Word Associations and Sentiment Analysis

Michael Lee
mdlee@uci.edu
Department of Cognitive Sciences,
University of California, Irvine

with thanks to Simon De Deyne, Jenny Shi and Simon Dennis
Brown uses vetoes to press for Prop. 30

The governor scolds legislators for needless projects and risks alienating fellow Democrats to show that Sacramento would spend tax money wisely.

buckre at 3:10 PM October 2, 2012
Gov Brown there is no way I’m going to let you tapdance your way into my wallet this year. We need real accountability and oversight in government. No more flinging money around at departments without knowing where in the world it went or how much was spent. There is no way any of these tax measures are passing this year.

condondeb at 2:09 PM October 2, 2012
To: Novotopia,

Good gosh! Are you blind? Do you read the news, follow the antics of our legislature? You cannot be that far out to lunch! This is not a question of Liberal versus Conservative. Governor Brown has a Democratic Majority in the California Legislature, which begs the question: What has Governor Brown and his legislature done to tangibly reduce spending at the State Level? Please toss a couple of legislative bills at me that honestly make an effort to gain control of State expenditures and pay outstanding bills. The heck with Schwartznegger, he is HISTORY, gonzo, out of office the current Governor is in power and is playing "Politics" with the citizens of California. Open your eyes and look around you.
Another boneheaded #Marlins trade 😐 RT @WSJ: Breaking: Tigers' Cabrera wins baseball's first Triple Crown in 45 years
on.wsj.com/QtxgKt

Great trade Marlins! RT @WPLGLocal10: Former #Marlin Miguel Cabrera becomes baseball's first triple crown winner in 45 years
Speculative Idea

- It would be useful to be able to gauge automatically (even just approximately) the sentiment of the comments
  - The valences of the words people use probably give some decent indication of sentiment
  - But when we read descriptions (or generate them for reading to express sentiment) we are “primed” also to think about the associations
    - So maybe the valences of the associated words also play a role in assessing sentiment

- Test this using valence databases, the Leuven word association database, and some representative descriptions where we have a ground truth
Looper (2012)

TOMATOMETER ©

94%

As thought-provoking as it is thrilling, Looper delivers an uncommonly smart, bravely original blend of futuristic sci-fi and good old-fashioned action.

AUDIENCE ©

90%

liked it

Average Rating: 4.3/5
User Ratings: 48,310

Critic Reviews for Looper

All Critics (204) | Top Critics (41) | Fresh (191) | Rotten (13)

The reasoning behind all this may not reward prolonged inspection, but Johnson is smart enough to press onward with his plot, leaving us with neither the time nor the desire to linger over the logic...

Full Review | Comment

October 1, 2012

Anthony Lane
New Yorker
Top Critic

A mind-bending ride that is not afraid to slow down now and again, to explore themes of regret and redemption, solitude and sacrifice, love and loss. It's a movie worth seeing and, perhaps, going back to see again.

Full Review | Comments (4)

September 28, 2012

Christopher Orr
The Atlantic
Top Critic

Writer-director Rian Johnson establishes himself as an original talent who clearly believes storytelling must prevail.

Full Review | Comments (3)

September 30, 2012

Richard Roeper
Richard Roeper.com
Top Critic

Looper has more heart than Brick and the 2008 con-man flick The Brothers Bloom. Both fine achievements, they could also be described as viscerally cerebral.

Full Review | Comment (1)

September 28, 2012

Lisa Kennedy
Denver Post
Top Critic
This book was recommended to me by a friend, who said he loved its wicked sense of humor. Named one of the best 100 novels of the last century by The Independent, as most of the highly favorable reviews here attest, but I find all this rather baffling. While not by any means a terrible book, Iain Banks's first novel is simply too messy and book requires that one have a high tolerance for detailed descriptions of cruelty to animals, including the mutilation and immolation of many rabbits and dogs. Some of the possibly macabre level — such as the narrator's description of an uncle's suicide gone terribly wrong — but most of it is simply too dark and literally described to be laughed thinking of the larger implications of any of the book’s violence. While I read "The Wasp Factory," I kept hoping for a denouement that would tie everything together and all I got was a transparent twist that lent nothing to the events that had preceded it, and seemed designed only to shock. In truth, the novel's twist is no more profound than that Banks really felt he was creating something on the level of an O. Henry story, but what he ended up with is a book that reads like a juvenile poison pen letter to all of
Our Starting Test Corpus

