Natural Language Processing 365

Data Science   Natural Language Processing   NLP Papers Summary

Day 236: NLP Papers Summary – A BERT Based Sentiment Analysis And Key Entity Detection Approach For Online Financial Texts

By Ryan   23rd August 2020   No Comments

Objectives and Contributions

Used BERT to perform sentiment analysis and key entity detection on financial texts. We consider key entity detection as a sentence matching task. We proposed an ensemble learning that outperformed SVM, LR, NBM, and BERT for two financial sentiment analysis and key entity detection datasets.

We combine sentiment analysis and key entity detection in a unified approach based on RoBERTa. We fine-tuned RoBERTa using different methods to implement sentiment analysis and key entity detection. With sentiment analysis, we focus specifically on the negative emotion information. With key entity detection, we were able to match each financial entity with its text to detect the key entities of negative information.

Methodology

Existing entity detection methods cannot detect key entities, let alone selecting the most relevant entities with tags. Therefore, we propose RoBERTa based sentiment analysis and key entity detection approach for online financial texts as follows:

1. Analyse the sentiment of text using RoBERTa (classic)
2. Get financial entity list and select key entities
3. Select the key entity with tags

In terms of step 2, for each piece of financial data, we use NER to get a list of entities. In the coarse-grained task, we detected some key entities related to the financial text from the entity list. As shown in the figure below, we feed each entity and the financial text into RoBERTa and use RoBERTa as a sentence matching model to determine whether each entity is a key entity. The number of models is determined by the number of entities. We would form a list of key entities of those entities that are predicted as key by RoBERTa.
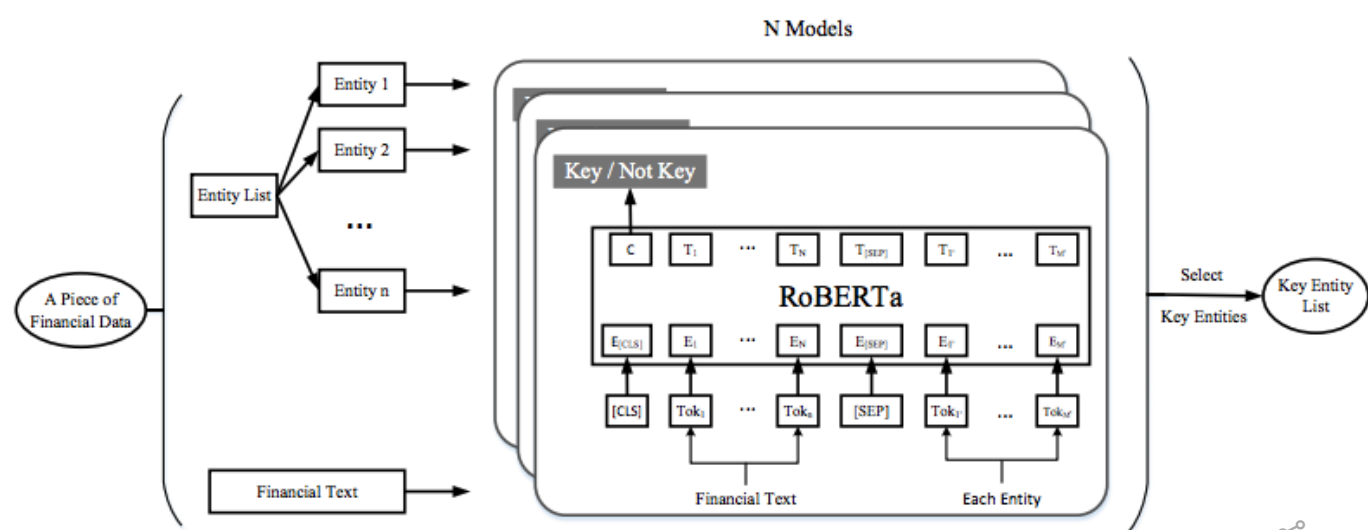

Fig. 2. Get financial entity list and select key entities.

In terms of step 3, in certain fine-grained tasks, we are required to select the most relevant key entities based on the tags of the financial text. Tags can be corruption, fraud etc. This is more of a machine reading comprehension (MRC) task and therefore, we treat each financial text as an MRC article and rewrite its tag as MRC questions. The answers predicted by the MRC model are considered to be key entities.

Experiments and Results

DATASET

We used two datasets: Negative Financial Information and Subject Deter-mination 2019 and Event Subject Extraction for Financial Field. The table below showcase the descriptive statistics of the dataset. Our evaluation metrics are accuracy rate for sentiment analysis and F1 score for key entity detection.

TABLE I.   DIFFERENCE BETWEEN DATASET 1 AND DATASET 2

Data set	Sentiment	Entity List	Tag
Dataset 1	Negative and Positive	Exist	Not Exist
Dataset 2	Negative	Not Exist	Exist

RESULTS

TABLE II.   RESULT OF SENTIMENT ANALYSIS

Model	Accuracy
$BERT_{BASE}$ to Vector + NBM	0.87750
$BERT_{BASE}$ to Vector + LR	0.91957
$BERT_{BASE}$ to Vector + SVM	0.94034
$RoBERTa_{BASE}$ to Vector + SVM	0.94762

Model	Accuracy
$BERT_{BASE}$-Single	0.95427
$RoBERTa_{BASE}$-Single	0.96638
Our Approach	**0.96774**

TABLE III.	RESULT OF KEY ENTITY DETECTION FROM DATASET 1
Model	F1 Score
BERT$_{BASE}$-Single	0.94537
RoBERTa$_{BASE}$-Single	0.94953
Our Approach	0.95082

Source: https://arxiv.org/pdf/2001.05326.pdf

Ryan

Data Scientist

Next Post

Previous Post

Day 235: Learn NLP With Me - Topic Modelling with LSA and LDA

Day 237: Learn NLP With Me - An Exhaustive Guide to Detecting and Fighting Neural Fake News using NLP

Data Science   **Natural Language Processing**   **NLP Papers Summary**

Day 365: NLP Papers Summary – A Survey on Knowledge Graph Embedding

Ryan
30th December 2020

Data Science   **Implementation**   **Natural Language Processing**

Day 364: Ryan's PhD Journey – OpenKE-PyTorch Library Analysis + code snippets for 11 KE models

Ryan
29th December 2020

Data Science   Ryan's PhD Journey

Day 363: Ryan's PhD Journey – Literature Review – Knowledge Acquisition – 1st Passes XIV

Ryan
28th December 2020