Cono: A Word Emotion Lexicon Built On Movie Scripts

Barbara Kiyomoto Katsuhito Sudoh Satoshi Nakamura
Nara Institute of Science and Technology
{kiyomoto.barbara.kb7,sudoh,s-nakamura}@is.naist.jp

Abstract

A common issue that language learners face is identifying word emotion in their non-native language. Translations that they reference may provide a literal translation, without consideration for the emotional information a word carries. There is a limited amount of word emotion lexicons available that could help with this problem, therefore we attempt to create a system that automatically classifies word emotion using movie scripts as training data. We then create a learner-friendly application called Cono, that allows language learners to view emotion lexicons in a more accessible way by providing analysis for their input text and then identifying emotion words along with their synonyms with emotion label predictions.

1 Introduction

A word may have many synonyms, where the synonyms might have very similar meanings, but different emotional sentiments. For example, **slim** and **lanky** are both synonyms for the word **thin**. However, only one of those synonyms could be used with positive sentiment to compliment one's figure, i.e., slim. A language learner may face difficulty in choosing the right synonym for the right context, if they are not aware of the emotional meaning. A tool that could help a language learner with this issue, aside from access to a native speaker, is an emotion lexicon.

An emotion lexicon could be used just like a dictionary, except it is used to find the emotional sentiment attached to a word rather than the literal meaning. However, there is a limited amount of existing lexicons that learners could access, so this research hopes to add to the current available resources. Furthermore, if a language learner finds that a word they tried to use in their non-native language does not contain the sentiment they intended, they would want to find a proper synonym with the right emotion meaning. Therefore, this research hopes to also present an effective UI for language use support.

This paper will begin with an introduction to well-known emotion and sentiment theories, before continuing with a reference to similar related studies and then an explanation of our method in creating our word emotion classification system with a final analysis of its inner workings in the evaluation and conclusion.

2 Background on Emotion Models

It is important to start with a clarification of the difference between the terms sentiment and emotion. Sentiment analysis focuses on the distinction of positive, negative, or neutral, whereas emotion analysis focuses on identifying specific emotions such as anger, sadness, or joy. Emotion analysis can be considered "a natural evolution of sentiment analysis" according to Staiano and Guerini (2014) [1]. From this point, the terms sentiment and emotion will be used accordingly.

In our research, we choose to focus on finding a way to classify synonyms in emotion categories rather than sentiment, in order to provide more varied information. Synonyms like **slim**, **lanky** and **thin** can certainly be respectively categorized into either positive, negative or neutral sentiment. However, if we add another synonym like **emaciated**, it can be more specifically differentiated from the other synonyms as a word that carries the emotions of fear and sadness.

With regard to emotion analysis, a commonly debated question is: how many emotions can you categorize emotion into? The answer is there is no single set of agreed-upon categories; however, there are several well-known emotion models that propose their own emotion categories. The most well-known being Ekman's six basic emotions [2]. Paul Ekman proposed that there were six basic and universal emotions, which all cultures could identify by facial expressions. These six categories being: anger, fear, happiness, sadness, disgust, and surprise.

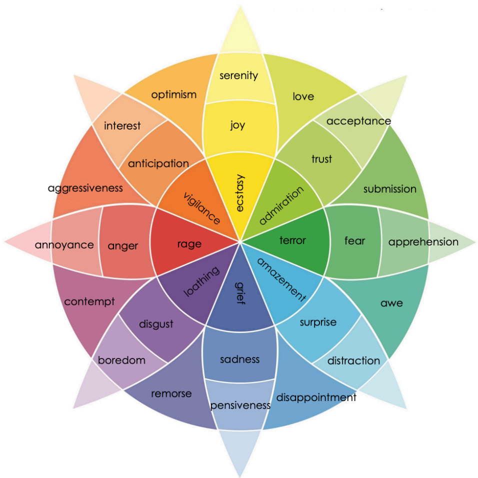

Figure 1 Plutchik's Wheel of Emotions

Another well-known model is Plutchik's wheel of emotions [3]. Plutchik proposed eight basic emotions of joy, trust, fear, surprise, sadness, disgust, anger, and anticipation. He then placed these basic emotions on a color wheel, including a spectrum of stronger and weaker variations for each emotion, as seen in Figure 1. For example, looking at the emotion of joy (in the yellow spectrum), the stronger variation is ecstacy and the weaker variation is serenity.

