Multi-class Classification of Tweets Based on Kindness Analysis

Charles Han, Wanzi Zhou, Xinyuan Huang
hcs@stanford.edu, wanziz@stanford.edu, xhuang93@stanford.edu

Introduction

Establishing a kindness assessment mechanism is very helpful for maintaining a healthy environment for social network. Former studies focused on spotting out Internet trolls, which is a binary classification; while we propose to analyze each tweet/user's kindness degree, which leads to multinomial classification for applications like a rewarding system or parent control. Using K-means clustering and Gaussian Mixture model we classify tweets into three categories. We further test our model on three US politicians: Barack Obama, Donald Trump and Hillary Clinton.

Data

Twitter has always been a great source for Natural Language Processing researchers. It has sufficiently large size of data, along with outstanding qualities - which comprises of real-life conversations, uniform length (140 characters), rich variety, and real-time data stream.

With Twitter API, we captured a random sample of tweets in continuous 24 hours in a regular day and picked out all the English tweets. Through the above procedures, we obtained 58295 tweets as our dataset for this project. We also collected two lexicons of positive words such as "amazing" and negative words such as "bastard", which has 723 and 236 words respectively. We clean the data by transforming all the letters into lower cases and neglecting the punctuations.

Feature

For every tweet we obtained from the dataset, we compare them to the words in the dictionary of both positive words and negative words, and obtain a 959 x 1 feature vector, where each value in the vector represents the number of times the word appears in a certain tweet. There are 354 words that are not found in all the collected tweets so they are neglected from the feature vector. We then use the 605 x 1 feature vector to implement the learning part.

Unsupervised Learning Models

K-means

Using prior knowledge, we set the initial centers to be the average vectors of positive, negative and neutral elements in all feature vector matrix. Then we iterate the two steps:

For every i, i=1,…,58295, set

$$c^{(i)} = \arg\min_j ||x^{(i)} - \mu_j||^2$$

For every j=1,2,3, set

$$\mu_j := \frac{\sum_{i=1}^{m} 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c^{(i)} = j\}}$$

till convergence.

Gaussian Mixture Model

We use three Gaussians representing positive, negative and neutral classifications.

Our goal is to maximize the log likelihood

$$l(\phi, \mu, \Sigma) = \sum_{i=1}^{m} \log \sum_{z^{(i)}=1}^{k} p(x^{(i)} | z^{(i)}, \mu, \Sigma) p(z^{(i)}, \phi)$$

To achieve this we use EM algorithm.

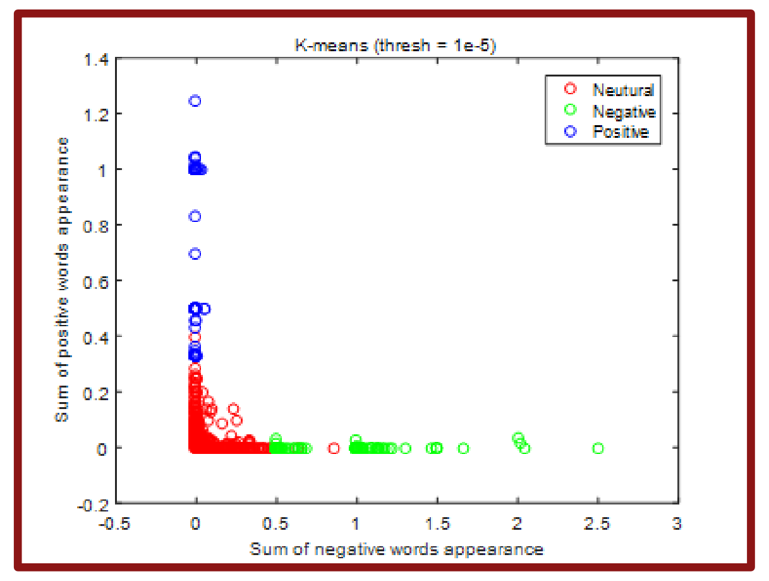
E-step: compute posterior distribution of $z^{(i)}$ given $x^{(i)}$ and parameters.

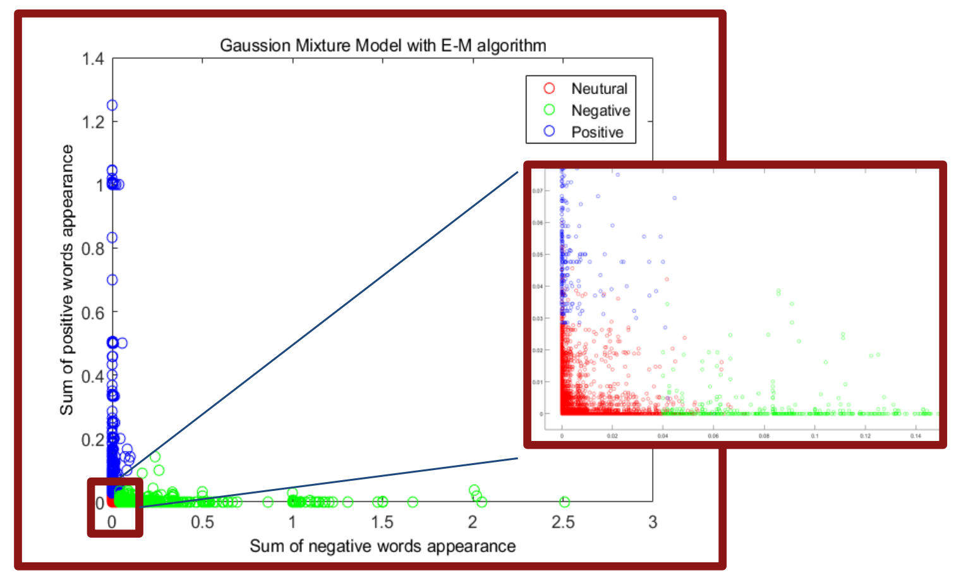
$$w_j{}^{(i)} = p(z^{(i)} = j | x^{(i)}; \phi, \mu, \Sigma)$$

M-step: we update the $\phi_j, \mu_j, and\ \Sigma_j$ for every j=1,2,3. Repeat until convergence.

