* is in the public domain, we can distribute our emotion lexicon without the burden of restrictive licenses. We annotated only those words that occurred more than 120,000 times in the Google n-gram corpus.³

Roget’s Thesaurus groups related words into about a thousand categories, which can be thought of as coarse senses or concepts (Yarowsky, 1992). If a word is ambiguous, then it is listed in more than one category. Since a word may have different emotion associations when used in different senses, we obtained annotations at word-sense level by first asking an automatically generated word-choice question pertaining to the target:

- Q1. Which word is closest in meaning to *shark* (target)?
• *car* • *tree* • *fish* • *olive*

¹Mechanical Turk: <http://www.mturk.com/mturk/welcome>

²Roget’s Thesaurus: <http://www.gutenberg.org/ebooks/10681>

³The Google n-gram corpus is available through the Linguistic Data Consortium: <https://www.ldc.upenn.edu>

The near-synonym (*fish*) is taken from the thesaurus, and the distractors are randomly chosen words. This question guides the annotator to the desired sense of the target word. It is followed by ten questions asking if the target is associated with positive sentiment, negative sentiment, anger, fear, joy, sadness, disgust, surprise, trust, and anticipation.

If an annotator answers Q1 incorrectly, then we discard information obtained from the remaining questions. Thus, even though we do not have objectively correct answers to the emotion association questions, many of the dubious annotations are likely to be eliminated. About 10% of the annotations were discarded because of an incorrect response to Q1.

Each term is annotated by five different people. For 74.4% of the instances, all five annotators agreed on whether a term is associated with a particular emotion or not. For 16.9% of the instances, four out of five people agreed with each other. The information from multiple annotators for a particular term is combined by taking the majority vote. The lexicon has entries for about 24,200 word-sense pairs. The information from different senses of a word is combined by taking the union of all emotions associated with the different senses of the word. This resulted in a word-level emotion association lexicon for about 14,200 word types. These files are together referred to as the *NRC Emotion Lexicon version 0.92*.

2 Applications

Automatic recognition of emotions is useful for a number of tasks, including the following:

1. Managing customer relations by taking appropriate actions depending on the customer's emotional state (for example, dissatisfaction, satisfaction, sadness, trust, anticipation, or anger) (Bougie, Pieters, and Zeelenberg, 2003).
2. Tracking sentiment towards politicians, movies, products, countries, and other target entities (Pang and Lee, 2008; Mohammad and Yang, 2011).
3. Developing sophisticated search algorithms that distinguish between different emotions associated with a product (Knautz, Siebenlist, and Stock, 2010). For example, customers may search for banks, mutual funds, or stocks that people trust. Aid organizations may search for events and stories that are generating empathy, and highlight them in their fund-raising campaigns. Further, systems that are not emotion-discerning may fall prey to abuse. For example, it was recently discovered that an online vendor deliberately mistreated his customers because the negative online reviews translated to higher rankings on Google searches.⁴
4. Creating dialogue systems that respond appropriately to different emotional states of the user; for example, in emotion-aware games (Velásquez, 1997; Ravaja et al., 2006).
5. Developing intelligent tutoring systems that manage the emotional state of the learner for more effective learning. There is some support for the hypothesis that students learn better and faster when they are in a positive emotional state (Litman and Forbes-Riley, 2004).
6. Determining risk of repeat attempts by analyzing suicide notes (Osgood and Walker, 1959; Matykiewicz, Duch, and Pestian, 2009; Pestian, Matykiewicz, and Grupp-Phelan, 2008).⁵
7. Understanding gender differences in communication styles through work-place and personal email (Mohammad and Yang, 2011).
8. Assisting in writing e-mails, documents, and other text to convey the desired emotion (and avoid misinterpretation) (Liu, Lieberman, and Selker, 2003).
9. Depicting the flow of emotions in novels and other books (Boucoulalas, 2002; Mohammad, 2011).
10. Identifying what emotion a newspaper headline is trying to evoke (Bellegarda, 2010).

⁴http://www.pcworld.com/article/212223/google_algorithm_will_punish_bad_businesses.html

⁵The 2011 Informatics for Integrating Biology and the Bedside (i2b2) challenge by the National Center for Biomedical Computing is on detecting emotions in suicide notes.

