

Emotion Recognition from Text

Computer Science and Engineering, POSTECH

Sohyeon Lee, 20140165

- Datamining Lab, prof. Hwanjo Yu

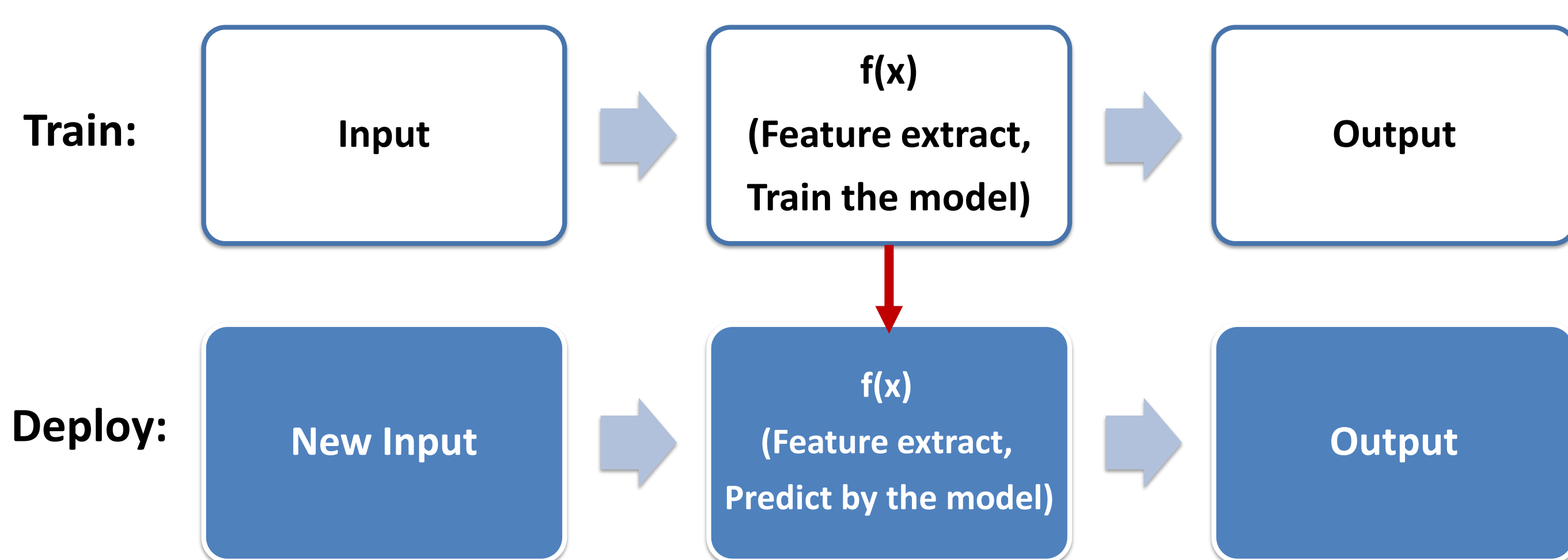
Introduction and Motivation

- Recognize emotion from text with machine learning.
- With this project,
 - We are able to understand thought or emotion about one person, product, company, accident subject, etc.
 - We are able to understand user's emotion in various kinds of SNS and messenger and develop new useful applications by using it.
 - We are able to detect and observe happiness of citizen, which depends on social, economic and environment factors.

Goal of Project

- Read text and recognize emotion with classifiers. Emotion labels include 7 kinds of emotion - Love, joy, surprise, anger, sadness, fear and neutral.
- Compare performances of all classifiers.

