Emotion classification for Japanese songs

Author：Ian Hsieh

➢ Abstract

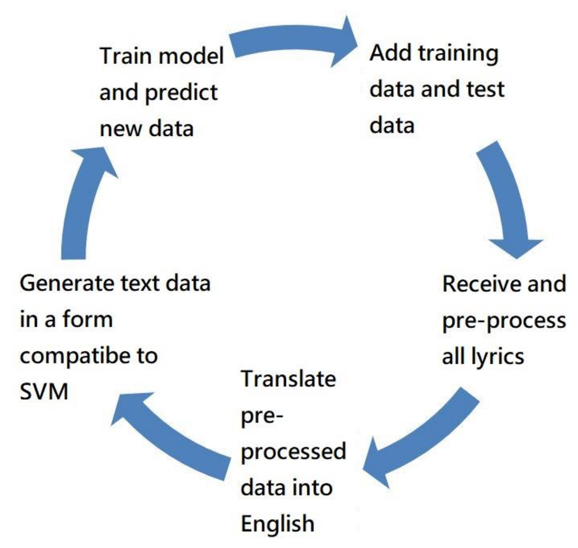
This project try to make a research in machine-translating approach, and the research is about: “Effective emotion classification for Japanese songs”. The research of emotion classification for songs has been saturated in recent years. Besides melody of songs, lyrics have been also proved as a significant indicator for classification. Based on the condition that this field has been widely researched in Chinese-English relationship, and has found out that relatively foreign cultures have different view toward lyrics of each different culture, this project mainly tries to find out whether Japanese-English relationship would reach same emotion from Chinese experimenter viewpoint.

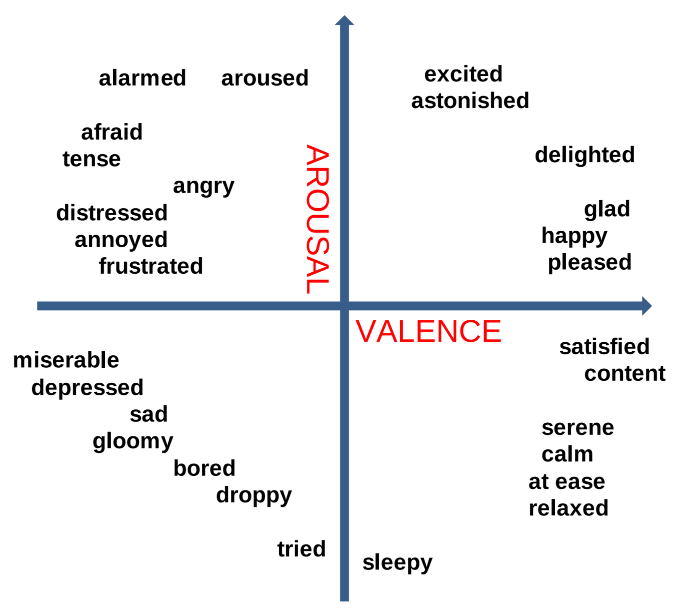
First, the system for this project will receive information of Japanese songs lyrics on certain website. There are 125 samples, 100 of them are training data, and 25 of them are test data. The system provide artists and songs name as indices of automatically receiving lyrics. Next, classify all songs manually based on valence-arousal quadrant diagram in Russell’s model[2]. Then pre-process all lyrics by translating Japanese into English, due to lots of foreign use of English in Japanese. It enhances the precision of prediction in the future.

The main approach of the project is Machine Learning. By utilizing the attributes of Support Vector Machine, it’s easy to partition lyrics into features and calculate their weight, train a model by the gold standard in self-defined 4 genres. Then gather a test data set which is not a gold standard, put the test data into the model we trained before to predict the model’s precision on emotion classification. We can find some trends after the experiment. One is the influence of cross-language attribute on the classification, and the other is the correlation between lyrics and emotion classification. By analyzing these two findings, we can determine how the research could be improved in the future.

