SemAxis: A Lightweight Framework to Characterize Domain-Specific Word Semantics Beyond Sentiment

Jisun An[1], Haewoon Kwak[1], Yong-Yeol Ahn[2]
[1]Qatar Computing Research Institute, Hamad Bin Khalifa University [2]Indiana University, Bloomington

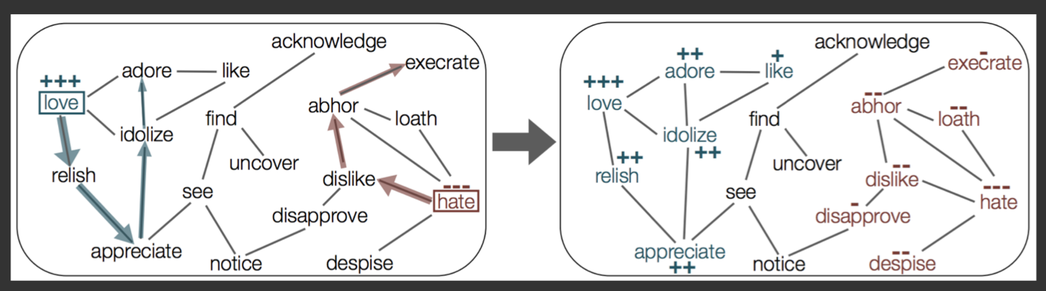
Meaning of words can change?

Yes. **Context** can strongly alter the meaning of words.
Fischer, 1958; Eckert and McConnell-Ginet, 2013; Hovy, 2015; Hamilton et al., 2016b

Kill in video games vs. news
Soft in sports vs. toys

Domain-specific sentiment lexicons

Hamilton et al., 2016a



Is it possible to generalize this idea to general word semantics other than sentiment?

Our key contributions

1. We propose a **general framework** to characterize the domain-specific word semantics.
2. We systematically identify **732 semantic axes** based on the antonym pairs in ConceptNet.
3. We demonstrate that SemAxis can capture **semantic differences** between two corpora.
4. We provide a **systematic evaluation** in comparison to the state-of-the-art, domain-specific sentiment lexicon construction methodologies.

Identifying 732 semantic axes

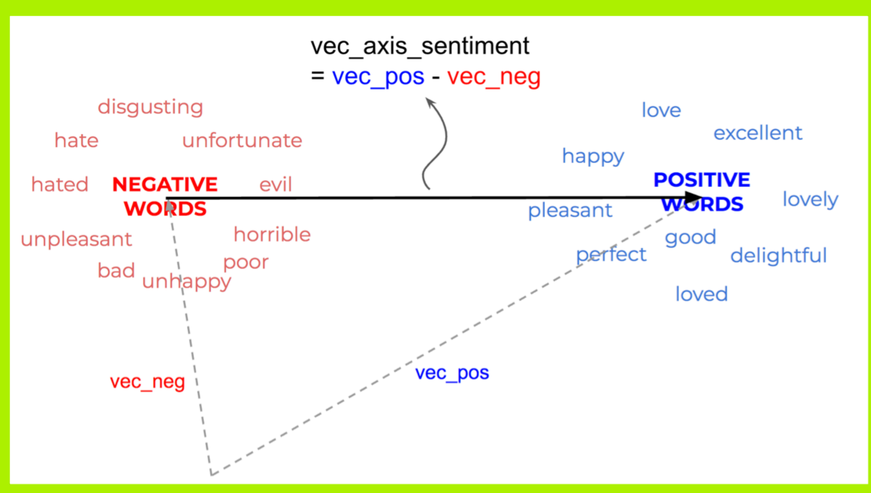
1. We begin with **a pair of antonyms**, called initial pole words. To build a comprehensive set of initial pole words, we compile a list of antonyms from **ConceptNet** 5.5 (Speer et al., 2017).
2. To further refine the antonym pairs, we create a **crowdsourcing** task by asking *Do these two words have opposite meanings?*

Challenges

1. **Small-sized corpus**: Pre-train a word embedding using a background corpus and update this reference model with the target corpora.
2. **Sensitivity to seed words**: Use ℓ closest words on the vector space as well as the two initial pole words.