Overall Structure

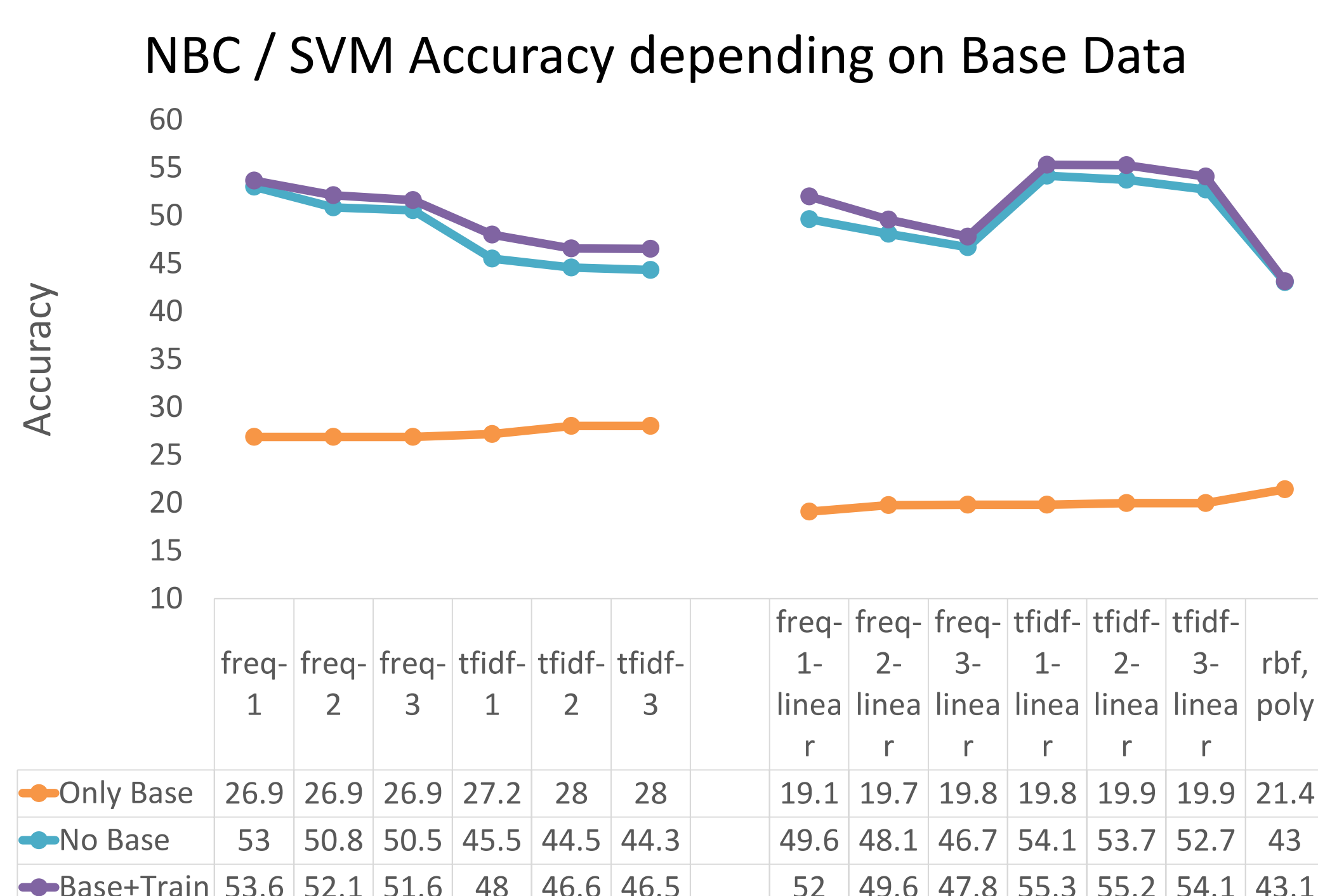


Method

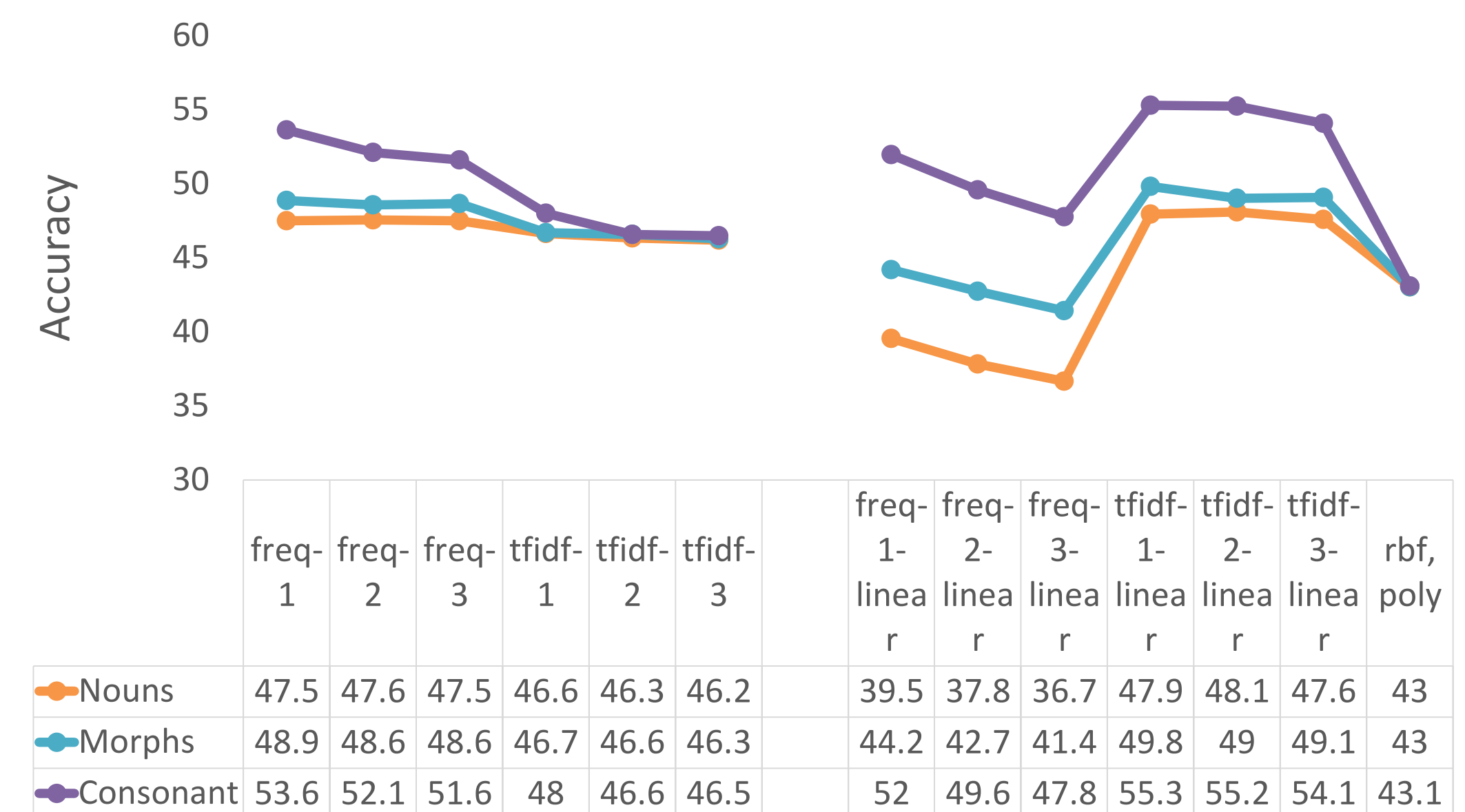
- Make base emotion data and crawl text data for train and test from Twitter.
 - Implement various kinds of textual emotion classifiers with various methods and evaluate.
- Consider **pre-processing** with Morphological analysis, etc.
 - Consider **feature extractor** with Word count, TFIDF, phrase, etc.
 - Implement **various classifiers**. - Naïve Bayes, SVM, RNN-LSTM
- Evaluate the performances
 - Calculate accuracy with 5-fold cross validation.

$$Accuracy = \frac{true\ pos + true\ neg}{true\ pos + false\ pos + false\ neg + true\ neg}$$