Results

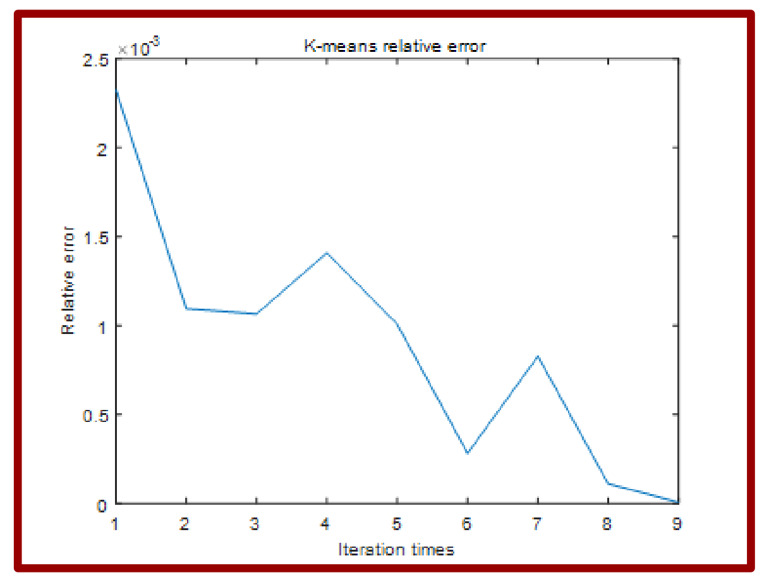
We performed k-means and Gaussian mixture model with EM algorithms and classified the tweets into three clusters. Since the feature dimensions are very high(605x1), we projected the result into 2-dimensions to visualize, where x,y dimensions represent the normalized sum of positive and negative feature number counts, respectively.

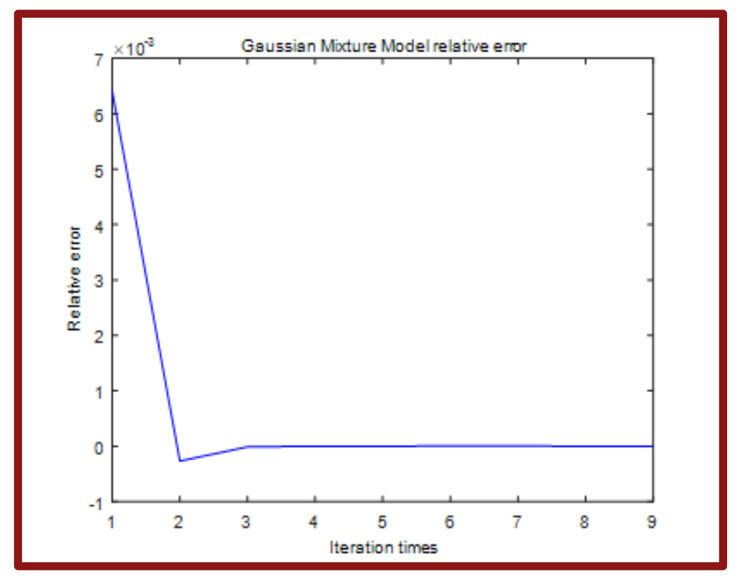




Multi-class classification on 58295 tweets

	K-means	GMM
Positive	110	1461
Negative	137	1129
Neutral	58048	55705





Model test on three US politicians

	Barack Obama	Donald Trump	Hillary Clinton
Positive	35	50	47
Negative	10	10	8
Neutral	155	140	145

Discussion

By extracting feature vector with a dictionary of positive and negative words and applying unsupervised learning models for multi-class classification, we find that most of the tweets online are neutral. With K-means and Gaussian mixture model, we find a proportion of 0.4% and 9% tweets are either positive or negative, respectively, which shows that Gaussian mixture model can better recognize positive or negative tweets. This result is because with hard assignment, K-means only realizes spherical clusterings, while GMM considers probability and incorporates the covariance structure of data and adjusts itself to elliptic clusterings.

Using our trained model to test on the recent 200 tweets of three US politicians Barack Obama, Donald Trump and Hillary Clinton, we show that they all follow the same pattern: while most of their tweets are neutral, their proportion of positive tweets are significantly higher than general public.

Future work

Current feature extraction method assigns same weight for each word and ignores the structure of the sentence or the context. To solve this issue, we will look into learning vector representations of the words. Also we want to know deeper in the logic gap between positive and negative words on a psychological level.

Reference

[1] Cheng, Justin, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. "Antisocial behavior in online discussion communities." arXiv preprint arXiv:1504.00680 (2015).
[2] ttp://www.frontgatemedia.com/a-list-of-723-bad-words-to-blacklist-and-how-to-use-facebooks-moderation-tool/
[3] http://www.the-benefits-of-positive-thinking.com/list-of-positive-words.html