

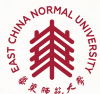
# Financial text emotion analysis model

# 金融文本情感分析模型

组长：唐小卉

组员：

- 王溢阳
- 杨茜雅
- 仲韦萱（汇报人）



ECNU

Copyright © 2021 ECNU Corporation. All rights reserved.

Tel: +86-021-62233586 Fax: +86-021-62606775

E-mail: [ecnu@ecnu.com.cn](mailto:ecnu@ecnu.com.cn) Http: <http://www.ecnu.edu.cn>

# 目录 Catalogue

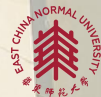
研究背景  
Background

问题定义  
Problem definition

模型应用  
Model application

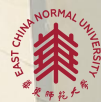
结果展示  
Result presentation

总结  
Summarize



ECNU

# 1.研究背景 Background



ECNU

## 什么是情感分析？

识别和提取文本材料中的主观信息。它主要用于了解人们在某个主题、产品或服务上的情绪倾向，通常分类为正面、负面或中性。

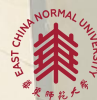
- 社交媒体：了解公众对品牌、产品、服务的看法
- 金融市场：预测股市动态或经济趋势

## 为什么要研究金融文本情感分析？

- 金融文本包含投资者的情绪以及公众对相关事件的态度
- 提取文本中蕴含的语义情感信息，帮助理解投资者态度，也会影响投资决策和市场走势
- 每天都会有大量的词汇和信息出现，手动分析这些语言非常困难，需要自动化的方法来处理

## 2. 问题定义

## Problem definition



ECNU

## 选择哪种情感分析任务？

- 二分类情感分析
- 多分类情感分析
- 方面级情感分析
- 语言模型任务

## 选择哪种应用模型方法？

- 基于预训练模型微调
- 特征提取

## 与其他模型相比，选择什么样的模型在短句分类的方面表现更好？

- 短句分类任务在资源消耗上与长文本处理有所不同

## 训练过程中参数对于分类性能的影响如何？

## 能否减少或避免“灾难性遗忘”

- 模型需要不断适应新的数据或情境，同时保留对历史数据的有效识别能力



## 金融文本情感分析的难点

- 使用专用词汇和语言以及缺乏标记的数据，导致通用模型不够有效。
- 现有模型易受到对抗样本的干扰导致模型结果出错
- 利用神经网络，需要大量的标记数据和标记资金文本片段，需要大量专业知识

## 使用预训练语言模型的理由

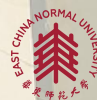
- 需要更少的标记示例
- 再特定的语料库上进一步训练

## 模型的核心思想

- 在更大的语料库上训练模型
- 在特定领域未标记的语料库上进一步提高模型能力

# 3.模型应用

## Model application



ECNU



# 数据集

## Sentiment Analysis for Financial News

简介:

数据集包含了从零售投资者角度来查看的财经新闻标题的情绪。

\*更多相关信息: Malo、Sinha、Takala、Korhonen 和 Wallenius (2014年): “债务好还是坏: 在经济文本中检测语义取向。”《美国信息科学技术协会期刊》。

### Data Structure of Dataset:

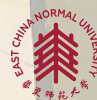
Number of Rows: 4846

Number of Columns: 2

Column Details:

