Chinese Sentence Meme Matcher Using Deep Neural Networks

Ying-Jung Chien, Chu-Chun Chang, Ting-Ru Liu, Jou-Min Liu, Ling-Yu Huang

***Abstract*—Although there are many English meme text generators, there is no matcher for matching memes and Chinese sentences. Therefore, we decided to construct a meme matcher, which takes Chinese sentences as input, and outputs the best matching meme photo. We have tried six different kinds of DNN models, including CNN, GRU, Bidirectional LSTM, Bidirectional GRU, and LSTM + CNN. Among these models, CNN has the best performance. After fine-tuning, our meme matcher can correctly classify most memes and is useful for people who want to pick a suitable meme instantly.**

I. INTRODUCTION

From time to time, people want to post a meme to a specific scenario. And it can be difficult for humans to find a suitable meme photo to match the scenario or mood immediately because there are too many memes to choose from. So we construct a meme matcher to find the right photo to help people make memes and speed up the rate at which people make funny memes. Since describing the scene or expressing the current emotion in words is more intuitive for humans, the method we propose takes Chinese sentences as input, tokenizes the input sentence by jieba, then classifies the label matching the tokenized data by CNN, and finally outputs the best matching meme photo corresponding to the input.

Input: 沒有搶到免運的你

Output: label 1

Figure 1: An example of our meme matcher. The input is a Chinese sentence, and the output is the label of the best matching meme photo.

II. METHODS

A. Dataset

The data we used is crawled from the website named Meme Warehouse: https://memes.tw/, which is a website where many netizens share meme photo templates and upload their meme creations.

We choose these 10 images to train the model for two reasons. The first reason is that they're popular so we can have enough data in the dataset. And the second is that they express different emotions from each other, thus reducing the ambiguity of classification. For these 10 pictures, we crawled the sentences in the meme creation shared by netizens as data and saved them as csv files. Then, we manually removed some of the sentences we crawl, such as advertisements, non-Chinese sentences, political spam, etc.

Figure 2: These are the 10 images we picked.

主人只給我吃10包貓飼料	1
解題課業上的數學 看不懂	1
我真的好想放棄	1
我能怎麼辦 我也很無奈啊	1
我沒有錢	1
對不起我上班會乖乖	1
看完進巨最後一集的我:	1
小時候當魚刺卡在喉嚨時:	1
啊～～～ 誰能來救救我啊	1
抱歉了錢錢 但我真的需要那個酷東西	1

Figure 3: An example of sentences and corresponding image and label.

B. Data preprocessing

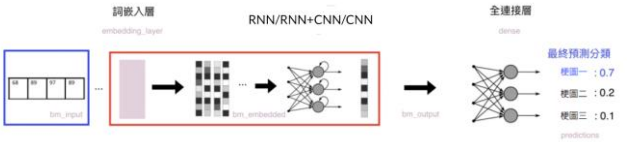
After we've got all the data we need, we start doing data preprocessing. First step is to tokenize. We tried both ckiptagger and jieba, and decided to use jieba. Jieba is one of the best Python Chinese word segmentation modules, and it uses dynamic programming to find the most probable combination based on the word frequency. After getting the

影音攝影剪輯師 Video Editing Designer 全職 · 每月 NT$28,000 - NT$38,000 【工作內容】 **投遞履歷請提供相關作品參考**
強逼人民硬吞萊豬核食謀誰之利？孫大千質疑蔡政府：叛國賣台 蔣萬安個性太像馬英九 名嘴曝2022藍營第一戰將 萊豬爭議未平核食接力？ 羅智強：公投若沒過大家等著吃核食 我就油
注意！功能更新！ 勁爆好用！ 文字底色功能上線！ 方便做出更豐富的圖喔！ 試試看喔！ 可以追蹤站長喔～ 第一時間得知新消息～
拜託各位按讚吧 我想看100讚後會發生什麼事 達標之後就會刪這張圖 所以沒有盈利 按讚不用錢
淺井 大樹 おつかれさまでした！！！！！ お好み焼き食べながら生放送見てます

Figure 4: Some examples of the removed sentences.

satisfying result of tokenization, we put Texts to sequences with max number word ten thousand.

