

Carolyn Lamb

CISC 871 final project

0549-8486

Using WordNet to Replicate an Affective Word List

For my final project I decided to build a program that would perform a limited version of the affect analysis done by Subsic & Huettner (2001). Rather than build word lists for a large number of different emotions, I chose a single emotion: insanity. After finding and compiling insanity-related words on WordNet, I wrote a Python script to locate these words in a text file and thus compute the text's centrality score with regards to insanity.

This document explains how to use the program and describes some informal experimental results. It also describes my experience with the process of finding words on WordNet for this purpose, which turned out to be much trickier than it appeared.

1. Background

Subasic & Huettner's affect analysis system treats each affect category - happiness, fear, love, and so on - as a fuzzy set. A text or word has a membership in each affect set. These memberships are each split into two separate measures ranging

from 0 to 1 - centrality, a measure of the directness or extent of relatedness of the word to the category; and intensity, a measure of the strength or weakness of the emotion in the affect category.

The essential component of Subasic & Huettner's system is their word list, which contains the centrality and intensity scores of each word in each affect category. Each text is then given a centrality and intensity score based on those of the words in the text. The centrality of a text is equivalent to the centrality of the most central word in the text, while the intensity of the text is calculated taking into account all of the words in the text with intensity scores of greater than 0.

Subasic & Huettner produced their word list by taking an already existing word list by another researcher and supplementing it "rather haphazardly" with words from an online thesaurus. However, they recognized that this method was not scalable or transferable, and planned to experiment with creating word lists from WordNet in the future. Other researchers, such as Andreevskaia & Bergler (2006), have also created word lists using WordNet; Andreevskaia & Bergler did this by starting with a small base list of words and searching for their synonyms and hyponyms. The resulting list still needed to be filtered and to have contradictions removed (e.g. words which were listed as both positive and negative).

2. The Program

`insanity.py` is a simple Python script meant to be run from the IDLE Python interpreter. When the `main()` method is run, the script prompts the user for the name of a text file. The script then assigns the text a centrality score. It lists the insanity-related words that were found in the text and their centrality scores.

Like in Subasic & Huettner, the centrality score of a text is computed by taking the highest centrality score of any individual word in the text. Centrality is separate from intensity (which is not measured here); rather than a measure of "how insane" the text is, it should be taken as a measure of how directly the text deals with insanity.

Centrality scores are assigned as "high", "medium", and "weak" based on the informal classification that Subasic & Huettner use in one of their own examples.

Included with `insanity.py` are ten sample texts (explained in more detail in part 5), but you can play around with the program and use any text you like. It may be slower with very long texts, since its time complexity is $O(n*m)$, where n is the number of words in the text and m is the length of the list of insanity-related words.

3. The Word List

Given the interest in constructing word lists using WordNet, I decided that I would use WordNet to build my own word list. However, misjudging the scale of the task by a couple of orders of magnitude, I determined that since I was only looking at the words related to a single concept, I could write down the words manually instead of writing a harvesting program. This is why the project took so long.

The general idea, in creating the word list, was to start with a pair of seed words - in this case, "insanity" and "insane". These seed words would each have a centrality of 1. Each word which was related to the seed words would then be added, and then each word which was related to those words, and so on. However, each relation would have a reduction factor: the more relations it took to navigate from "insanity" or "insane" to the word in question, the lower its centrality score would be. While Andreevskaja & Bergler suggest other methods for calculating centrality - such as the number of relations each word has to other words in the list - this one was chosen as it seemed relatively straightforward to implement.

Andreevskaja & Bergler's program looked only at synonyms and hyponyms. However, WordNet offers a variety of different

relations between words besides these, and synonymy and hyponymy alone are not sufficient to connect the words "insanity" and "insane" with all the words that would seem to relate to them. For example, most names of specific mental illnesses cannot be reached, using only synonyms and hyponyms, from "insanity" or "insane". Therefore, I broadened my horizons and took into account other relations between words.