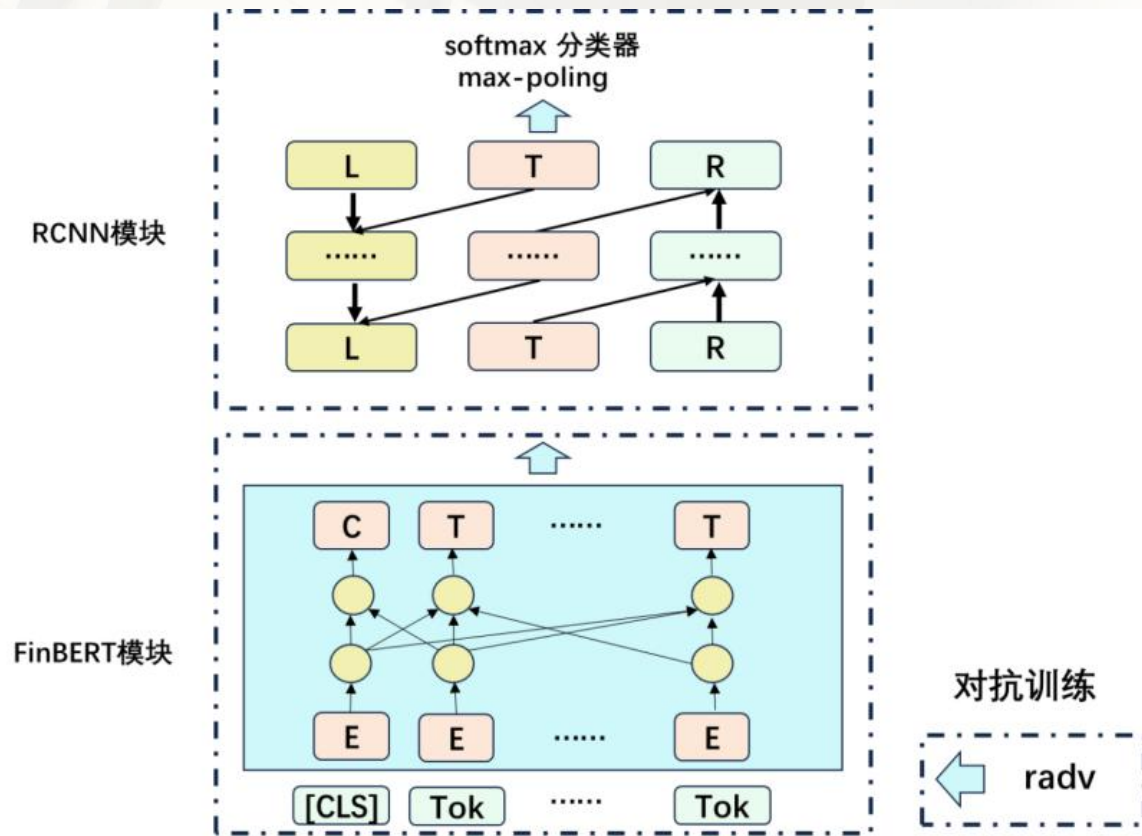
label: object, Non-Null Count: 4846

content: object, Non-Null Count: 4846

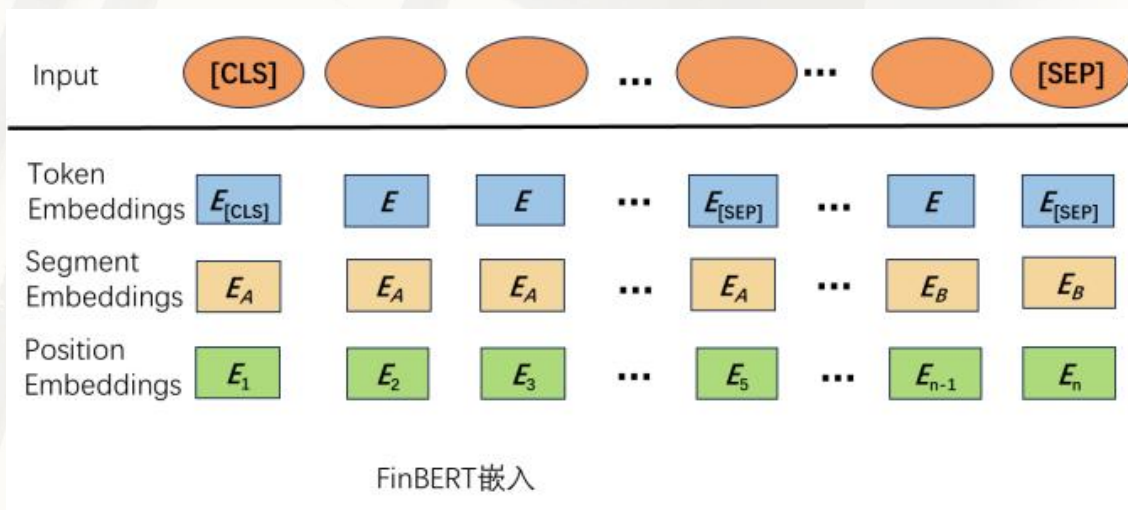


ECNU

## 模型概念图



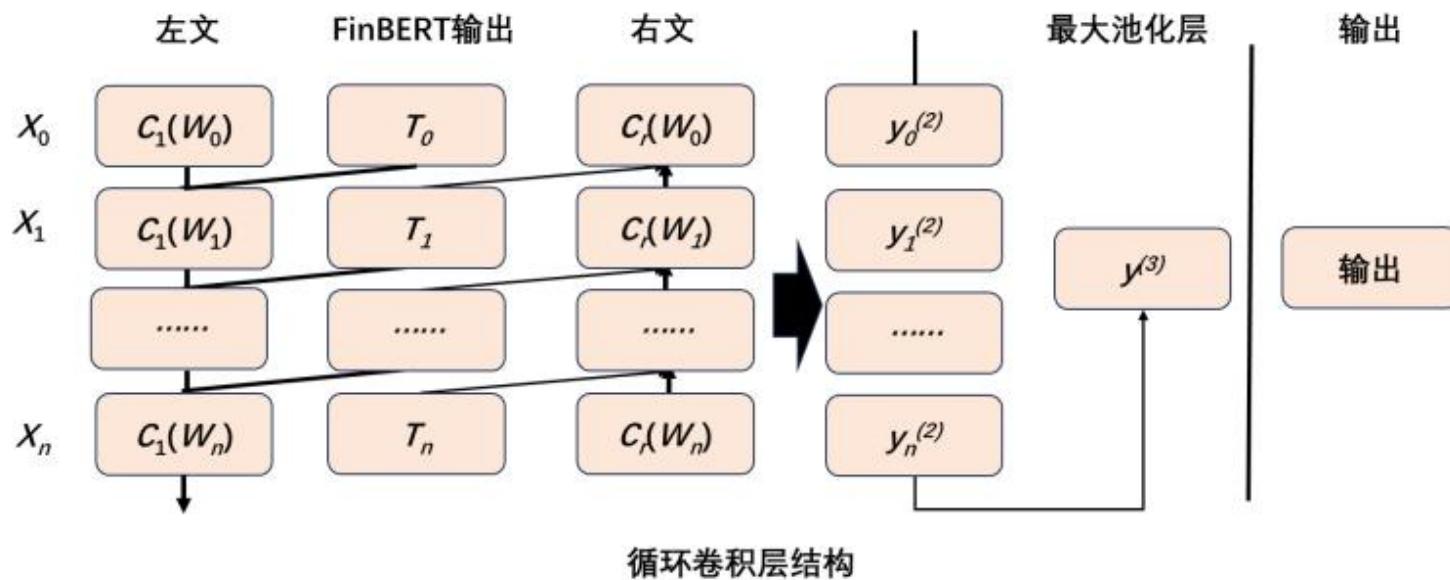
## FinBERT模块



$$[E_{\text{word}} = E_{\text{tok}} + E_{\text{seg}} + E_{\text{pos}}]$$

$$[T_1, T_2, \dots, T_n]$$

## RCNN模块



## 对抗训练

$$[r_{\text{adv}} = \arg \min \log p(y|x + r; \theta)][\text{subject to } \|r\| \leq \epsilon]$$

- 在给定输入  $x$  和模型参数  $\theta$  的情况下，预测标签  $y$  的概率
- $r$  是加在输入上的扰动，且此扰动在足够小的范围内被严格限制。目标是寻找能够最大化模型预测误差的最优扰动
- 对抗性扰动被添加到输入嵌入中，以在嵌入空间中构造新的对抗性文本实例



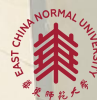
## 模型参数设置

参数	含义	数值
hidden_size	BERT模型和LSTM的隐藏层大小	768
num_labels	分类任务的标签数量	3
dropout	Dropout层的丢弃率	0.1
epsilon	对抗训练中扰动的大小	1e-5
lr	AdamW优化器的学习率	5e-5
num_epochs	训练的总轮数	3
criterion	交叉熵损失函数	