- We created a small test corpus of 200 reviews by
  - Considering 5 product types on Amazon
    - Books, sports, cosmetics, appliances, instruments
  - Chose 10 items at random from each product type
    - e.g., The Wasp Factory, Gossip Girl, The Fortress, ... for books
  - Picked 4 reviews, with a range of scores, for each item
Stop-Words

• Remove stop-words
  – Total of 571 common English words removed from analysis
    – m, mainly, many, may, maybe, me, mean, meanwhile, merely, might, more, moreover, most, mostly, much, must, my, myself

• This was the first free database I found from a Google Search

• No stemming, or any other pre-processing at all besides making lower case and removing punctuation
I don't recommend this cymbal at all. Kinda sounds like a gong. A fair representation of what this sounds like would be if you used the bottom high hat as a crash.

I'm a beginner so I thought I'd buy this guitar since it's one of the cheaper ones that I've found online and it had good reviews. I got it and it was such a bad guitar that I returned it less than hours upgraded to a cheap Fender. The strings were so far from the neck that it would have been very hard and frustrating to learn on this and there was buzzing in the first and second frets. Don't waste your time or money on this guitar. Spend a little more and get something worth playing and learning on.

This stuff is worthless. You can barely see it when you first put it on. It has the consistency of water. Way overpriced. Doesn't last. I would give it zero stars if possible.

Benetint is my old reliable. The product is excellent and gives healthy looking glow to lips and cheeks. I hate to travel with it even to work and back because the bottle is always leaky. You also have to be careful opening it if you have been traveling or carrying it in your makeup case because even a little drip or two can really ruin your clothes. I have never found another product that can live up to Benetint. I wish they would come up with smarter packaging though.
Valences

• Find valences of remaining words in a standard valence dataset
  – 2477 words, on a -5 ... +5 scale from very negative to very positive
    – mad -3, maddening -3, made-up -1, madly -3, madness -3, mandatory -1, manipulated -1, manipulating -1, manipulation -1, marvel 3, marvelous 3, marvels 3, masterpiece 4, masterpieces 4, matter 1, matters 1, mature 2, meaningful 2, meaningless -2, medal 3, mediocrity -3, meditative 1, melancholy -2, menace -2, ...

• This was the first free database I found from a Google Search that wasn’t annoying
  – We saw yesterday how in Dutch this database could be made much more extensive on Mechanical Turk
Associations

- Find the valences of every *associate* of a word, using the Leuven database

<table>
<thead>
<tr>
<th>Word</th>
<th>Valence</th>
<th>Avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>glad</td>
<td>3</td>
<td>2.33</td>
</tr>
<tr>
<td>associated</td>
<td>happy</td>
<td>3</td>
</tr>
<tr>
<td>associated</td>
<td>happy</td>
<td>3</td>
</tr>
<tr>
<td>associated</td>
<td>sad</td>
<td>-2</td>
</tr>
<tr>
<td>associated</td>
<td>happy</td>
<td>3</td>
</tr>
<tr>
<td>associated</td>
<td>delighted</td>
<td>3</td>
</tr>
<tr>
<td>associated</td>
<td>scream</td>
<td>-2</td>
</tr>
<tr>
<td>associated</td>
<td>happy</td>
<td>3</td>
</tr>
<tr>
<td>associated</td>
<td>joyous</td>
<td>3</td>
</tr>
</tbody>
</table>

| associated| honour  | 2   |
| associated| loyalty | 3   |
| associated| fear    | -2  |
| associated| appreciate| 2 |
| associated| like    | 2   |
| associated| kind    | 2   |
| associated| love    | 3   |
| associated| healthy | 2   |
| associated| authority| 1  |
| associated| courtesy| 2   |
Schwinn Thrasher Adult Bicycle Helmet