I made two major attempts at forming a word list. I felt the need to end the first one prematurely because of a couple of unforeseen problems. For one thing, words kept pouring in which seemed completely unrelated to the concept of insanity. I blamed this on the use of hypernyms, derivationally related forms, and homonyms, many of which caused rapid branching out into other topics that did not seem related to insanity. More worrying still were the problems caused by my choice of reduction factors. In the old word list, synonyms and some derivationally related forms had a reduction factor of 0, meaning a word would have the same centrality score as all its synonyms. While this makes intuitive sense, it resulted in sections of the word list that expanded while they were being worked on. Working on a word at a given centrality would result in many synonyms and derivationally related forms being added to the list of words at that centrality, and then the synonyms of the derivationally related forms, and then the derivationally related forms of the synonyms

of the derivationally related forms, and so on. I realized there was no guarantee that this process would always terminate. So after listing all the words with centralities ranging from 1.0 to 0.3, I abandoned this word list (attached as words-old.csv).

In my second attempt (attached as words.csv) I tried to address the shortcomings of the first one. Every relation, even synonymy, had a reduction factor, and hypernyms and homonyms were excluded altogether. However, a few other relations were given more attention. For example, the relation "domain term category" produced an abundance of words relating to mental disorders, so it was given a smaller reduction factor, as were some other relations. The new word list also included multi-word phrases (and thus, their own hyponyms), which I had previously avoided.

Strangely, words-old.csv is not smaller than words.csv, even taking into account that words-old was ended prematurely. Words-old contains 743 words with scores from 0.3 to 1. Words contains 2773 words, 812 of which have scores from 0.3 to 1 (the others range from 0.05 to 0.25).

Both attempts appeared unsatisfactory while I was working on them. They included many unrelated words while excluding ones that ought to have been related. For instance, neither list contained words like "madhouse" or "asylum". In the second list, many mental illnesses appeared only at very low centrality levels, mixed in with hordes of seemingly unrelated words.

4. Remarks about the word lists

Manually creating both word lists was probably a poor choice in terms of how to use my time, but it did produce a great familiarity with some of the issues involved in constructing such lists, which might be glossed over during automatic retrieval.

Part of the problem is that not all relations of a given type, on WordNet, are of the same relevance. For example, the "derivationally related form" relation can produce strong and obvious connections - like "mad" to "madness" - but many derivationally related forms veer off into another domain entirely. For instance, "delirious" meaning "excited" is derivationally related to "delirious" meaning "experiencing delirium" - which leads to the addition of all sorts of other words meaning "excited". Apart from cutting out such relations altogether, there is no obvious way of distinguishing between these relations without direct human judgement. A few other relations, such as "similar to", behaved in the same way.

Especially at low centrality scores, these relations produce a jumble of words relating to different mental states, particularly states involving strong emotions, rather than words which are obviously related to insanity. This is actually quite similar to an effect that Subasic & Huettnner produced

deliberately - the "fuzzy thesaurus", in which some affect categories are given fuzzy membership in other affect categories. For example, if the "fear" category has a membership of 0.5 in the "insanity" category, then we would expect words with a membership in "fear" to also have at least a small membership in "insanity". This is exactly what occurs with my word lists, though through a different process. For example, in words.csv the word "crazy" (meaning "insane") is derivationally related to the word "crazy" (meaning "foolish"), which leads to "craziness" and then to words like "foolery" and "japery". Through "craziness", more or less the entire "humor" fuzzy set has been incorporated into these word lists.

Using WordNet relations with reduction factors instead of using a fuzzy thesaurus results in a similar inclusion of sets within other sets, but this inclusion occurs in a different and arguably more nuanced manner. With Subasic & Huettner's fuzzy thesaurus, assuming there was a connection between humor and insanity, a word's centrality to insanity would be equal to its centrality to humor or the fuzzy membership of humor in insanity, whichever was lower (unless, of course, the word had a higher insanity centrality score before the humor fuzzy set was included). But in words.csv, some types of humor seem to be more central to insanity than others. The humor words with the highest centrality are those which relate to foolishness and bizarre

actions, which are arguably traits associated with the insane. Words referring to clever and controlled humor, such as "wit" and "satire", receive much lower scores. A fuzzy thesaurus on its own would not produce such finely graded distinctions.

Regardless of the relations used, these word lists tend to expand exponentially as they move to lower centralities. The second word list, for example, contains 318 words at 0.5 or above, 696 from 0.25 to 0.45, and 1759 at 0.2 or below.

5. Testing the word lists

After spending considerable time on the second word list, I hoped to show experimentally that it performed better than the first. Unfortunately, this is not quite what occurred.

