Political Depolarization of News Articles Using Attribute-Aware Word Embeddings

Ruibo Liu, Lili Wang, Chenyan Jia, Soroush Vosoughi

Mind, Machine, and Society Lab @ Department of Computer Science, Dartmouth College

ICWSM **2021**

Motivation

1. Our everyday news are politically biased.
2. Reader tend to consume like-mind arguments.
3. We need a tool to depolarize the biased news.

Contributions

1. We show that word embeddings can capture bias.
2. We present a framework to depolarize bias.
3. We evaluate our model emperically.

Dataset

We colllect around 360k articles from 22 media outlets, covering 11 hot topics on social issues.

Model

Our model has two main modules:

1. **Dynamic Attribute-aware Embeddings.**
2. Text Annealing Depolarization Algorithm (**TADA**).

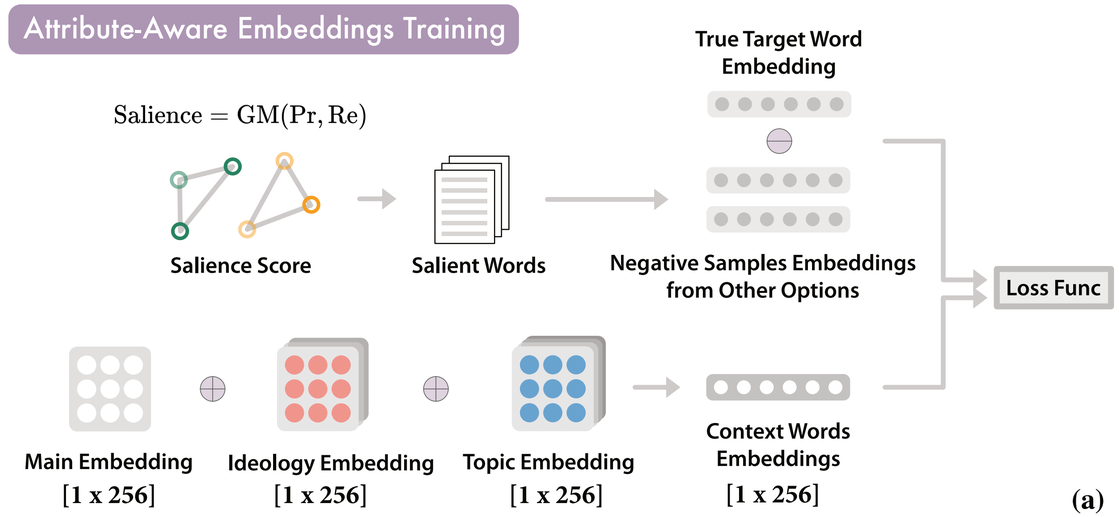
Evaluation

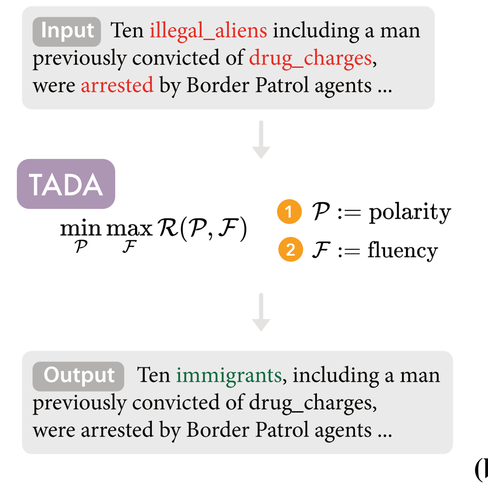
1. We successfully depolarized 90.1% of paragraphs in semi-automatic mode and 78.3% of paragraphs in fully-automatic mode.

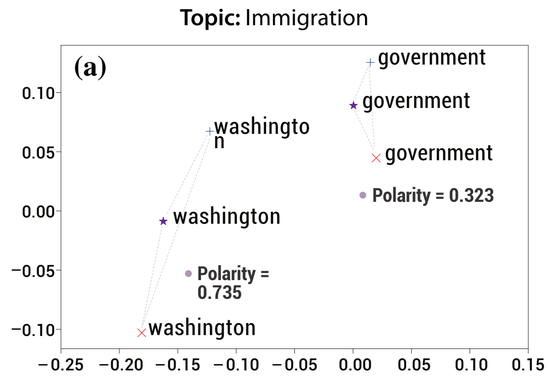
2. 81.2% of the testers agree that the non-polar content infor mation is well-preserved and 79% agree that depolarization.

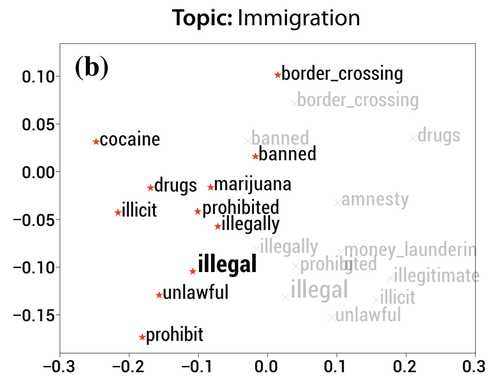
Overview of Model

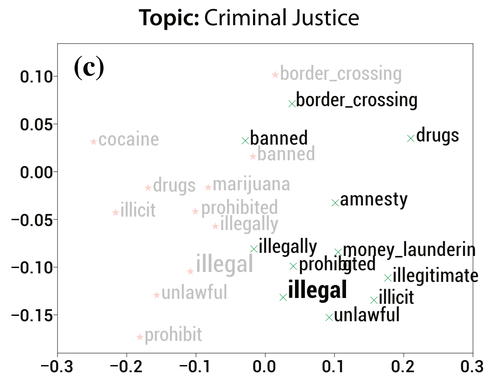
(a) Dynamic Attribute-aware Embeddings (b) TADA











In the left subfigure, we plot 2D projections of the words "government" and "washington" from liberal, neutral and conservative within the same topic "Immigration", as a way to visualize the polarity of the words.

In the two right subfigures, we plot the ten nearest words to the word "illegal" in two different topics ("Immigration" and "Criminal Justice"), to verify that our attribute-aware word embeddings manage to capture topic information.