After reading other reviews for this and people saying they had a large head and it fit them I pulled the trigger and purchased this item. It arrived today and I put it on. Sort of the darn thing doesn't fit my head. I turned the dial to make it as large as possible and even at the largest setting it just doesn't sit right on my head and I'm not a large guy at all. The helmet itself looks decent but I have to give it low rating because it doesn't fit and it's not comfortable to me even at the largest setting. Guess I will need to run down to the local store and get a helmet after I can try them on for size. This adult adjustable does not fit all head sizes.
Amazon Results

- Blue results are for mean valence of presented words
- Green results are for mean of \{mean presented words, mean associates\}

Books: $r = 0.49$

Sports: $r = 0.56$

Cosmetics: $r = 0.50$

Appliances: $r = 0.32$

Instruments: $r = 0.71$

Books: $r = 0.53$

Sports: $r = 0.57$

Cosmetics: $r = 0.55$

Appliances: $r = 0.43$

Instruments: $r = 0.70$

- Around 1000 movies from 4 different reviewers, on a 4-point scale
- [http://www.cs.cornell.edu/People/pabo/movie-review-data/](http://www.cs.cornell.edu/People/pabo/movie-review-data/)

This sci-fi'er left a lot to be desired. It didn't work out in 3-d, its original intention of how it was to be shown, as the public's taste for that fad cooled, or in the conventional 2-d, how it was actually shown when released. For the first 30 minutes or so I felt I was stuck in a detention classroom for dummies (ah! even if these machines could do a 1,000 times the work man can do, they were also a 1,000 times deadlier than man, and this reactionary story dwells on the negative side of what machines could do, even if the machine in this instance is a magnificent brain. The action in the film, which is none too thrilling, is in watching how many of the 150 workers will get killed and what novel ways are the killings to be done. If the film was funny or its technology was more interesting, or if the killings weren't so melodramatic, and the acting not so staid, and the patriotic jingoism wasn't so shrill, I might have found this film bearable. But since that wasn't the case, I found the film to be an academic bore.
Standard Movie Corpus Results

- Blue results are for mean valence of presented words
- Green results are for mean of {mean presented words, mean associates}

Dennis+Schwartz: $r = 0.19$

James+Berardinelli: $r = 0.37$

Scott+Renshaw: $r = 0.35$

Steve+Rhodes: $r = 0.39$
• Detecting sentiment for things with complicated semantic content (books, movies) might be very different from simpler products
Good and Bad Decisions I’ve Made in Leuven

- Detecting sentiment for things with complicated semantic content (books, movies) might be very different from simpler products
- While the goal is to avoid intensive pre-processing or parsing, so that the method is very scalable, there are likely simple things that can be done to improve the results
Good and Bad Decisions I’ve Made in Leuven

- Detecting sentiment for things with complicated semantic content (books, movies) might be very different from simpler products
- While the goal is to avoid intensive pre-processing or parsing, so that the method is very scalable, there are likely simple things that can be done to improve the results
- A much better valence database seems likely to help a lot, by giving not only more words but more information about words like variability
Good and Bad Decisions I’ve Made in Leuven

- Detecting sentiment for things with complicated semantic content (books, movies) might be very different from simpler products
- While the goal is to avoid intensive pre-processing or parsing, so that the method is very scalable, there are likely simple things that can be done to improve the results
- A much better valence database seems likely to help a lot, by giving not only more words but more information about words like variability
- Using the Leuven association database was a wonderful decision, because the data are perfect in every way, and the people responsible deserve a medal
Good and Bad Decisions I’ve Made in Leuven

- Detecting sentiment for things with complicated semantic content (books, movies) might be very different from simpler products.
- While the goal is to avoid intensive pre-processing or parsing, so that the method is very scalable, there are likely simple things that can be done to improve the results.
- A much better valence database seems likely to help a lot, by giving not only more words but more information about words like variability.
- Using the Leuven association database was a wonderful decision, because the data are perfect in every way, and the people responsible deserve a medal.
- The rich valence (and associate valence) distribution information is currently crudely collapsed to means, and could surely be handled better (structured representation, sequential sampling).