Exploratory investigation of word embedding in song lyric topic classification: promising preliminary results

Kahyun Choi
School of Information Sciences
University of Illinois at Urbana Champaign
ckahyu2@illinois.edu

J. Stephen Downie
School of Information Sciences
University of Illinois at Urbana Champaign
jdownie@illinois.edu

ABSTRACT

In this work we investigate a data-driven vector representation of word embedding for the task of classifying song lyrics into their semantic topics. Previous research on topic classification of song lyrics has used traditional frequency based text representation. On the other hand, empirically driven word embedding has shown sensible performance improvment of text classification tasks, because of its ability to capture semantic relationship between words from big data. As averaging the word vectors from a short text is known to work reasonably well compared to the other comprehensive models utilizing their order, we adopt the averaged word vectors from the lyrics and user's interpretations about them, which are short in general, as the feature for this classification task. This simple approach showed promising classification accuracy of 57%. From this, we envision the potential of the data-driven approaches to creating features, such as the sequence of word vectors and doc2vec models, to improve the performance of the system.

CCS CONCEPTS

• **Information systems → Content analysis and feature selection**;

KEYWORDS

Song lyrics, Subject, Topic, Metadata, Word embedding, Classification

ACM Reference Format:
Kahyun Choi and J. Stephen Downie. 2018. Exploratory investigation of word embedding in song lyric topic classification: promising preliminary results. In *Proceedings of IEEE/ACM Joint Conference on Digital Libraries (JCDL2018).* ACM, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

The underlying topics of song lyrics have long been considered as useful metadata for the browsing and searching songs [5]. However, due to the poetic nature of lyrics it is difficult to label the songs based on their topics even for humans especially when it comes to a digital music library with big dataset. This inherent difficulty of creating a labeled dataset calls for an automatic topic labeling/classification system. However, once again training such a supervised learning system is challenging as well for the same reason: understanding the meaning of song lyrics is difficult for a machine learning algorithm, too. This problem was first addressed in [2][3], where user interpretations in addition to song lyrics were proposed as a new feature for this classification task. Because user interpretations contain richer topic-related information in a more straightforward form compared to the ambiguous song lyrics, the classification systems using user interpretations as well as the lyric text showed superior performance to its counterpart based solely on the lyrics. To this end, they collected a large amount of comments from the website, `songmeanings.com` where music lovers discuss about song lyrics and share their interpretations about them.

This paper uses a more advanced way to extract some higher level features from both the song lyrics and user interpretations. Instead of the primitive Term-Frequency (TF) representation that the preceding topic classification systems were based on, we propose to use a more advanced word embedding representation, `fastText` [1]. While the TF representation can effectively describe the simple statistics of the bag-of-words representation of a document, the independent and sparse nature of the elements of the TF vector requires the feature to be too high dimensional. Word embedding techniques can address this issue by learning a model that converts a word into a dense vector representation, which should subsequently be able to recover the other neighboring words [6]. In this way, the vector representation can encode the co-occurrence information of the words spread in the large corpus. Often, this model is represented as an artificial neural network, and consequently a deep neural network architecture is also introduced to learn this embedding more effectively.

We test out `fastText` to create the vector representation of words used in song lyrics and user interpretations. By using them we create feature vectors that are known to be more manageable and semantically meaningful. Eventually, topic classification tasks benefit from these new features.

2 EXPERIMENTS

2.1 Datasets and Preprocessing

We follow the same experimental setup used in [3]. The dataset of 800 popular songs was labeled by experts in `songfacts.com`. The balanced dataset has 8 categories in total: places, sex, ex-lover, drugs, war, parent, religion, and death. Song lyrics and user comments are collected from `songmeanings.com` where millions of user comments about meanings of song lyrics are posted.

For the word embedding, we used the word vectors trained on English Wikipedia using `fastText` [1]. The pre-trained vector is

JCDL2018, June 2018, Forth Worth, Texas USA

ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

	TF	Weighted Average of Word Vectors	Average of Unique Word Vectors
Accuracy	59.00	56.60	52.00

Table 1: Classification accuracy

trained based on the skipgram model [6] and has 300 dimensions. Because this model utilizes sub-word information, it showed better performance than others in multiple tasks including human similarity judgment and word analogy tasks than `word2vec` models. As an already trained neural network model `fastText` predicts the 300 dimensional vector representation of a given word through its ordinary feedforward process.