It is also possible to view emotion as a spectrum, rather than a set of emotion categories. The Valence-Arousal-Dominance model, or VAD model, proposed by Russel and Mehrabian [4] views emotion as points in 3D space along the axes of **valence**(pleasure-displeasure), **arousal**(calmness-excitement), and **dominance**(control-lack of control).

3 Related Studies

3.1 Automatic Emotion Lexicon Creation

In the field of emotion lexicon generation, a dilemma besides choosing the appropriate emotion model to use is how to create a lexicon while reducing the high cost of human annotation.

DepecheMood is one Emotion Lexicon that is automatically created from news articles on rappler.com. For each article on this website, readers can vote for one of eight emotion labels, thus creating a large set of documents with emotion labels collected through crowd-sourcing. Using distributional statistics and unsupervised learning, Staiano and Guerini [1] calculate the most likely emotion labels for words based on their frequency in each article.

DeepMoji [5] is another system that works around the lack of annotated data. It is trained on a dataset of tweets containing emojis to choose the most fitting emoji from a set of 64 to match a tweet. Although it does not perform automatic emotion classification of words, it performs a similar function of sentence level emotion classification without the need for annotation. This research illustrates the possibilities of using emojis as a type of emotion model, and the benefits to using already emotion-rich data.

3.2 Emotion Analysis in Education

In this research, we also examine how emotion analysis and emotion lexicons have been used in the classroom to aid learning and teaching. Emotion analysis with facial recognition on students has been used in the system 4 Little Trees [6]. Student facial expressions are analyzed during class and categorized into Ekman's six basic emotions with an 85% accuracy rate. The system is used for teachers as a way to monitor student emotions for better classroom management and learning support. However, such uses of emotion analysis have been criticized as violations of privacy of one's emotions, in addition to causing unwanted consequences such as discrimination when errors occur[7]. While facial expression analysis may be a sensitive topic, it is also possible to use sentiment or emotion lexicons to analyze student text, e.g. Feng and Qui [8] construct a sentiment lexicon in the educational field, that can be used to analyze positive or negative sentiment in student text.

Although emotion analysis is used on students by educators, little effort is shown in making emotion data available to learners for self-study purposes rather than management purposes. In contrast to what has already been done with emotion lexicons, we consider learner-based emotion lexicon availability when we present our results.

4 Methodology

This research hopes to find a method to automatically create an emotion lexicon based on a word's frequency scores among various labelled documents. Rather than requiring emotion annotation by annotators as the main part of the process, we would like to observe if a word's emotion label can be estimated based on its use frequency in various movie genres. Our inspiration is the distributional hypothesis, which posits that "there is a correlation between distributional similarity and meaning similarity"[9]. Since a movie in a certain genre is likely to trigger a specific emotion (such as a horror film activating the fear emotion, or a comedy activating joy), we hypothesize that a pattern of occurrence can be found for emotion words across movie genres. Based on this hypothesis, we use supervised learning to predict emotion labels on words based on their use frequency in various movie genres.

4.1 Collecting the Training Data

Movie scripts were collected to be used as training data using the method by Ramakrishna et al. (2017) [10]. In total, 448 movie scripts were obtained, along with their genre tags. The script data was then organized into 19 genre categories. Since a single movie often had more than one genre tag, most of the movie scripts exist in several genre categories. The number of scripts in each genre is shown in Table 2 in the Appendix.