C. Model



Convolutional Neural Network (CNN)

CNNs are very similar to ordinary Neural Networks — they are made up of neurons that have learnable weights and biases. The name 'convolutional neural network' indicates that the network employs a mathematical operation called convolution. Convolutional networks are a specialized type of neural networks that use convolution in place of general matrix multiplication in at least one of their layers. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity.

Long Short-Time Memory (LSTM)

LSTM networks are a special kind of RNN, which is capable of learning long-term dependencies. LSTMs have a key structure called cell state, which is like a conveyor belt. It runs straight down the entire chain, with only some minor linear interactions.

The LSTM does have the ability to remove or add information to the cell state, carefully regulated by structures called gates.

Gates are a way to optionally let information through. They are composed out of a sigmoid neural net layer and a pointwise multiplication operation. An LSTM has three of these gates, to protect and control the cell state..

Gated Recurrent Unit (GRU)

GRUs are a gating mechanism in recurrent neural networks. The GRU is like a LSTM with a forget gate, but has fewer parameters than LSTM, as it lacks an output gate. GRUs have been shown to exhibit better performance on certain smaller and less frequent datasets.

Bidirectional Long Short-Time Memory (LSTM/ GRU)

Bidirectional recurrent neural networks (BRNN) connect two hidden layers of opposite directions to the same output. With this form of generative deep learning, the output layer can get information from past (backwards) and future (forward) states simultaneously.

LSTM + CNN

The CNN LSTM architecture involves using Convolutional Neural Network (CNN) layers for feature extraction on input data combined with LSTMs to support sequence prediction.