Song lyrics and user comments were broken down into words and only alphabetic letters were kept to remove irrelevant words including typos and user IDs. Next, the words were lemmatized to group words with the same meaning and common stopwords in English were deleted.

2.2 Feature Representations and Classification

Among many classifiers, we used naïve Bayes for this preliminary experiments since it performs almost as good as Support Vector Machines (SVM) in the previous study [3] while it is faster than SVM. Considering the small size of the dataset, we excluded deep learning-based classification algorithms, although we believe that they could perform better than other classifiers once a bigger dataset is available for training. We performed 10-fold cross validation and reported the average accuracy.

Since `fastText` produces a vector per word, a document, i.e. lyrics of a song or a user comment, is represented with a sequence of vectors, whose length varies depending on the number of words. Instead of treating this sequential input data as it is, we use their weighted sum as the representative word vector, where weights are proportional to the frequency of the words. This proposed feature aggregation is based on the observation that averaging word vectors work reasonably well when classifying short texts [4].

3 RESULTS

Table 1 compares accuracy of classification systems with different feature representations. The classifier based on term frequency yielded the best accuracy, 59% [3]. When the word vectors were averaged while considering frequency of each word, the classifier performed pretty well with 56.6% accuracy. On the other hand, when frequency of words were ignored, the classification accuracy dropped to 52%. This indicates that the frequency of words still plays a role in topic classification. The confusion matrices also show how categories are confused with each other. The classification based on TF penalized "places" heavily (Figure 1) while one based on word vectors favors the category (Figure 2).

4 CONCLUSION

We took the first step towards utilizing word embedding for lyric topic classification task. Because lyrics and comments are short, averaging the word vectors kept topic information quite a lot and performed well on the topic classification tasks. However, we believe that more sophisticated ways to describe song lyrics and the user interpretations can improve the classification performance. Taking sequences into account or applying `doc2vec` are among them.

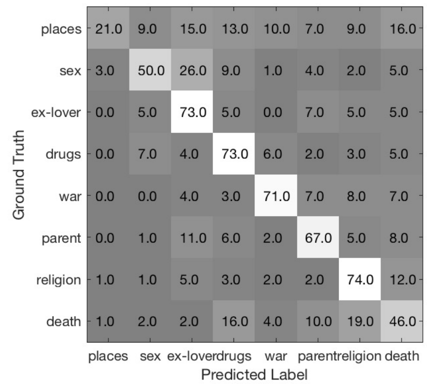

Figure 1: Confusion matrix of the classifier with TF

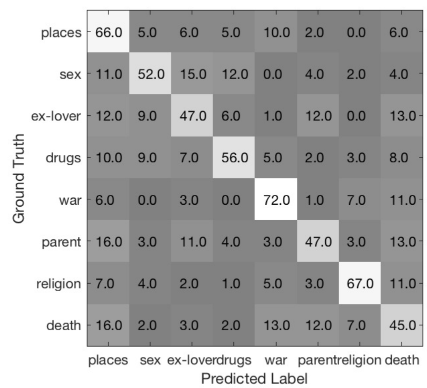

Figure 2: Confusion matrix of the classifier with averaged word vectors where frequency of words were allowed

REFERENCES

[1] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching Word Vectors with Subword Information. *arXiv preprint arXiv:1607.04606* (2016).

[2] Kahyun Choi, Jin Ha Lee, and J Stephen Downie. 2014. What is this song about anyway?: Automatic classification of subject using user interpretations and lyrics. *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries* (2014), 453–454.

[3] Kahyun Choi, Jin Ha Lee, Xiao Hu, and J Stephen Downie. 2016. Music Subject Classification Based on Lyrics and User Interpretations. *Proceedings of the American Society for Information Science and Technology* (2016), 1–10.

[4] Cedric De Boom, Steven Van Canneyt, Thomas Demeester, and Bart Dhoedt. 2016. Representation learning for very short texts using weighted word embedding aggregation. *Pattern Recognition Letters* 80 (2016), 150–156.

[5] Jin Ha Lee and J Stephen Downie. 2004. Survey Of Music Information Needs, Uses, And Seeking Behaviours: Preliminary Findings. *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)* 2004 (2004), 441–446.

[6] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).