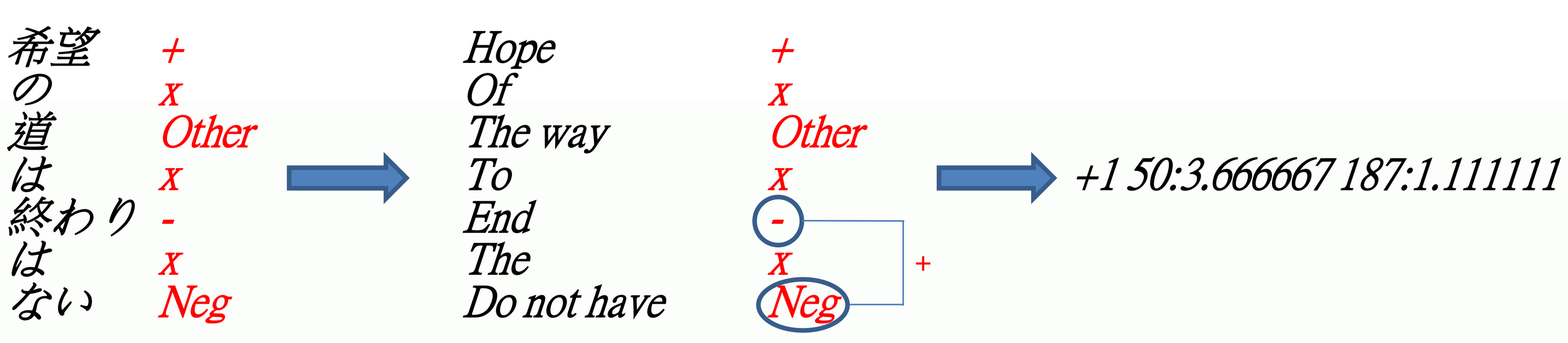
➢ Experiment details

Method





As figure in the left described, receive lyrics from “mojim.com” first by using Html Agility Pack in C# to automate the process. The XPATH of the site changes frequently, so authentication is needed in order to gain correct data. After receiving the data, mark the data into gold standard by self-defined 4 genres of Russell’s model (figure in the right). It is mainly based on lyrics to do the classification. First quadrant is tagged as positive and exciting song area, second quadrant is negative and exciting area, third quadrant is negative and calm area, and the fourth quadrant is positive and calm area.

