Utilizing Figurative Language Examples for Recognizing Japanese Sentences with Metaphors

Mateusz Babieno Rafał Rzepka Sho Takishita Kenji Araki
Graduate School of Information Science and Technology, Hokkaido University

1. Abstract

In this paper, we introduce results of a classification experiment designed to recognize sentences containing metaphors as the first step of recognizing figurative expressions in Japanese text. For the experiments we have utilized existing set of figurative expressions and constructed one which should consist of mostly literal sentences. The former consists of almost 10,000 unique entries from a simile dictionary used in previous studies [1]. The latter was created by retrieving data from three different sources: Japanese Wikipedia, local assembly minutes and news articles. After testing three basic text classification methods (Naive Bayes, Support Vector Machines and an Artificial Neural Network) we confirmed very high precision and recall (94-98% F-score depending on the training-testing data size ratio) achieved by all models with slightly higher results for SVM classifier. The experiments suggest that the simile examples can be successfully used for extracting sentences containing figurative expressions for further research on Japanese metaphor recognition and generation.

2. Background

Contrary to back then popular view, metaphors have been proven as not mere rhetorical figures used almost solely in poetry but rather as a profound cognitive mechanism construing large part of any given utterance and widely noticeable in everyday language. Metaphorical expressions, or to put it more broadly, figurative language, is so prevailing in our daily communication, that people can hardly distinguish it from literal language. Probably most explicitly expressed by Lakoff and Johnson in their Metaphors We Live By [2], this discovery has become a turning point in scholarship, mainly linguistics. From that point on, vast number of research papers and whole books have been dedicated solely to the analysis of this phenomenon.

Partly thanks to all the attention given it by general linguistics, metaphor analysis has already become quite a popular topic among scholars dealing with natural language processing. The largest part of research concerning the figurative language is performed on European languages, mostly English and utilizes rules [3]. Bulat et al. [4] has recently proposed the first metaphor identification method using representations constructed from embedding-based property norms (approx. 75% accuracy). Tsvetkov and colleagues [5] have shown that their method using subject-verb-object and adjective-nouns can be used for English, Spanish, Farsi and Russian. When it comes to Japanese language, probably the first trial to automatically draw a line between idiomatic and literal meanings was made by Hashimoto et al. [6] who proposed a lexical knowledge base for idiom recognition and achieved 90% accuracy. Dybala et al. [7] have concentrated on metonymical comparisons and their web-corpus based method showed that occurrences of source and target pairs is able to achieve 74% accuracy. Yoshimura et al. [8] have proposed a method for recognizing similes by using sensory associations and their method’s accuracy was 67.5%. To the authors’ best knowledge, this work is the first trial for utilizing statistical methods for figurative language recognition. Although they currently recognize only sentences with metaphorical expressions, the results seem to be very promising for acquiring means for big-data approaches to figurative language processing.

3. Datasets for Classification

The dataset used in our experiment comprises two large subsets: one of the figurative expressions and the other of the literal ones. The former was created using a dataset from [1], which limits a large metaphor dictionary examples to similes. After preprocessing, we gained 9,152 unique expressions. Besides removing duplicates, we excluded noise: punctuation marks and content in brackets. The non-metaphorical set comprises material taken from three different sources, namely the current Japanese Wikipedia dump [9], local assembly minutes [10] and articles from Livedoor News [11]. In this case preprocessing included erasing XML tags, removing all ASCII characters (as this encoding standard does not include Japanese characters), punctuation marks, and sentences including less than 6 or more than 86 characters (this is the length of respectively the shortest and the longest sentence in our metaphorical dataset and we wanted to limit surface differences between both sets to minimum). We then took 30,000 random expressions from each set, and concatenated those, which gave us the sum of 90,000 non-metaphorical expressions from three different sources. The final step was to randomly choose 9,152 of such expressions, which is the number of items in corresponding metaphorical dataset.