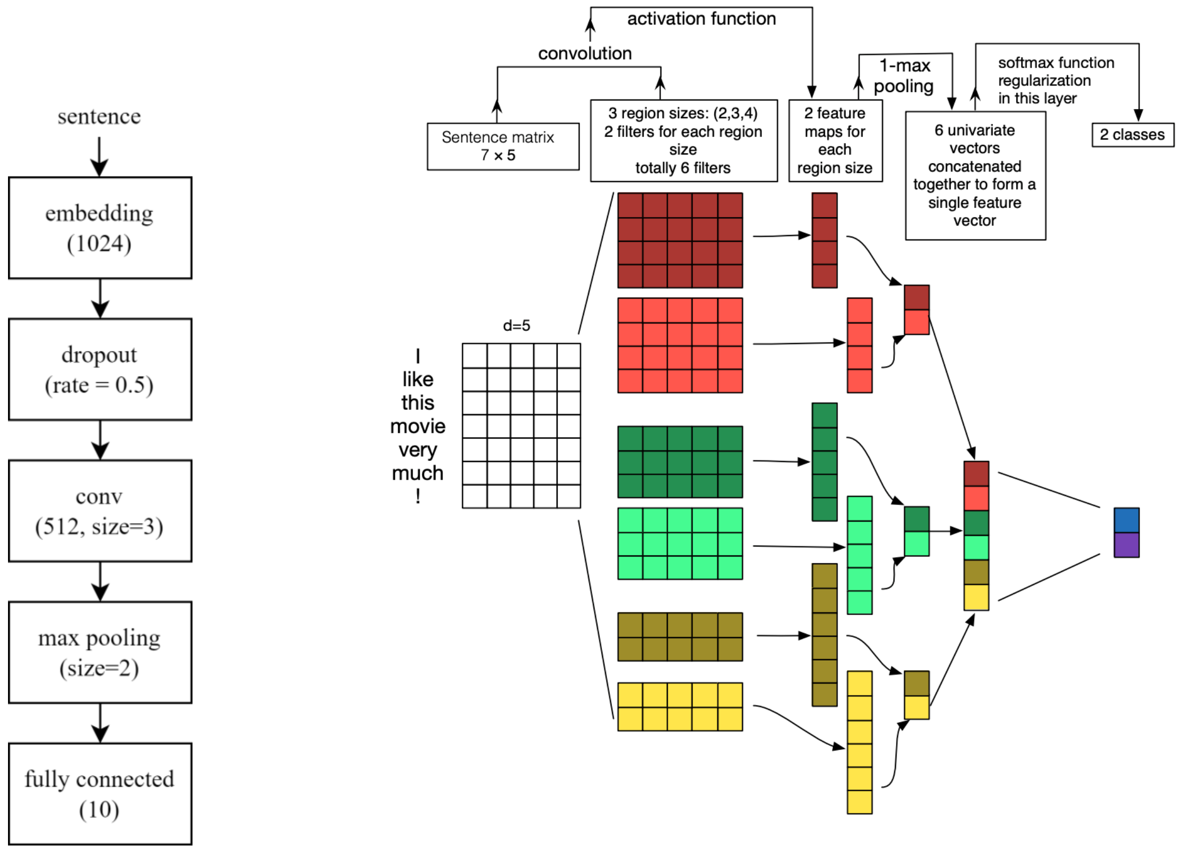


Figure 5: Our CNN model architectures

In accuracy evaluation, we find that CNN has outstanding performance among all these different approaches. Therefore we then focus on building CNN model.

In our model, after embedding the layer, we reshape the data and put it into a one dimensional CNN model with a single layer, since it actually performs remarkably better than other more complicated multi-layer models.

Then we shaped data into a matrix, and added convolutional layers with different kernel sizes. The concept of the kernel size is like an n-gram model in natural language. The model focuses on n words each time, so the kernel size would be n multiply embedding dim. After each CNN convolutional layer with different n, we do max pooling with size two respectively and concatenate them together. Flatten layer and dropout layer are then added to receive the result of the max pooling layer. Finally, put it into a dense layer with a softmax activation function, and the dense layer classifies our result into one of the ten meme pictures.

Our CNN model is shown in Figure 5.

III. Results

Training parameters	Value
Optimizer	Adam
Learning rate	0.001
Loss function	Categorical cross entropy
Activation function	ReLU

Table 1: the training parameters in our CNN model

The training parameters in our model are listed in Table 1. Finally, the model achieves 63.11% validation accuracy.

IV. Discussion

From Table 2, we can see that the models with CNN perform better than the others. There are several factors that might be the reason for our result. Yin et al. showed that when sentence lengths are small, the performance of CNN and RNN are comparable.[3] The length of a meme sentence is usually shorter compared to articles. Besides, CNN is good at extracting position-invariant features, so it performs better on tasks like keyphrase recognition, which our meme matcher is doing.

Further observation

When the embedding dimension is smaller than 128, the performance of CNN model drops dramatically. On the other hand, the accuracy didn't change much when we increased the embedding dimension from 512 to 1024.

Model	Accuracy (%)	Learning rate	Batch
LSTM	57.92	0.001	128
GRU	49.18	0.003	128
Bidirectional LSTM	55.19	0.001	128
Bidirectional GRU	53.01	0.003	128
LSTM + CNN	61.59	0.001	128
CNN	63.11	0.001	128

Table 2: Best results or CNN, GRU, Bidirectional LSTM, Bidirectional GRU, LSTM + CNN in Meme classification.

Sentence length has a relatively small impact on performance. Since CNN performs a keyphrase recognition, as long as the keyphrase of the sentence is included in the sentence length, smaller sentence length can perform as well as larger sentence length.

V. Conclusion

In this work, we run extensive experiments for sentence classification. Our results add to the well-established evidence that a simple CNN with one layer of convolution performs remarkably well for the classification task in NLP. Our meme matcher can achieve good performance for classifying most memes and is convenient for people who want to find a matching meme instantly.

VI. Author Contribution Statements

Y.J.C. (20%): data analysis, study design, programming of the artificial neural network, and writing.
C.C.C. (20%): data collection, data analysis, study design, writing, and final presentation.
T.R.L. (20%): data analysis, study design, writing, and programming of the artificial neural network.
J.M.L. (20%): data collection, data analysis, writing, and study design.
L.Y.H. (20%): data analysis, study design, final presentation, writing, programming of the artificial neural network.

VII. Reference

[1] Zhang, Y., & Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification.

[2] Yoon Kim. (2014). Convolutional neural networks for sentence classification.

[3] Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative study of CNN and RNN for natural language processing.