MULTI-DIMENSIONAL SENTIMENT ANALYSIS

Team members: Jenna Kudaimi, Hiep (Roger) Nguyen, Rudraaksh Karthick Bhuvaneswari, Aniket Shirodkar, Krupali Soni

Topic of interest
Sentiment Analysis for Emotion Detection

Research question
How much can sentiment analysis infer information, rather than only positivity/negativity, to predict emotions (multiple spectra) underneath the text, given that corpus by modern English speakers regardless of their background and level of education?

Summary
This research proposal explores the multidimensional nature of sentiment analysis for emotion detection, aiming to move beyond the binary of positivity/negativity. We intend to investigate the predictive power of sentiment analysis on variegated expressions from diverse modern English speakers, with no bias towards background or education level.

To navigate the complexity of sentiment and emotion in text, we will deploy multiple machine learning algorithms, namely Support Vector Machines, Regression, Random Forests, and probably k-means clustering for preliminary analysis. These methods will be tailored to manage the nuances of emotion detection, informed by insights from contemporary literature that highlight the shift from keyword-based to deep learning approaches in sentiment analysis.

With an acute awareness of the potential biases that traditional unsupervised word embeddings may introduce, our approach will involve contextual and connotative multi-dimensional analysis to cultivate a more refined emotion detection model. The model development will be informed by state-of-the-art practices such as those presented in the reviewed sources, which emphasize the efficacy of ensemble methods and deep learning in capturing complex emotional states.

Our dataset selection process will emphasize on the judicious selection of datasets that represent a wide demographic to ensure inclusivity and robust training. We will be critically aware of sources of bias, from data imbalance, lexicon construction, or algorithmic predispositions, in our data and methods, implementing measures to mitigate their effects.

In summary, our goal is to answer the question of how effectively sentiment analysis can identify and classify the complex landscape of human emotions from text by using sophisticated machine learning techniques while accounting for a diverse population of English speakers. This will not only advance the field of sentiment analysis but also enhance the understanding and application of AI in processing human language and emotions.

Types of machine learning algorithm(s)

- Support Vector Machines: https://www.educative.io/answers/how-to-use-svm-for-sentiment-analysis
- Regression: https://towardsdatascience.com/sentiment-analysis-using-logistic-regression-and-naive-bayes-16b806eb4c4b
- Random Forests: https://mlarchive.com/machine-learning/sentiment-analysis-with-random-forest/
- k-means clustering (emotion): https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/

Datasets

- https://www.kaggle.com/datasets/prajwalkanade/sentiment-analysis-word-lists-dataset
- https://www.kaggle.com/datasets/parulpandey/emotion-dataset

Sources of bias

- **Data Representation:** Biases may come from the lack of representation of training data and user demographics related to culture, language use, slang, and idioms that may not be universally understood or represented in the training set.
- **Lexicon and Annotation Biases:** Biases may be caused by the subjective choice of word, reflecting the annotators' cultural and personal perspectives.
- **Class Imbalance:** Biases may also be led by an overrepresentation of certain sentiments or emotions, leading to less accurate predictions for the underrepresented classes.
- **Feature Selection:** Biases may definitely appear in the features selection for training the sentiment analysis models, especially if they capture irrelevant or misleading aspects of the data that correlate with unintended attributes.
- **Socio-Cultural Factors:** Biases may arise due to the failure to account for the varying ways different cultures express emotions in text, resulting in biased emotion detection that does not accurately reflect the intended sentiment of speakers from different backgrounds.

Citations

- https://www.jstage.jst.go.jp/article/jaciii/27/1/27_84/_pdf
- https://www.mdpi.com/2076-3417/13/7/4550
- https://ai.stanford.edu/~amaas/papers/multisent-techreport-2011.pdf
- https://link.springer.com/article/10.1007/s13278-021-00776-6
- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3291299
- https://dl.acm.org/doi/abs/10.1145/3057270
- https://oneai.com/learn/sentiment-analysis-vs-emotion-detection