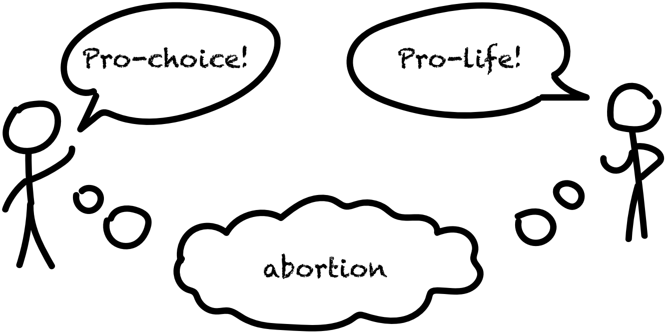
Learning Polarity Embedding

Zhiping (Patricia) Xiao, Zhicheng Ren, Zijie Huang
University of California, Los Angeles

CS263, Spring 2020

Outline

Introduction


Pro-choice!
Pro-life!
abortion

- Topic-modeling?
- Machine-translation?
- Embedding-learning?

Methodology

Word Embedding Settings

Our goal: learn a structured embedding.

$$v = \left[v^{(n)}; v^{(p)}\right] \in \mathbb{R}^{d_n} \times \mathbb{R}^{d_p}$$

where ideally,

- $v^{(n)}$ captures the semantic meanings without any polarity information;
- $v^{(p)}$ captures only the polarity info.

Model — Vanilla GN-GloVe

Learning Gender-Neutral Word Embeddings

$J = J_G + \lambda_D J_D + \lambda_E J_E$, where J_G is the standard un-averaged GloVe loss, J_D has two versions:

$$J_D^{L_1} = -\Big\| \sum_{w \in \Omega_F} w^{(g)} - \sum_{w \in \Omega_M} w^{(g)} \Big\|_1$$

$$J_D^{L_2} = \sum_{w \in \Omega_F} \|\beta_1 \mathbf{e} - w^{(g)}\|_2^2 + \sum_{w \in \Omega_M} \|\beta_2 \mathbf{e} - w^{(g)}\|_2^2$$

where $\mathbf{e} \in \mathbb{R}^k$ is a all-one vector, $\beta_1 = -1$, $\beta_2 = 1$.

$$J_E = \sum_{w \in \Omega_{neutral}} (v_g^T w^{(n)})^2$$

where v_g **per epoch** is: $v_g = \frac{1}{|\Omega'|} \sum_{(w_F, w_M) \in \Omega'} (w_F^{(n)} - w_M^{(n)})$

Discussion: Is GN-GloVe Enough?

Consider the difference between standard gender dataset and our Twitter political dataset.

- Pair quality?
- Noise?
- Relations between pairs?

The experimental results showed that it doesn't perform well in practice.

- Avoid using pairs, using sets instead.
- Taking average (divided by vocabulary size / involved words count).

$$J_D^{L_1} = -\left\| \frac{1}{|\Omega_D|} \sum_{w \in \Omega_D} w^{(g)} - \frac{1}{|\Omega_R|} \sum_{w \in \Omega_R} w^{(g)} \right\|_1$$

$$J_D^{L_2} = \frac{1}{|\Omega_D|} \sum_{w \in \Omega_D} \|\beta_1 \mathbf{e} - w^{(g)}\|_2^2 + \frac{1}{|\Omega_R|} \sum_{w \in \Omega_R} \|\beta_2 \mathbf{e} - w^{(g)}\|_2^2$$

$$J_E = \frac{1}{|\Omega_N|} \sum_{w \in \Omega_N} (v_g^T w^{(n)})^2$$

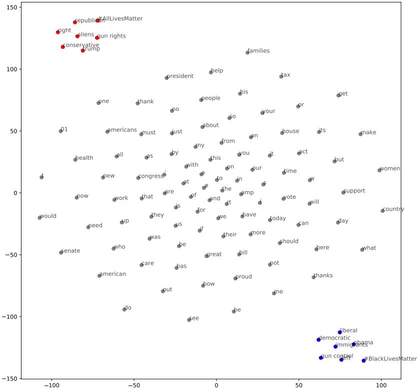


Figure: GN-GloVe. All dimensions, TSNE results of the all $d = 200$ dimensions plotted.

Experiments — Word-Level Case Study II

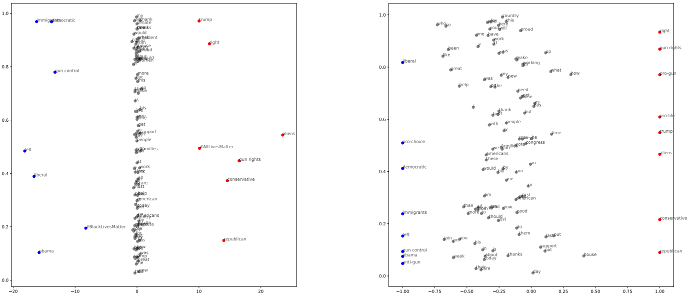


Figure: GN-GloVe. The learned polarity dimensions. Left: $J_D^{L_1}$, Right: $J_D^{L_2}$.

Experiments — Word-Level Case Study III

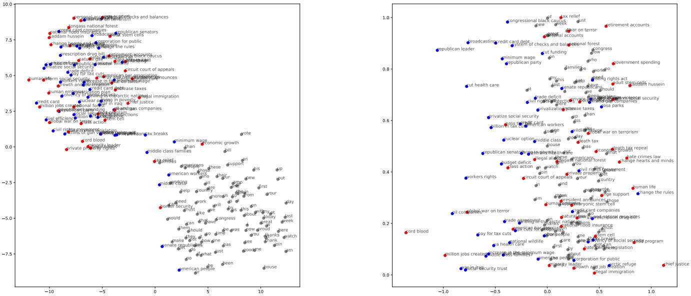


Figure: Set-Based GN-GloVe. The learned polarity dimensions. Left: TSNE of all dimensions, Right: visualization of the polarity dimension $d_p = 1$ only.

Experiments — Sentence-Level Classification

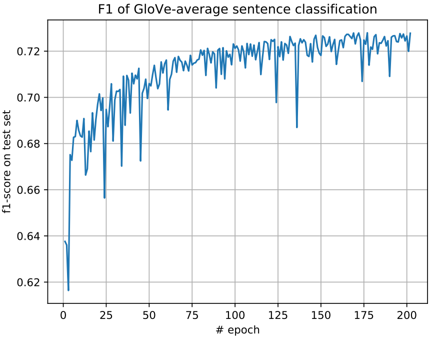


Figure: Word-Average as sentence representation, and then classify the sentences by simply an ordinary 3-layer MLP. Note that by simply using GloVe pre-trained embeddings, the accuracy of sentence-level prediction is also around 70%. It shows that the embeddings learned are not good enough.

Discussion: TODO

- Multi-Task Learning.
  - Task 1: J_G
  - Task 2: J_D
  - Task 3: J_E
- Other architecture.

Code & Data

Github Link

- Our implementation (code & data): `https://github.com/PatriciaXiao/CS263_PolarityWordEmbedding`
- GN-GloVe implementation (code & data): `https://github.com/uclanlp/gn_glove`
- Additional tool: AutoPhrase `https://github.com/shangjingbo1226/AutoPhrase`