Complex Emotion Analysis of Song Lyrics

Darren Edmonds
Department of Computer Science
Johns Hopkins University
dedmon13@jhu.edu

Text-based sentiment analysis has become increasingly popular in recent years, in part due to its numerous applications in fields such as marketing, politics, and psychology (Rambocas and Pacheco, 2018; Haselmayer and Jenny, 2017; Provoost et al., 2019). However, the vast majority of sentiment analysis models are built to identify net positive or negative sentiment rather than more complex, ambiguous emotions such as anticipation, surprise, or nostalgia (Jongeling et al., 2017). As a result, current models usually fail to portray the coexistence of multiple emotions within a text sample, resulting in limited characterization of a human's true emotions. Songs are often created to elicit complex emotional responses from listeners, and thus are an interesting area of study to understand nuanced emotions (Mihalcea and Strapparava, 2012).

This ongoing project examines a variety of approaches to the multi-emotion classification problem, in which we want to determine the presence of multiple emotions in a sample of text. Given that inter-annotator agreement for text-based sentiment analysis most often produces an agreement rate between 70-90% (Diakopoulos and Shamma, 2010; Bobicev and Sokolova, 2017; Takala et al., 2014), the main goal of this project is to build an emotion classification model that can consistently detect the presence of multiple emotions in text.

A data set consisting of 600 song names, artists, and lyrics was created by searching through a Spotify playlist consisting of 800 songs, both lyrical and instrumental, and collecting available lyrics from LyricFind, Genius, and MusixMatch[1]. Each song was then labeled with an array of size 8; indices respectively corresponded to the emotions of joy, trust, fear, surprise, sadness, disgust, anger, and anticipation, with each array index containing a value of 1 or 0 to indicate an emotion's presence or absence.

The basis for label selection was provided by Plutchik's Theory of Emotion, which postulates that all emotions are combinations of the 8 core emotions present in our label (Plutchik, 2001). As a result, the label can lead to additional classification models for emotions which are theorized to be dyads of the core emotions (e.g, $P_{Love} = P_{Joy} * P_{Trust}$, or $P_{Aggressiveness} = P_{Anger} * P_{Anticipation}$).

Before model-building, song lyrics in the data set were cleaned by removing punctuation and capitalization. During model selection, Naive Bayes was first chosen due to its widespread applications in text classification and sentiment analysis (Raschka, 2014). 5-fold cross validation was then implemented to train various Naive Bayes classifiers for each emotion. The best-performing model here was a Multinomial Naive Bayes model using a TF-IDF vectorizer, which detected anger, fear, disgust, and surprise in test data with over 80% accuracy, as well as the presence of anticipation with 71% accuracy.

To explore a second approach to the multi-emotion classification problem, lyrical data was then transformed into a feature vector of length 9 using the NRC Emotion Lexicon, which contains binary indicators regarding the presence or absence of Plutchik's 8 core emotions in 14182 common English words (Mohammad and Turney, 2013). This occurred by iterating through a song's lyrics, counting each word present in the NRC Emotion Lexicon as well as its emotional classification, and storing this information in the feature vector. For example, the feature vector $[5, 10, 1, 9, 4, 2, 2, 3, 28]$ would correspond to a song's lyrics that contained 28 words (not necessarily all distinct) which were present in the NRC Emotion Lexicon. Of these words, 5 were associated with joy, 10 with trust, 1 with fear, etc.

Given its robustness to outliers and its ability to deal with imbalanced data classification (Chen et al., 2004), a Random Forest model was chosen to analyze this data. The best-performing Random Forest model on the transformed data had 50 estimators; this model detected the presence of surprise in test data with 76% accuracy, as well as the presence of anger, fear, and disgust with over 70%

[1] Links to the websites used: https://www.lyricfind.com/ https://genius.com/ https://www.musixmatch.com/

accuracy.

Some shortcomings of this project include the fact that dataset labels are based off of the judgment of only one person, as well as the tendency of both Naive Bayes and Random Forest models to disproportionately misclassify the minority class. Next steps for this project include imposing a higher-weighted penalty on misclassification of the minority class, as well as exploration of oversampling and undersampling techniques. We also plan to increase the quality and size of the labeled dataset through a combination of active learning strategies and crowdsourcing additional contributions.

References

Victoria Bobicev and Marina Sokolova. 2017. Inter-annotator agreement in sentiment analysis: Machine learning perspective. In *RANLP*, pages 97–102.

Chao Chen, Andy Liaw, Leo Breiman, et al. 2004. Using random forest to learn imbalanced data. *University of California, Berkeley*, 110(1-12):24.

Nicholas A Diakopoulos and David A Shamma. 2010. Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1195–1198.

Martin Haselmayer and Marcelo Jenny. 2017. Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. *Quality & quantity*, 51(6):2623–2646.

Robbert Jongeling, Proshanta Sarkar, Subhajit Datta, and Alexander Serebrenik. 2017. On negative results when using sentiment analysis tools for software engineering research. *Empirical Software Engineering*, 22(5):2543–2584.

Rada Mihalcea and Carlo Strapparava. 2012. Lyrics, music, and emotions. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 590–599, Jeju Island, Korea. Association for Computational Linguistics.

Saif M Mohammad and Peter D Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4):344–350.

Simon Provoost, Jeroen Ruwaard, Ward van Breda, Heleen Riper, and Tibor Bosse. 2019. Validating automated sentiment analysis of online cognitive behavioral therapy patient texts: An exploratory study. *Frontiers in psychology*, 10:1065.

Meena Rambocas and Barney G Pacheco. 2018. Online sentiment analysis in marketing research: a review. *Journal of Research in Interactive Marketing*.

Sebastian Raschka. 2014. Naive bayes and text classification i-introduction and theory. *arXiv preprint arXiv:1410.5329*.

Pyry Takala, Pekka Malo, Ankur Sinha, and Oskar Ahlgren. 2014. Gold-standard for topic-specific sentiment analysis of economic texts. In *LREC*, volume 2014, pages 2152–2157.