Basics of our framework, SemAxis

1. Building a word embedding of a given corpus
2. Defining a semantic axis and computing its vector

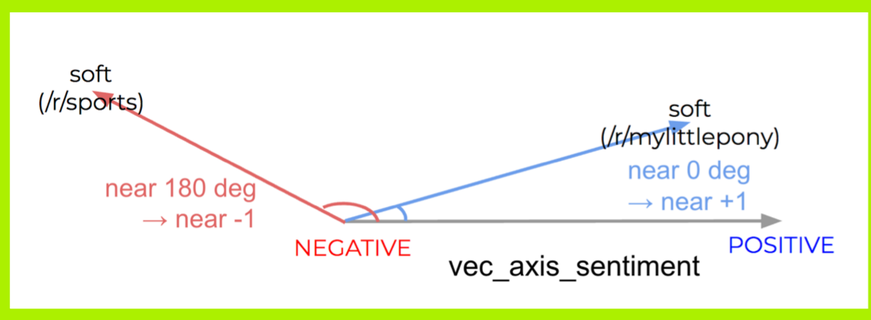


$$\mathbf{V}^{+}=\frac{1}{n}\sum_{1}^{n} v_i^{+}$$

$$\mathbf{V}^{-}=\frac{1}{m}\sum_{1}^{m} v_j^{-}$$

$$\mathbf{V}_{\text{axis}} = \mathbf{V}^{+} - \mathbf{V}^{-}$$

3. Projecting word onto a semantic axis



$$score(w)_{\mathbf{V}_{\text{axis}}} = \cos(v_w, \mathbf{V}_{\text{axis}}) = \frac{v_w \cdot \mathbf{V}_{\text{axis}}}{\| v_w \| \| \mathbf{V}_{\text{axis}} \|}$$

Evaluation

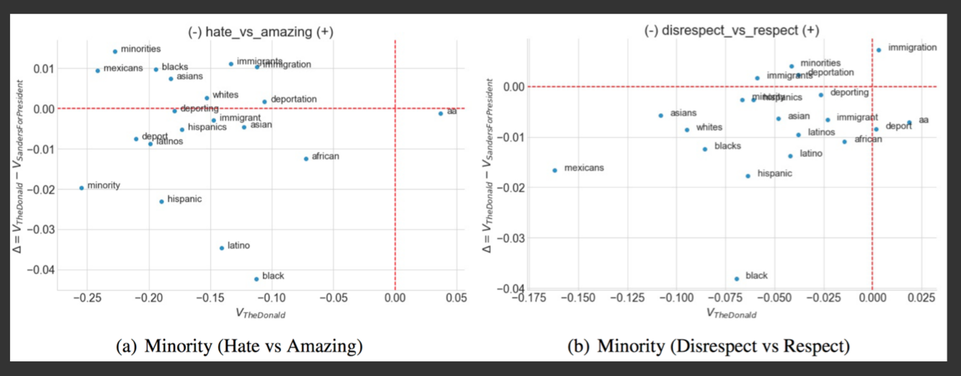
SemAxis outperforms others on both Standard English and Twitter datasets across all measures.

Domain	Positive pole words	Negative pole words
Standard	good, lovely, excellent, fortunate, pleasant, delightful, perfect, loved, love, happy	bad, horrible, poor, unfortunate, unpleasant, disgusting, evil, hated, hate, unhappy
Twitter	love, loved, loves, awesome, nice, amazing, best, fantastic, correct, happy	hate, hated, hates, terrible, nasty, awful, worst, horrible, wrong, sad

Standard English			
Method	AUC	Ternary F1	Tau
SEMAXIS	**92.2**	**61.0**	**0.48**
DENSIFIER	91.0	58.2	0.46
SENTPROP	88.4	56.1	0.41
WordNet	89.5	58.7	0.34
Twitter			
Method	AUC	Ternary F1	Tau
SEMAXIS	**90.0**	**59.2**	**0.57**
DENSIFIER	88.5	58.8	0.55
SENTPROP	85.0	58.2	0.50
Sentiment140	86.2	57.7	0.51

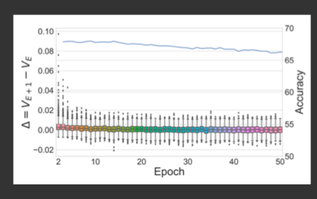
SemAxis in the wild

We compare supporters of Donald Trump (/r/The_Donald) and Bernie Sanders (/r/SandersForPresident), and examine the semantic differences in minority issue based on different axes.

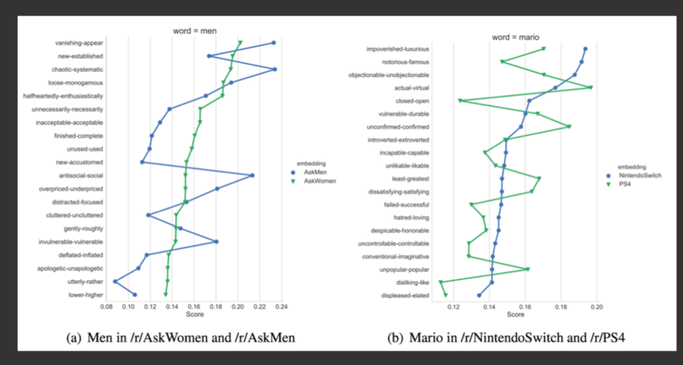


(a) Minority (Hate vs Amazing) (b) Minority (Disrespect vs Respect)

Category	Reddit20M	Google300D
World	28.34	70.2
family	94.58	90.06
Gram1-9	70.21	73.40
Total	67.88	77.08



SemAxis can find, for a given word, a set of the best semantic axes.
We map the target word on our predefined 732 axes and rank the axes based on the projection values on the axes.



(a) Men in /r/AskWomen and /r/AskMen (b) Mario in /r/NintendoSwitch and /r/PS4