Institute of Informatics
Logics and Security Studies

Automatic Expansion of Affective Norms of Words Lexicon

Samira Shaikh, Kit Cho, John Lien, Tomek Strzalkowski, Ting Liu, Laurie Feldman, George Aaron Broadwell, Boris Yamrom

January, 2015

ILS Tech Report: 16

Abstract

The present study describes an automated method for expanding the Affective Norms of Words (ANEW) lexicon by conferring valence ratings over the synonyms of existing words, as identified in Wordnet. We follow the method described in Liu et al. (2014) to derive our expansion from an existing English lexicon and also extend the method to derive valence scores of words in language other than English. The result is expanded lexicons with valence scores in four different languages, which can be used across several research strands where valence scores of words may be required.

ANEW Lexicon

The ANEW lexicon (Bradley and Lang, 1999) was developed to provide a set of normative emotional ratings for a large number of words in the English language. Subjects rated words on the dimensions of pleasure, arousal, and dominance. This lexicon by Bradley and Lang consisted of 2466 human-rated words. In the Metaphor project, we made use of this lexicon for automated determination of affect of metaphorical expressions, using a novel method called Affect calculus (Strzalkowski et al. 2013). The number of words in the existing human-rated lexicon is not sufficient to cover the extent of natural language expressions typically encountered when working with large amounts of text. To overcome these issues, we sought to first expand the existing English language lexicon to provide more coverage. An additional goal was also to use the expanded lexicon to then derive similar lexicons for other languages we were interested in – Spanish, Russian and Farsi.

Expansion Method

Our expansion method follows that adopted by Liu et al. (2014) for their automated expansion of the MRC psycholinguistic database. We use WordNet (Miller, 1995), a large English lexical database with over 150,000 words, hierarchically organized in synsets that capture semantically equivalent words. It is thus reasonable to assume that if one element of a synset has a known valence score, all other words in this synset should have the same or closely related scores, and can be added to the expanded lexicon with the inherited valence ratings. It is important to note that a word can have multiple senses defined in Wordnet and some of them may be less commonly used. It may not be appropriate to assign the same imageability ratings to all synsets of the word, because the valence ratings by humans in the original lexicons are assumed to represent words in their most common usage, which is typically captured by the first synset. Consequently, we only utilize the top-ranked (most frequent) sense for expansion.

While doing the expansion, we keep track of the original words and the words that come from WordNet synoynms of the original word. Each word added to this set will be the synonym of one or more original words. When a synonym is added to the set, we simply keep track of a list of words it originally derived from. Keeping track of the original words is later used in the procedure for calculating the expansion scores. Once our expansion set is completely populated, we can begin calculating the affect scores for each of the newly discovered terms. We do this by taking the average of the words mapped to their newly expanded terms. By doing this, we can see that if the expanded term is a synonym of only one word, then it simply inherits the score of

the word from which it was derived. After finishing the expansion, the original words, the expanded words, and their corresponded scores are consolidated into a single list.

For extending the lexicons to different languages, such as Spanish, Russian and Farsi, we make use of the Google Translate API and attempt to translate each source word from English to the corresponding language. When a translation is returned via the Google Translate API, we add the translated word to the corresponding language lexicon and assign the valence score from the original English word to the translated word.

The automatic expansion and translation method has resulted in significant expansion of the coverage in each language we have worked with. The details of the number of words are described in the next section.

Details of Lexicons

	English	Spanish	Russian	Farsi
ANEW+	22,671 words	13,154 words	16,485 words	14,448 words

In the table above, we provide the number of words currently in the lexicons we have developed. Each word in the lexicons is accompanied by a part of speech tag and a valence score.

The valence scores follow the same range as the original ANEW lexicon, with 1 being the lowest (most negative) and 9 being the highest (most positive). This scale of ratings allows us to experiment with positive and negative ranges as well as ranges of neutral values.

Validation of Expansion and Translation

Our goal is to do systematic validation of these resources that we have automatically developed. We plan to conduct several validation experiments using human subjects to test the validity and robustness of our expansion and translation methods. These experiments are being planned and will be the basis of a separate publication.

Acknowledgments

References

Bradley, M.M., & Lang, P.J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.

Liu, Ting, Kit Cho, George Aaron Broadwell, Samira Shaikh, Tomek Strzalkowski, John Lien, Sarah Taylor, Laurie Feldman, Boris Yamrom, Nick Webb, Umit Boz, Ignacio Cases, Ching-sheng Lin. (2014) Automatic Expansion of the MRC Psycholinguistic Database Imageability Ratings. In Proceedings of the *Ninth International Conference on Language Resources and Evaluation (LREC'14)*.

Tomek Strzalkowski, Samira Shaikh, Kit Cho, George Aaron Broadwell, Laurie Feldman, Sarah Taylor, Boris Yamrom, Ting Liu, Ignacio Cases, Yuliya Peshkova, Kyle Elliot. (2014). Computing Affect in Metaphors. In Proceedings of Second Workshop on NLP in Metaphor, at the 2014 Association of Computational Linguistics, Baltimore, MD.

Warriner, A.B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior Research Methods, 45, 1191-1207.