# 4.结果展示

## Result presentation



ECNU

## 对比实验

### BiLSTM

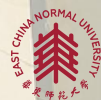
- BiLSTM层配置为64个单元，通过其双向结构，模型能有效学习文本数据的前向和后向依赖关系

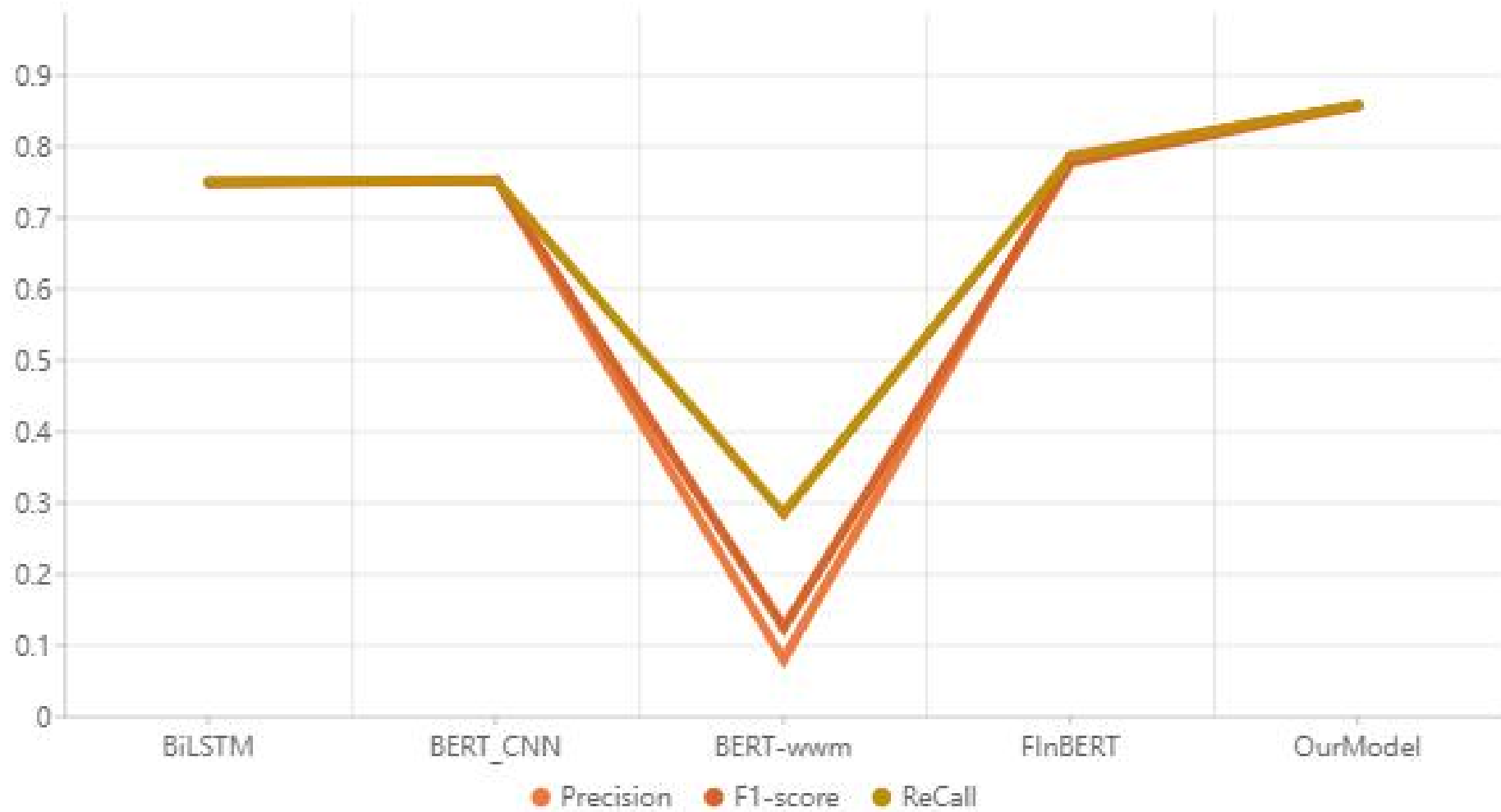
### BERT\_CNN

- 在 BERT 模型下游引入CNN 提取词级特征,，使用预训练的BERT模型作为基础来提取文本的深层语义特征，并在此基础上通过一维卷积层 Conv1D 进一步提取关键的局部特征

