KAGGLE PROJECT REPORT

PENGCHENG JIANG AND GANG LI

Abstract. I finished the project and got good results.

Contents

Date: 2021-07-17.
1991 *Mathematics Subject Classification.* Artificial Intelligence.
Key words and phrases. kaggle,slides,notebook.

1. Problem Definition

Last year, in the Toxic Comment Classification Challenge, you built multi-headed models to recognize toxicity and several subtypes of toxicity. This year's competition is a related challenge: building toxicity models that operate fairly across a diverse range of conversations.

Here's the background: When the Conversation AI team first built toxicity models, they found that the models incorrectly learned to associate the names of frequently attacked identities with toxicity. Models predicted a high likelihood of toxicity for comments containing those identities (e.g. ”gay”), even when those comments were not actually toxic (such as ”I am a gay woman”). This happens because training data was pulled from available sources where unfortunately, certain identities are overwhelmingly referred to in offensive ways. Training a model from data with these imbalances risks simply mirroring those biases back to users.

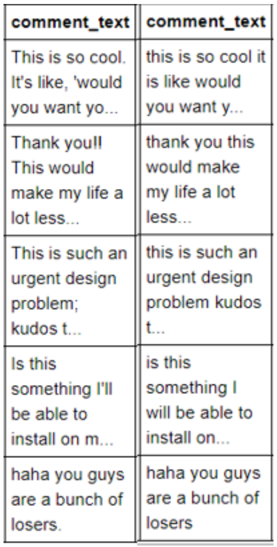
In this competition, you're challenged to build a model that recognizes toxicity and minimizes this type of unintended bias with respect to mentions of identities. You'll be using a dataset labeled for identity mentions and optimizing a metric designed to measure unintended bias. Develop strategies to reduce unintended bias in machine learning models, and you'll help the Conversation AI team, and the entire industry, build models that work well for a wide range of conversations.

2. Text preprocessing

- Count the total number of words contained in all texts, the maximum and minimum number of words contained in a text
- Check for missing data
- Change abbreviations to full:isn't -¿ is not(via dictionnary)
- clean˙numbers
- Find all non alphabetic characters and clean˙special˙chars
- Solve the problem of misspelling words
- lower

3. Embedding

- What tokenizer does is actually very simple. It divides the words it sees into spaces, and then uses numbers to correspond one by one. Then we take the first num˙ Words is the word with the highest frequency, others are not recognized.
- First learn the dictionary of the text, and then get the corresponding relationship between words and numbers, and then convert the text into a number string through this relationship, and then use the padding method to make up the number string to the same degree, then you can proceed to the next step : embedding
- collections.counter,pytorch:torchtext.vocab,
- The embedding layer is the same as word2vec. Whether it is skip gram or cbow model, they infer each other from the context and the current, so we consider the relationship between the preceding and the following.
- glove.42B.300d.txt



comment_text	comment_text
This is so cool. It's like, 'would you want yo...	this is so cool it is like would you want y...
Thank you!! This would make my life a lot less...	thank you this would make my life a lot less...
This is such an urgent design problem; kudos t...	this is such an urgent design problem kudos t...
Is this something I'll be able to install on m...	is this something I will be able to install on...
haha you guys are a bunch of losers.	haha you guys are a bunch of losers

FIGURE 1. data cleaning

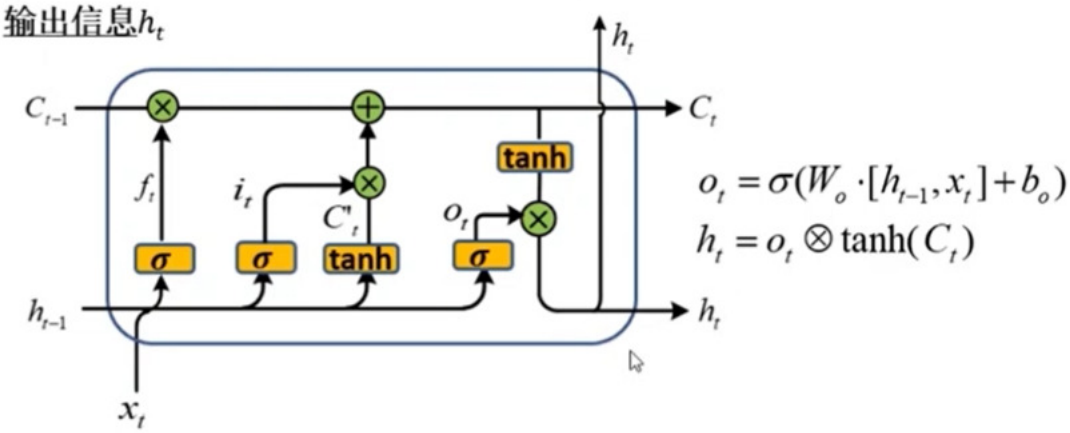


FIGURE 2. LSTM

4. MODULE

- Long short term memory (LSTM) is a special RNN, which is mainly used to solve the problem of gradient disappearance and gradient explosion in the process of long sequence training. In short, LSTM can perform better in longer sequences than ordinary RNN
- BiRNN:In practical problems, there are also problems that not only rely on the previous sequence, but also rely on the subsequent sequence for prediction. For those problems, we need to use bidirectional RNN (birnn)
- embed˙size, num˙hiddens, num˙layers = 300, 100, 2

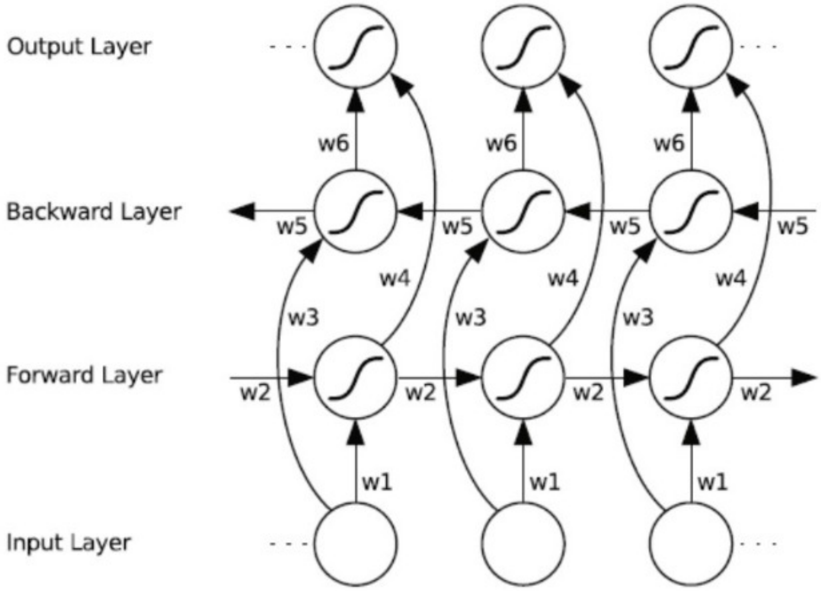


Figure 3. BIRNN

5. Conlusion

Through text processing and glove embedding, the model achieves good results

(A. 1) School of Computer Science,, JiLin University, jilin 130000, China
Email address, A. 1: `pcjiang@tulip.academy`

(A. 2) School of Information Technology, Deakin University, Geelong, VIC 3216, Australia
Email address, A. 2: `gang.li@deakin.edu.au`