We then collect labelled emotion words from the NRC Emotion Lexicon [11]. This lexicon contains more than 14,000 words with crowd-sourced binary labels for positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. This lexicon contains words that are not emotion-bearing as well, marked with binary (0) labels for each emotion. To create our training dataset, we find the TF-IDF frequency scores of each word from the NRC Emotion Lexicon, in each movie genre. TF-IDF (term frequency-inverse document frequency) was used because scores for content words (meaning-bearing words) are higher, and scores for function words (mostly grammatical words) are lower. The resulting training dataset consists of the NRC Emotion Lexicon words with TF-IDF scores for each word in 19 movie genres.

4.2 Training the Prediction Model

To predict emotion on words, 8 models were trained on the NRC Emotion Lexicon with TF-IDF scores to predict binary labels for the emotions of positive, negative, anger, disgust, fear, joy, sadness, and surprise. The emotions of anticipation and trust from the NRC Emotion Lexicon are excluded, as we plan to represent the results of the research with emojis that are easily interpretable by language learners. We predict for Ekman's six emotions, which are universally identifiable facial expressions [2] and therefore adaptable to emoji, as well as sentiment.

Table 1 Model performance on test data

	K-Means F1	SVM(RBF) F1
Positive	73.92%	76.80%
Negative	63.94%	66.40%
Anger	77.56%	73.50%
Joy	87.34%	91.97%
Sadness	76.25%	74.85%
Disgust	79.67%	80.09%
Surprise	89.74%	94.14%
Fear	74.00%	68.81%

Emotion prediction was performed for each emotion with the method of K-means with the Minkowski distance metric, and Support Vector Machine (SVM) in the RBF kernel. Each model was trained with 80% of the dataset, and 20% of the dataset was then used for testing the performance of the model. In preliminary experiments, there was an issue with recall due to the data imbalance between positive and negative binary labels. To resolve this issue, emotion word data was duplicated in the training data nine times. For sentiment prediction, the training data only contains two duplicates of sentiment words, since increasing the duplicates too many times caused an overwhelming amount of false positives. The performance of the perfected models on the test data is shown in Table 1.

4.3 Creating a New Emotion Lexicon

The SVM model performs the best on test data, so this is the model used to create a new emotion lexicon. We hope to aid learners particularly with the identification of differences in synonym meaning, therefore our next step was to collect synonyms of words in the NRC Emotion Lexicon from WordNet. Using the same method as when creating the training dataset, the TF-IDF frequency scores of the synonym words were found in scripts for each movie genre. Inputting the synonyms with their TF-IDF frequencies into our model, the emotion predictions were collected for each synonym in 8 categories of sentiment and emotion: positive, negative, anger, disgust, fear, joy, sadness, and surprise. If a synonym is predicted to have no sentiment or emotion, it is assumed the word is neutral and labelled as such. We will refer to the resulting dataset as the Cono dataset.

4.4 Creating the Cono Application

One main focus of this research has been to develop larger emotion lexicons without the need for copious

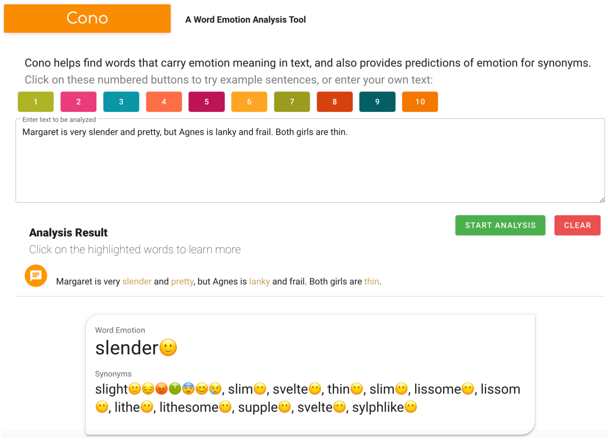

Figure 2 The Cono Tool

amounts of annotation, while another vital motivation is to make emotion lexicon data more readily available to language learners. This section describes the Cono application tool (shown in Figure 2), which we have created as an aid to learners in identifying word emotion within text.