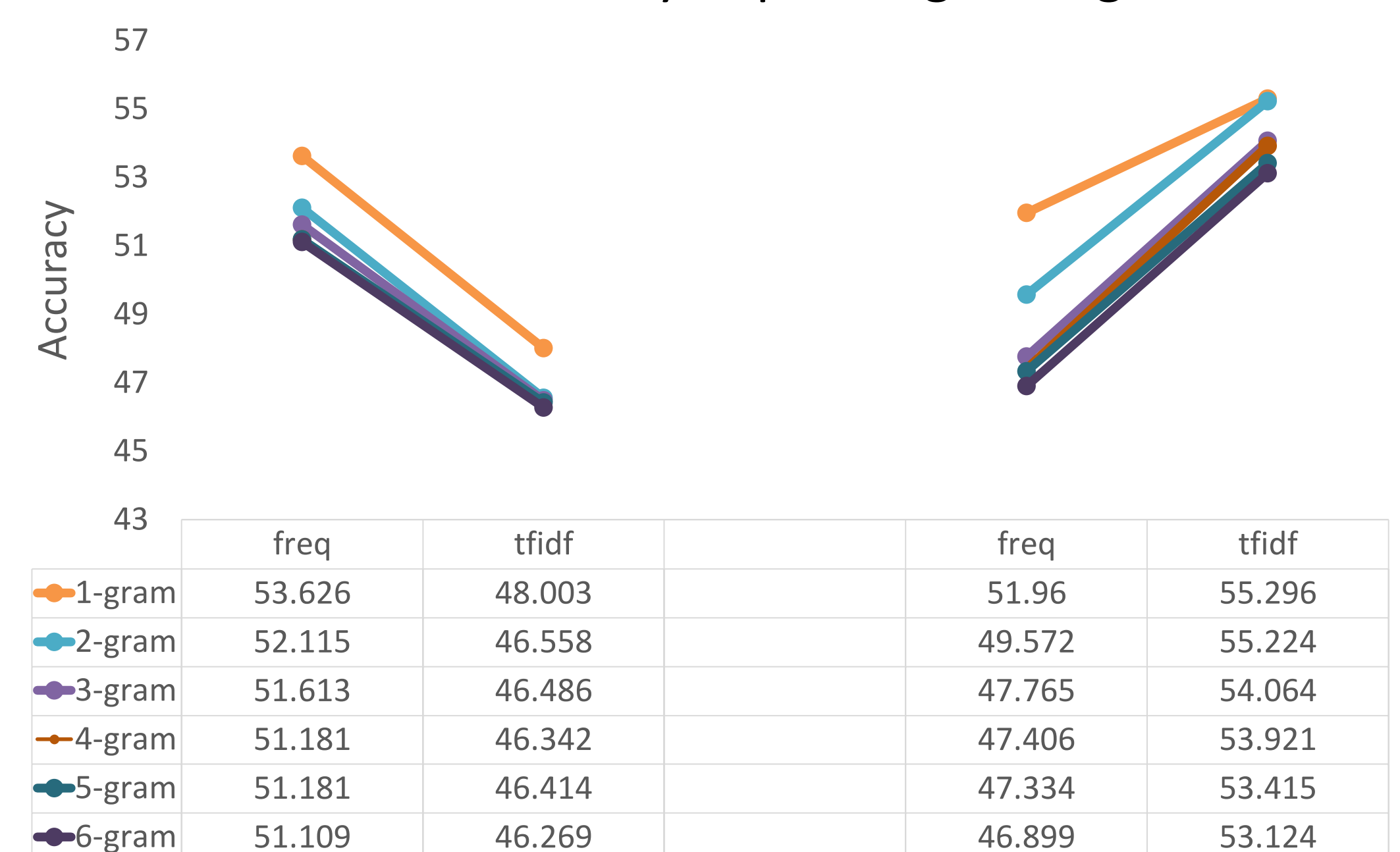
Result



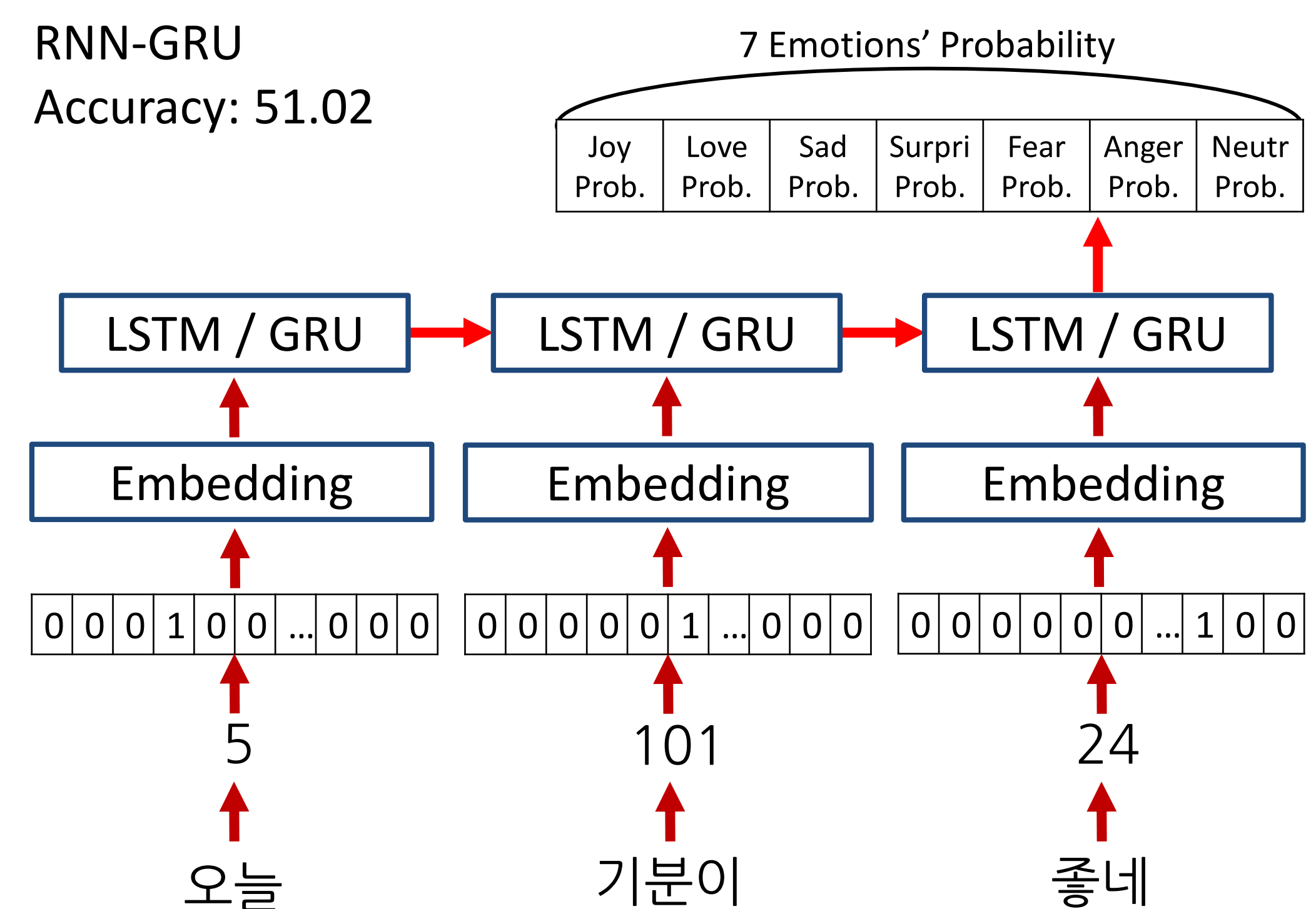
NBC / SVM Accuracy depending on Nouns, Morphs, Consonant



NBC / SVM Accuracy depending on N-gram



RNN-GRU
Accuracy: 51.02



Conclusion and Discussion

In result, 1-gram TFIDF of linear SVM is best performance classifier and second one is RNN classifier. Against expectation, RNN is worse than SVM because of not enough data. Also, 2 or 3 gram classifiers perform worse than 1 gram classifiers. It is because that there exists many meanings in one relation on words, not just one meaning.

Because machine learning is based on training the model and predicting by the model, data amount has a strong influence on the accuracy. In this project, when data is increased from 635 to 1,384, accuracy is increased by about 5 % (52.686 %→55.296 %) in 1-gram TFIDF of linear SVM.

Classifiers can be improved much more if we distinguish emoticon and punctuation mark to extract textual features.

References

- Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat, "Emotions from text: machine learning for text-based emotion prediction", 2005.
- Chetan R. Chopade, "Text Based Emotion Recognition: A Survey", International Journal of Science and Research (IJSR), 2013, pp.409-414.
- The Emotion Machine, <http://www.theemotionmachine.com/classification-of-emotions>.
- Microsoft, <https://www.microsoft.com/developerblog/real-life-code/2015/11/30/Emotion-Detection-and-Recognition-from-Text-using-Deep-Learning.html>.