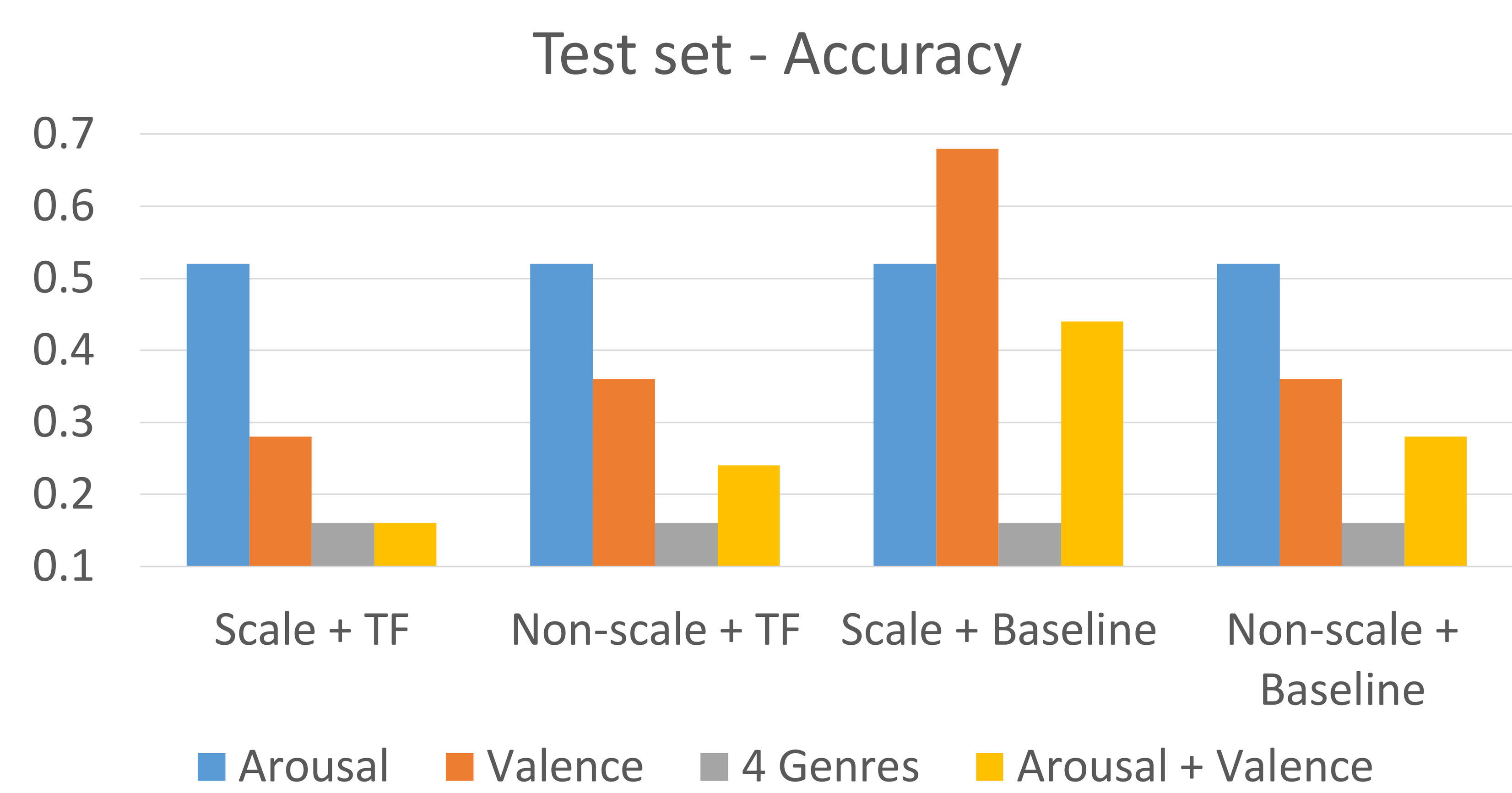


This project adopted word-level feature in SVM training. And it’s proved to be more predictive by adopting bag-of-word(unigram) approach rather than bigram approach. The diversity of songs nowadays confirms this fact. So it’s necessary to remove noise in Japanese sentence segmentation process(kuromoji). After translating into English, stemming and lemmatization are the normalization process which can restore word to its root state. Also remove stop-words which are not helpful in the classification process, like the, at, of. Then pack negative words with adjacency words as a bigram feature. The figure in upper right shows the format of LIBSVM. “+1” is the classification tag, and there’s “[Feature Number] : [Feature Weight]” structure.

$$WFO(t,c_i) = P(t|c_i)^{\lambda}\{\max\left(0, log\frac{P(t|c_i)}{P(t|\overline{c_i})}\right)\}^{1-\lambda}$$

There’s two approaches for calculating weight. One is the Baseline approach, to find emotional, positive and negative words in emotion dictionary[4][5] and put weight on them. · Neutral ones will be put a average weight on. The other is the Term-frequency approach, using the Weighted Frequency and Odds[3] formulation in the figure shown above. When λ equals 0, we get the best performance.

Result



There are four way to build prediction models. As figure shown above, 12 models are built to test the precision. The scaling is to normalize the feature weights between [-1, 1]. TF and Baseline are two main methods of calculating weights; Arousal and Valence are the specific model type built for classification. And “Arousal + Valence” simply integrated the result for 2 models. “4 Genres” directly put all 4 types into model training.

We find some trends in the result. The approach of this project is suitable for scaled data. “Scale + Baseline” and “Arousal + Valence” has the best performance, far better than 4 Genres approaches. On arousal-valence prediction, we can see good performance on valence part. We can also find out that there are few samples in this experiment, and the data sets within every quadrants are not uniformly distributed. We have many first, fourth quadrant data set in training set, but have many second, third quadrant data set in test set. Expanding emotion dictionary and improving translation quality can also help the research fit in more diverse lyrics in the future. Weight calculating method may still need detailed adaption in order to enhance the precision significantly.

Reference

[1] Abhishek Singhi and Daniel G. Brown, On Cultural, Textual And Experiential Aspects Of Music Mood, ISMIR 2014.

[2] Xiao Hu and J. Stephen Downie, When Lyrics Outperform Audio For Music Mood Classification: A Feature Analysis, ISMIR 2010.

[3] Shoushan Li, Sophia Yat Mei Lee, Ying Chen, Chu-Ren Huang and Guodong Zhou, Sentiment Classification And Polarity Shifting, Coling 2010.

[4] Minqing Hu and Bing Liu, Mining And Summarizing Customer Reviews, KDD 2004.

[5] Bing Liu, Minqing Hu and Junsheng Cheng, Opinion Observer: Analyzing And Comparing Opinions On The Web, WWW 2005.