In creating the user interface (UI) of Cono, the use of complicated jargon is avoided to provide a scaffolded explanation of the purpose of the tool. Our explanation states that this is "A Word Emotion Analysis Tool" that will help find emotion-carrying words in text, and their synonyms with emotion prediction labels. Users can select example sentences to try first, or proceed to enter their own text for analysis. The system first highlights words in the text that belong to the NRC Emotion Lexicon. These words are shown with an emoji next to each word, based on the emotion label provided in the NRC Emotion Lexicon. Ekman's six basic emotions are universally distinguished by facial expressions, which is why emoji are shown as a visual aid. The visual representation will assist English learners in creating emotion associations with words.

After finding emotion words in the text, the synonyms from the Cono dataset are shown with the emotion predictions (also represented in emoji form) provided from our model. To ensure that the meaning of the emoji representations is clear, a pop-up legend is provided, showing each emoji and the emotion it represents.

5 Evaluation

Eight volunteers participated in a user evaluation experiment in order to judge the use-ability and helpfulness of the Cono application. The demographics were 3 female and 5 male participants; 7 participants were non-native English speakers, and 1 participant was a native speaker. Based on TOEIC scores, the English proficiency levels are distributed amongst the evaluators as such: 3 evaluators with professional proficiency, 1 evaluator with working proficiency, 3 evaluators with basic working proficiency, and 1 evaluator with advanced elementary proficiency.

The overall response to the UI was positive. 75% of the evaluators found the purpose of the application to be clear, and found no issues with the website features or visual design. A few issues noticed by the other 25% included the inability to input long text, or the discomfort in having to click on the legend each time. Suggestions given included providing dictionary definitions and example use cases of synonyms in sentences for easier comparison.

To evaluate model predictions, evaluators were asked to compare labels from the NRC Emotion Lexicon with labels from the Cono dataset, and rate them on a five-point scale from Very Accurate to Very Inaccurate. Regarding NRC Emotion Lexicon labels, 3 evaluators voted that the labels were Very Accurate, and 5 voted them to be Accurate. For emotion labels from the Cono dataset, 3 evaluators rated the labels as Accurate, 4 selected Inaccurate, and 1 evaluator selected neither Inaccurate nor Accurate. Many evaluators noted excessive neutral labels in the Cono dataset. As this label was chosen only after a word was predicted not to have other emotion meanings, this part of the system will be adjusted. Additional information from the user evaluation is provided in Figures 3 to 7.

6 Conclusion

This paper described the construction of the Cono dataset and Cono UI, with consideration to previous research on emotion modeling and emotion lexicons, and educational support. In future work, changes will be made to improve the prediction model and UI based on feedback. Easy accessibility for learners to emotion lexicons will remain a fundamental focus. The novelty point of this research as well was showing the possibilities of emotion lexicons as learner support. Many existing emotion lexicon datasets are built for use in opinion-mining and emotion analysis in research or business marketing, while here we show their use in language education.

References

[1] Jacopo Staiano and Marco Guerini. Depeche mood: a lexicon for emotion analysis from crowd annotated news. In **Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**, pp. 427–433, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[2] Dylan Evans. **Emotion: A Very Short Introduction**. Oxford: Oxford University Press, 2003.

[3] Robert Plutchik. 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**, Vol. 89, No. 4, pp. 344–350, 2001.

[4] James Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. **Journal of Research in Personality**, Vol. 11, pp. 273–294, 09 1977.

[5] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing**, pp. 1615–1625, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[6] Madhumita Murgia. Emotion recognition: Can ai detect human feelings from a face? **Financial Times**, May 2021.

[7] Kate Crawford. Time to regulate AI that interprets human emotions. **Nature**, Vol. 592, No. 7853, pp. 167–167, April 2021.

[8] Xiang Feng and Longhui Qiu. The construction of sentiment lexicon in educational field based on word2vec. In **Annual Proceedings of Selected Research and Development Papers Presented at the Annual Convention of the Association for Educational Communications and Technology**, Vol. 1, pp. 129–146, Jascksonville,Florida, November 2017.

[9] Magnus Sahlgren. The distributional hypothesis. **The Italian Journal of Linguistics**, Vol. 20, pp. 33–53, 2008.