<table>
<thead>
<tr>
<th>Data size ratio (train:test)</th>
<th>Measure</th>
<th>Naive Bayes</th>
<th>SVM</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%:10%</td>
<td>Precision</td>
<td>0.976</td>
<td>0.987</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>0.904</td>
<td>0.988</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td>F-score</td>
<td>0.939</td>
<td>0.987</td>
<td>0.982</td>
</tr>
<tr>
<td>80%:20%</td>
<td>Precision</td>
<td>0.980</td>
<td>0.987</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>0.892</td>
<td>0.980</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>F-score</td>
<td>0.934</td>
<td>0.984</td>
<td>0.980</td>
</tr>
<tr>
<td>70%:30%</td>
<td>Precision</td>
<td>0.984</td>
<td>0.989</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>0.899</td>
<td>0.983</td>
<td>0.978</td>
</tr>
<tr>
<td></td>
<td>F-score</td>
<td>0.939</td>
<td>0.986</td>
<td>0.982</td>
</tr>
</tbody>
</table>

P-value ≤ 0.05
4. Experiment and Results

Classification was performed in Python and scikit-learn library was utilized [12]. We conducted three sets of experiments, subsequently changing the proportion of original data incorporated in test and training subsets for each method. The proportion was set as respectively 10%, 20% and 30% of the whole set taken as test data. We have used standard state-of-the-art models, namely Naïve Bayes (multinomial), Support Vector Machines (Linear SVC) and Artificial Neural Nets (multi-layer perceptron with stochastic gradient descent, 10,000 max. iterations). As shown in Table 1, SVM acquired the best results independently of the corpora size (which seems to be rather typical in case of classification performed on two groups with little noise in data). In comparison with the neural network, SVM is not only slightly more efficient but also much faster which can be crucial when dealing with larger amount of data (5-fold cross-validation was used; experiments were performed on 2010 MacBook Pro, 2.53 GHz Intel Core i5 processor with 8 GB memory).

5. Error Analysis and Considerations

The dictionary examples-based and the non-metaphorical expression sets of our choice appeared to build a precise model for figurative language recognition. For the first trial we have chosen news, Wikipedia articles and minutes to produce less errors and simplify analysis of the erroneous classifications. Below we provide two examples of common mistakes. As originally labelled as literal but predicted as metaphorical was for example the following sentence: カメムシの大群は、日々を追って次第に移動していたようである(It seems that large swarms of stink bugs have changed locations as the time went by). Japanese original from Wikipedia entry includes an idiom to follow the sun, which is a metaphorical expression for elapsing days. An opposite case example is: 昼寝の最中に声をかけられた女みたいにどうくなり見つめる(Gaze at sth as a girl called out in the middle of the afternoon nap), where a figurative expression was mislabeled as non-metaphorical. In both examples Japanese grammatical clues for similes (ようである/みたいに; as/like) are present, therefore to test the role of semantic features one should test the data after limiting the data to e.g. nouns, verbs and adjectives.

5. Conclusions and Future Work

Over 90% precision acquired by all tested models proves that classification of figurative and literal language can be performed successfully without resorting to human-made annotation. Promising results listed above encourage us to use SVM model in the future experiments, which are going to be conducted on less homogeneous data, however the results might be better for the artificial neural network when the input contains more noise. Our ultimate goal is to create a powerful metaphor generation model and designing an efficient classifier similar to the one described in this paper should be treated as an important step towards acquiring vast number of figurative language examples. This could also help building a large repository of metaphorical expressions and sentences including figurative language similar to Amsterdam Metaphor Corpus [13]. Such a contribution could provide a precious source for researchers who want to study or process Japanese figurative language.

For the near future work, we plan to try to implement older [14] and latest [16] methods reported as useful for other languages, mainly English. One of them is a novel method for identification of metaphors with word vectors introduced in [17] and it seems to be less dependent on a given language.

References