11. Re-ranking and categorizing information/answers in online question–answer forums (Adamic et al., 2008). For example, highly emotional responses may be ranked lower.
12. Detecting how people use emotion-bearing-words and metaphors to persuade and coerce others (for example, in propaganda) (Kövecses, 2003).
13. Developing more natural text-to-speech systems (Francisco and Gervás, 2006; Bellegarda, 2010).
14. Developing assistive robots that are sensitive to human emotions (Breazeal and Brooks, 2004; Hollinger et al., 2006). For example, the robotics group in Carnegie Melon University is interested in building an emotion-aware physiotherapy coach robot.

3 Related Work

This section describes some of the other existing word–emotion association lexicons. The WordNet Affect Lexicon (WAL) (Strapparava and Valitutti, 2004) has a few hundred words annotated with a number of affect categories including the six Ekman emotions (joy, sadness, anger, fear, disgust, and surprise).⁶ General Inquirer (GI) (Stone et al., 1966) has 11,788 words labeled with 182 categories of word tags, including positive and negative polarity.⁷ Affective Norms for English Words (ANEW) has pleasure (happy–unhappy), arousal (excited–calm), and dominance (controlled–in control) ratings for 1034 words.⁸

4 NRC Emotion Lexicon: version 0.92

The National Research Council of Canada reserves all rights to the NRC Emotion Lexicon.

4.1 Terms of use

1. If you use this lexicon in your research, then please cite this paper and/or any of the following papers: Mohammad and Turney (2013, 2010); Mohammad and Yang (2011); Mohammad (2011).⁹ Below is how this paper can be cited:

NRC Emotion Lexicon, Saif M. Mohammad and Peter D. Turney, *NRC Technical Report*, December 2013, Ottawa, Canada.

Links to the papers are available at <http://www.purl.org/net/NRCemotionlexicon>.

2. If you use this lexicon in an application, please credit NRC appropriately. For example a web application making use of the NRC Emotion Lexicon should acknowledge the use the lexicon in the home page.

4.2 Format

The appendix shows the list of terms in the NRC Emotion Lexicon and their association with each of the ten affect categories studied—the eight basic emotions and positive and negative sentiment. Each line has the following format:

TargetWord AffectCategory:AssociationFlag AffectCategory:AssociationFlag ...
where,

- *TargetWord* is a word for which emotion associations are provided.
- *AffectCategory* is one of eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, or disgust) or one of two polarities (negative or positive).

⁶WAL: <http://wndomains.fbk.eu/wnaffect.html>

⁷GI: <http://www.wjh.harvard.edu/~inquirer>

⁸ANEW: <http://csea.php.ufl.edu/media/anewmessage.html>

⁹Our recommended citation is Mohammad and Turney (2013).

– *AssociationFlag* has one of two possible values: 0 or 1. 0 indicates that the target word has no association with affect category, whereas 1 indicates an association.

The Appendix presents the word-level emotion lexicon. If you wish to obtain the sense-level lexicon or you would like more detailed information such as the individual annotations by each of the annotators, then contact the National Research Council Canada.

