Image Recommendation for Social Media

Web Retrieval and Mining Final Project

王思傑/資工四/b02902010
曹爗文/資工四/b02902030
CSIE, NTU, Taipei, Taiwan

黃晴/資工四/b02902042
楊尚珉/資工四/b02902067
CSIE, NTU, Taipei, Taiwan

***Abstract*—There are a variety of social medias available through the Internet. Some of them are pure texts, while others come along with several pictures. This report is aimed to figure out a mechanism to recommend images to social media users who lack handy beautiful pictures. We conquered the task by implementing Neural Networks and word embedding. First, convert the context to word embedding. Then the neural network would classify the text into five distinct sentiment classes, including anger, joy, sadness, love and fear. We pick up the sentiment class and some specific keywords from the text as the query sentence for Google Image Search. Finally, we evaluate our work through questionnaires.**

Keywords—Image Recommendation, Sentiment Analysis, Neural Network.

I. Introduction

Social Media has been an inevitable trend nowadays. Instagram, Twitter and Facebook users worldwide contribute billions of online texts. Some social media posts require users to upload a picture. Yet the lack of suitable pictures would attenuate the desire of using the social media. A solution to the situation is through the help of image recommendation. Imagine that you are trying a new social media, and this system would analyze your text in the post just to recommend awesome pictures. Just a simple click among the candidates, and a picture would be appended to the post. According to our survey, **___**% of users would be glad of the new breakthrough. The report is in the form of short paper. We would first introduce the preprocessing of the text and data collection. After that, detailed implementation would be given in Method. Evaluation would be presented in Section 4, and Conclusion and Helpful Links would be at the end of the paper.

II. PREPROSSING AND DATA

A. DATA

We applied two datasets together in this job. The first data source is given in Helpful Links [1], and we select "Sentiment Analysis: Emotion in Text" data. There are 13 classes and 40000 data in total. The 13 classes are: Empty, Sadness, Enthusiasm, Neutral, Worry, Surprise, Love, Fun, Hate, Happiness, Boredom, Relief and Anger. The number of corresponding data number is given in Table I. In our work, we merge 13 classes to 5. We eliminate Neutral first, and then put Anger and Hate to one class, and Happiness, Surprise, Relief, Fun and Enthusiasm to another. Sadness, Boredom, and Empty would be a class, while Love and Worry go independently. Therefore, the first dataset would be classified to five classes after the work.

For the second dataset, we select "International Survey on Emotion Antecedents and Reactions (ISEAR)". The second dataset owns 7 tags: Joy, Fear, Anger, Sadness, Disgust, Shame and Guilt. The number of corresponding data number is given in Table II. We put Joy and the previous class Happiness together, while Fear goes with Worry. Anger and Disgust merge with previous Anger, and Shame and Guilt go with Sadness. The number of corresponding data number after merging is given in Table III, and number of training and testing data is 28377 to 3155.

TABLE I. SENTIMENT DATASET

Emotion Type	***Number of data***
Empty	827
Sadness	5165
Enthusiasm	759
Neutral	8638
Worry	8459
Surprise	2187
Love	3842
Fun	1776
Hate	1323
Anger	110
Happiness	5209
Boredom	179
Relief	1526

a. Number of Data for Each Emotion

TABLE II. ISEAR DATASET

Emotion Type	***Number of data***
Joy	1092
Fear	1093
Anger	1094
Sadness	1094

Emotion Type	*Number of data*
Disgust	1094
Shame	1094
Guilt	1091

b. Number of Data for Each Emotion

TABLE III. OUR DATASET AFTER MERGING

Emotion Type	*Number of data*
Anger, Disgust	3621
Joy, Happiness	12548
Sadness, Guilt, Shame	9450
Love	3842
Fear, Worry	9552

c. Number of Data for Each Emotion

B. TEXT PREPROCESSING

We apply the Twitter version of GloVe [3] as our word embedding. Then select 100 as the word dimension. We use the zero vector for the words not in GloVe.

III. METHOD

Our main framework about the problem is a mixture model based on Sentiment Analysis, TF-IDF keywords and Google Image Search. We take the result of Sentiment Analysis and TF-IDF keywords as a query, and deliver it to Google Image Search. For Sentiment Analysis part, we applied 3 Recurrent Neural Networks. The first network is to bi-classify whether the text emotion is positive (Joy, Love) or negative (Sadness, Anger, Fear). After that, two networks are specific to classify the sub-classes of positive and negative emotions respectively. Each network shares the same construction: one middle layer with 200 LSTMs, and we choose the class with the highest probability. Accuracy of three networks is given in Table IV.

TABLE IV. OUR DATASET AFTER MERGING

Network Type	*Accuracy*
Bi-classify (Positive/Negative)	0.7908
Positive subclass classification (Joy/Love)	0.8060
Negative subclass classification (Sadness/Fear/Anger)	0.6088

d. Accuracy for Each Network

After Sentiment Analysis, the emotion of text is predicted by the network. Then we select 3 keywords of the text through TF-IDF procedure. After the whole process, we regard the sentiment and three keywords as query, and deliver the query to Google Image Search. Lack of ground truth lead to human evaluation for our work, and this would be described in the following section.

IV. EVALUATION

Absence of ground truth of the recommendation triggers the design of questionnaires. The link of our questionnaire [4] is provided. The survey consists of into three parts. First of all, we calculate the willingness of owning such a system. 54 out of 59 people are glad to be recommended, and which leads to the ratio of 91.5%. Secondly, we compare the result of Image Search with sentiment and without sentiment. Questioners are given a context and two recommended images, which stand for the two scenarios mentioned. Questioners would select the image which he/she thinks is more related to the context. Four contexts in total brings out the result Table V. In average, 72% people think the result is better with the appending of sentiment.

TABLE V. RATIO OF BETTER PERFORMANCE WITH SENTIMENT

Question	*1*	*2*	*3*	*4*	*Avg.*
Ratio	89.8%	84.7%	61%	52.5%	72%

e. Ratio of Better Performance with Sentiment

The last part of the survey presents 5 search results of our system and questioners score the performance between 1-10. Analysis and details are in Table VI.

TABLE VI. SCORE OF THE PERFORMANCE OF OUR SYSTEM

Question	*1*	*2*	*3*	*4*	*5*	*Avg.*
Score	6.07	5.68	7.20	7.44	8.15	6.91

f. Score of performance

CONCLUSION

The recommendation of images for social medias is implemented through Sentiment Analysis, TF-IDF keywords and Google Image Search. We confirm the better performance of the system with the Sentiment Analysis Recurrent Neural Network. Future work would be consolidating the structure and accuracy of RNN. Evaluation implies the above average performance.

HELPFUL LINKS

[1] Data Source 1: “Sentiment Analysis: Emotion in Text”: https://www.crowdflower.com/data-for-everyone/.

[2] Data Source 2: International Survey On Emotion Antecedents And Reactions (ISEAR) directed by Klaus R. Scherer and Harald Wallbott: http://www.affective-sciences.org/en/home/research/materials-and-online-research/research-material/.

[3] GloVe: https://nlp.stanford.edu/projects/glove/.

[4] Questionnaire: https://docs.google.com/forms/d/e/1FAIpQLScd8paiZ1ToFLuoaECJ6gI-GhYO-O5NaiBxmedu9F41iQxb6w/viewform.