[10] Anil Ramakrishna, Victor R. Martínez, Nikolaos Malandrakis, Karan Singla, and Shrikanth Narayanan. Linguistic analysis of differences in portrayal of movie characters. In **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1669–1678, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[11] Saif M. Mohammad and Peter D. Turney. Crowdsourcing a word-emotion association lexicon. **Computational Intelligence**, Vol. 29, No. 3, pp. 436–465, 2013.

A Appendix

A.1 Movie Script Data

Table 2 Number of scripts in each movie genre

Action	299
Adventure	194
Animation	47
Comedy	338
Crime	221
Documentary	5
Drama	536
Family	62
Fantasy	132
History	44
Horror	128
Music	23
Mystery	117
Romance	186
Science Fiction	166
Thriller	345
TV Movie	7
War	35
Western	14

A.2 Cono Application Evaluations

On a scale of 1 to 5, with 1 being a low evaluation and 5 being a high evaluation, this section shows the evaluation results of the user experiment.

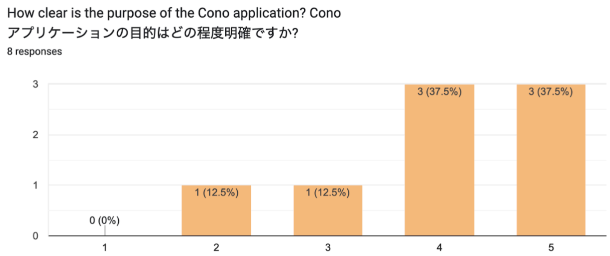

Figure 3 Evaluation of application purpose clarity

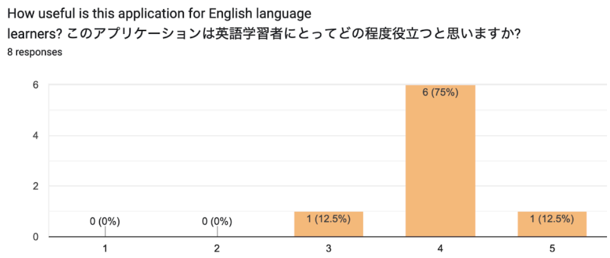

Figure 4 Evaluation of usefulness of the application for learners

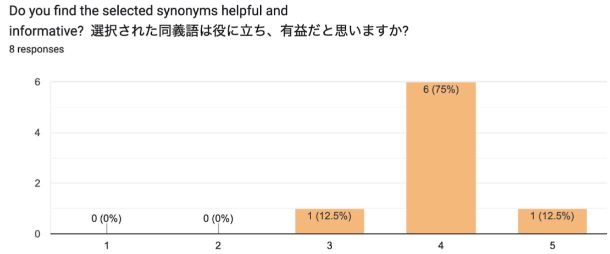

Figure 5 Evaluation of how helpful is the synonym selection

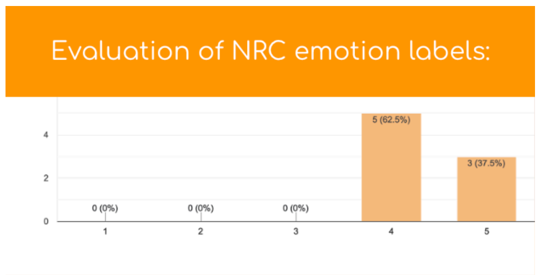

Figure 6 Evaluation of NRC Emotion Label Accuracy

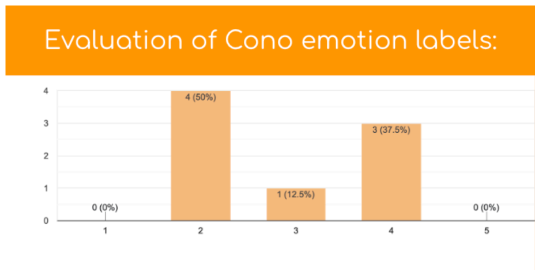

Figure 7 Evaluation of Cono Emotion Label Accuracy