References

- Adamic, L. A.; Zhang, J.; Bakshy, E.; and Ackerman, M. S. 2008. Knowledge sharing and yahoo answers: everyone knows something. In *Proceeding of the 17th international conference on World Wide Web, WWW '08*, 665–674. New York, NY, USA: ACM.
- Bellegarda, J. 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*.
- Boucouvalas, A. C. 2002. Real time text-to-emotion engine for expressive internet communication. *Emerging Communication: Studies on New Technologies and Practices in Communication* 5:305–318.
- Bougie, J. R. G.; Pieters, R.; and Zeelenberg, M. 2003. Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services. Open access publications from tilburg university, Tilburg University.
- Breazeal, C., and Brooks, R. 2004. Robot emotions: A functional perspective. In *Who Needs Emotions*. Oxford University Press.
- Callison-Burch, C. 2009. Fast, cheap and creative: Evaluating translation quality using amazon's mechanical turk. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, 286–295.
- Ekman, P. 1992. An Argument for Basic Emotions. *Cognition and Emotion* 6(3):169–200.
- Francisco, V., and Gervás, P. 2006. Automated mark up of affective information in english texts. In Sojka, P.; Kopecek, I.; and Pala, K., eds., *Text, Speech and Dialogue*, volume 4188 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. 375–382.
- Hollinger, G.; Georgiev, Y.; Manfredi, A.; Maxwell, B. A.; Pezzementi, Z. A.; and Mitchell, B. 2006. Design of a social mobile robot using emotion-based decision mechanisms. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 3093–3098.
- James, W. 1884. What is an emotion? *Mind* 9:188–205.
- Knautz, K.; Siebenlist, T.; and Stock, W. G. 2010. Memose: search engine for emotions in multimedia documents. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval, SIGIR '10*, 791–792. New York, NY: ACM.
- Kövecses, Z. 2003. *Metaphor and Emotion: Language, Culture, and Body in Human Feeling (Studies in Emotion and Social Interaction)*. Cambridge University Press.
- Litman, D. J., and Forbes-Riley, K. 2004. Predicting student emotions in computer-human tutoring dialogues. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, ACL '04*. Morristown, NJ, USA: Association for Computational Linguistics.
- Liu, H.; Lieberman, H.; and Selker, T. 2003. A model of textual affect sensing using real-world knowledge. In *Proceedings of the 8th international conference on Intelligent user interfaces, IUI '03*, 125–132. New York, NY: ACM.

- Matykiewicz, P.; Duch, W.; and Pestian, J. 2009. Clustering semantic spaces of suicide notes and newsgroups articles. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing, BioNLP '09*, 179–184. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Mohammad, S. M., and Turney, P. D. 2010. Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*.
- Mohammad, S. M., and Turney, P. D. 2013. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence* 29(3):436–465.
- Mohammad, S., and Yang, T. 2011. Tracking Sentiment in Mail: How Genders Differ on Emotional Axes. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, 70–79. Portland, Oregon: Association for Computational Linguistics.
- Mohammad, S. 2011. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 105–114. Portland, OR, USA: Association for Computational Linguistics.
- Ortony, A., and Turner, T. J. 1990. What’s basic about basic emotions? *Psychological Review* 97:315–331.
- Osgood, C. E., and Walker, E. G. 1959. Motivation and language behavior: A content analysis of suicide notes. *Journal of Abnormal and Social Psychology* 59(1):58–67.
- Pang, B., and Lee, L. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2(1–2):1–135.
- Parrot, W. 2001. *Emotions in Social Psychology*. Psychology Press.
- Pestian, J. P.; Matykiewicz, P.; and Grupp-Phelan, J. 2008. Using natural language processing to classify suicide notes. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing, BioNLP '08*, 96–97. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Plutchik, R. 1962. *The Emotions*. Random House.
- Plutchik, R. 1980. A General Psychoevolutionary Theory of Emotion. *Emotion: Theory, research, and experience* 1(3):3–33.
- Plutchik, R. 1994. *The psychology and biology of emotion*. New York: Harper Collins.
- Ravaja, N.; Saari, T.; Turpeinen, M.; Laarni, J.; Salminen, M.; and Kivikangas, M. 2006. Spatial presence and emotions during video game playing: Does it matter with whom you play? *Presence: Teleoperators and Virtual Environments* 15(4):381–392.
- Snow, R.; O’Connor, B.; Jurafsky, D.; and Ng, A. 2008. Cheap and fast - but is it good? Evaluating nonexpert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2008)*, 254–263.
- Stone, P.; Dunphy, D. C.; Smith, M. S.; Ogilvie, D. M.; and associates. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.
- Strapparava, C., and Valitutti, A. 2004. WordNet-Affect: An Affective Extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-2004)*, 1083–1086.
- Velásquez, J. D. 1997. Modeling emotions and other motivations in synthetic agents. In *Proceedings of the fourteenth national conference on artificial intelligence and ninth conference on Innovative applications of artificial intelligence, AAAI’97/IAAI’97*, 10–15. AAAI Press.
- Yarowsky, D. 1992. Word-sense disambiguation using statistical models of Roget’s categories trained on large corpora. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, 454–460.