For testing, I used the ten text files attached, as follows:

1. An excerpt from the Wikipedia article on insanity. This is the most central text I could think of.
2. An excerpt from an article on bipolar disorder in a psychiatric journal. This deals directly with mental illness and should receive a high centrality score.
3. A summary of a Victorian law regarding insane persons. This should receive a high score as well.
4. An excerpt from an essay by Amanda Gannon on her experiences with bipolar disorder. This was chosen because it

uses more poetic and indirect language to talk about mental illness, but to a human reader, the topic of the excerpt should still be clear. Ideally this would get a moderately high score.

5. A blog post by a competitive cyclist describing a race in which he felt very angry. Because of the strong emotions involved, one might expect this text to get a moderate score, but it should not be very high, since there is nothing actually abnormal about these emotions.

6. An excerpt from an article about Steve Jobs. The excerpt was chosen because it describes some of Jobs' more unpleasant personality traits. Like 5, this text might receive a moderate score, but the traits it describes are merely unpleasant, not actually insane, so the score should not be particularly high.

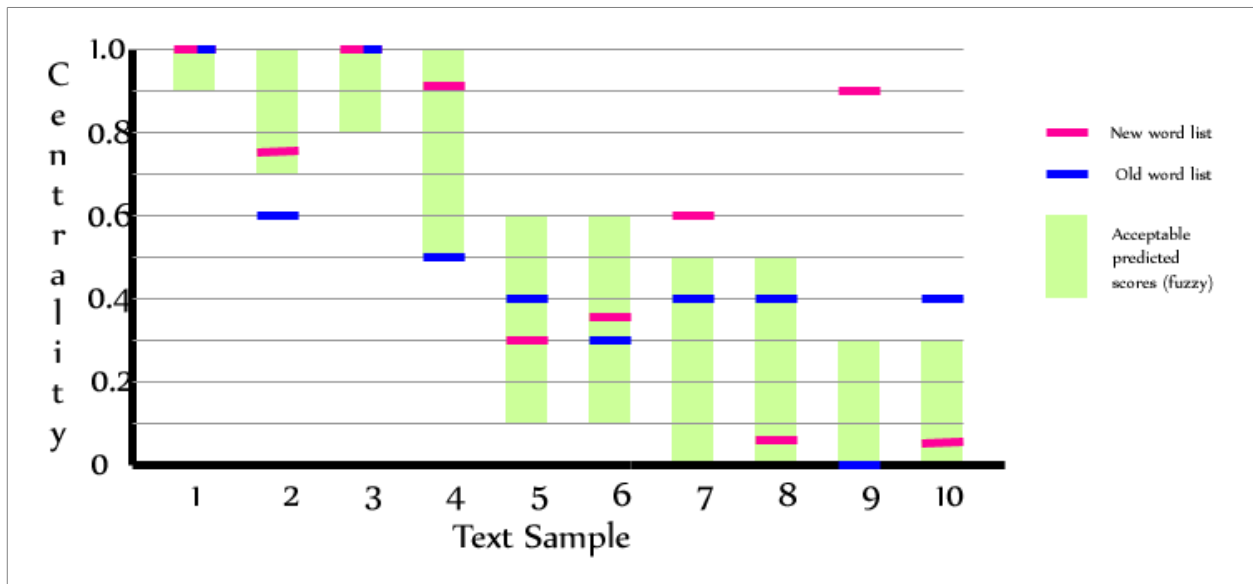
7. A description from a professional clown's web site. This was chosen because of specific weaknesses in the word lists, both of which incorporated many words describing humor. While a moderate score for this excerpt might be justified using the "fuzzy thesaurus" logic described in part 3, it should not be as high as the texts that actually describe mental illness.

8. Several descriptions of puzzles from an online catalog. Like 7, this was chosen because of a specific weakness in words-old.csv, which contains many words relating to puzzles and riddles. It should not have a particularly high score.

9 and 10. Two articles about matters related to engineering

careers from IEEE publications. These articles were chosen as a “control group”: they contain nothing significant that relates either to mental illness or to any other notable mental state, and they should receive low or nonexistent centrality scores.

The following chart shows how each word list rates each text:



Since there is no obvious ground truth with regards to the centrality of each text, the green areas of the chart show acceptable range of scores for each text - that is, scores at which the program’s rating of the text would make intuitive sense. Both the word lists perform within this area for most texts, but both diverge at times. Words-old.csv, which does not contain many key phrases like “mental illness” and “bipolar”, underestimates the centrality of Text 2, and the words.csv overestimates the centrality of Text 7 (giving too much weight to

words like "excitement" and "clown").

Each word list performs badly on one of the "control" texts due to poor handling of polysemy: the existence of words with more than one separate meaning. The old word list overreacts to the presence of the word "creep" in Text 10 - a word which is on the list with a moderate centrality score because it can refer to a person whose behavior is unusual and unsettling. However, in Text 10, "creep" is used as a verb to describe rising salaries. The new word list similarly overreacts to the presence of the word "certified" in Text 9. "Certified" has a high centrality score because it can mean "certified as insane" - but in Text 9, the word refers to professional certification for an engineer.

Subasic & Huettnner claim that their fuzzy affect categories deal well with polysemy by representing all possible meanings of the word, and that by taking into account all words in a document, their methods will give a nuanced picture of the overall meaning of a text. But Subasic & Huettnner's own method is to take the centrality of the single most central word in a text to represent the centrality of the entire text. This method is easily swayed by a single word with a very central meaning, like "certified", and there is no mechanism for moderating the effect of a single such word.

One crude but instructive compensatory mechanism, at least for short texts like the ones used as examples, would be to

remove the single most central word from consideration - either in all cases, or in cases where the most central word has a much higher score than the second most central. This would reduce Text 9's centrality, using words.csv, to 0.45 (because of the moderately central word "thrilled"). However, it would also reduce the centrality of Text 4 to 0.4, as Text 4's high score depends largely on the single word "wild"; the less ambiguous words in the text, like "disorder" and "depression", have much lower centrality. While Text 4 does use the word "wild" in a way that refers to uncontrollable, mentally unstable behavior, other texts (such as nature articles) would use this word in ways not related to insanity at all. As Text 4 and 9's scores show, taking the other words in the text into account will not necessarily resolve this problem.

Thus, while both word lists usually approximate the desired result, both are unreliable, and neither consistently outperforms the other. Part of this is attributable to problems with both of the word lists, which both include many unrelated words while excluding ones that ought to have been related. Another part is attributable to the problem of polysemy, which the method of fuzzy sets as used in Subasic & Huettner's paper does not actually resolve.

6. How do we improve these word lists?

My experiences trying to make a "better" word list show that these intuitive attempts at improvement will not necessarily be successful. However, with computer programs that automatically trawl WordNet for related words, many more attempts can be made in less time. The human creator of an automatically generated word list may not be as aware of the details of the list, its failings, and the relations involved with these failings, but they will also be able to produce the lists much more quickly and to potentially improve them through trial and error.

Thus, one way to build a better word list might be to mix careful human judgement with the repeated use of an automatic program. A user might keep track of the relations used to get from the central topic of "insanity" to other highly related topics (for example, specific mental illnesses, or symptoms such as hallucinations) and assign a low reduction in centrality to those relations while penalizing others. This could help pare down the word list and bring the user's preferred words to the forefront.

Alternately, one could abandon WordNet and experiment with other word list construction methods. For example, one might get good results with a statistical analysis: pre-judging texts as being highly central or unrelated and then working out which words are used more often in the central texts.

Neither of these methods really addresses the issue of polysemy. Without some method of guessing the likelihood of different meanings of a word being used, problems like the use of the word "certified" will continue to crop up. However, a statistical analysis-based word list might actually mitigate this problem by downplaying the importance of highly ambiguous words. If "certified" is widely used in ways that have nothing to do with insanity, then it will presumably not be used more often in insanity-related texts than unrelated texts, and thus, a statistical analysis will not pick it up. Texts which do use the word "certified" to mean "certified as insane" will presumably use many other words which are strongly and unambiguously related to insanity.

7. References

- Andreevskaja, A. & Bergler, S., "Mining WordNet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses", In Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, IT., 2006
- Subasic, P. & Huettner, A., "Affect analysis of text using fuzzy semantic typing", IEEE Transactions on Fuzzy Systems, Vol.

9, No. 4, August 2001, pp. 483-496