### BERT\_wwm

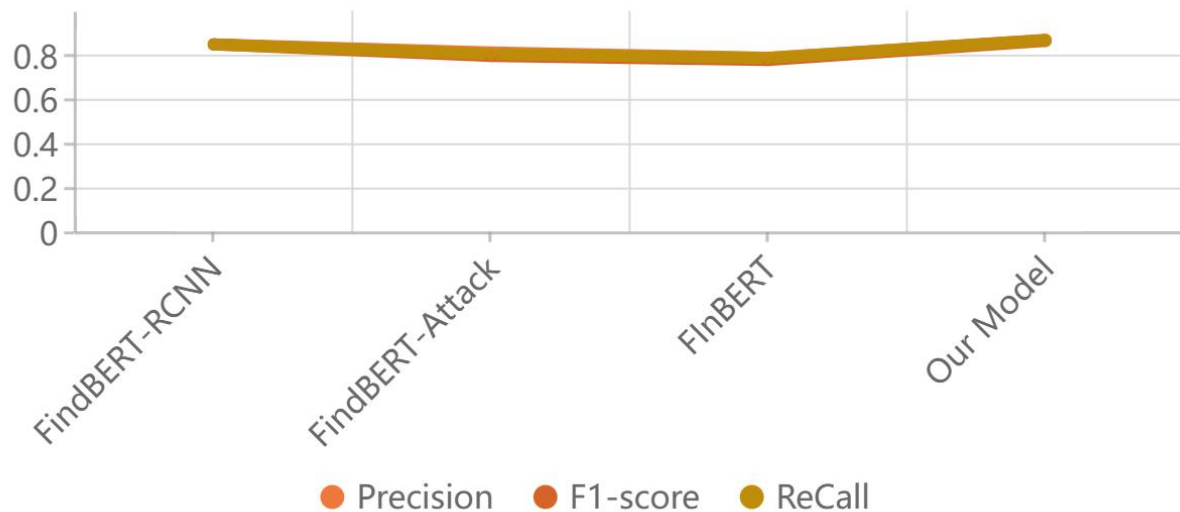
- 整合了全词覆盖 Whole Word Masking 策略和命名实体及词性标注的特征增强





## 消融实验

- FinBERT: 金融领域的预训练模型
- FinBERT\_RCNN: 预训练模型下游引入RCNN模型提取关键信息
- FinBERT\_对抗训练: 预训练模型引入对抗训练
- Our\_Model: 我们使用的模型



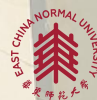
## 持续优化

精确度从原有模型的86.65%提高到92.17%，F1分数从86.12%提高到91.83%，召回率也从86.51%提升至92.04%。

Model	Precision	F1-score	Recall
NEW_Model	92.17%	91.83%	92.04%

# 5.总结

## Summarize



ECNU



## 项目内容

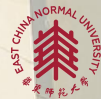
- 在金融文本情感分析领域改进FinBERT模型
- FinBERT模型的优势在于其专门针对金融领域的语言特性进行了优化
- 探讨不同的训练周期和模型配置对最终结果的影响，通过多次实验，结合RCNN架构和对抗训练技术、

## 改进方向

- 语境敏感性和语言歧义
- 适应金融语言快速变化的能力
- 对标记数据的依赖
- 跨不同金融语境的泛化能力

## 未来展望

- 增强特征工程
- 跨语言和跨域适应性



# Q&A



**ECNU**

Copyright © 2021 ECNU Corporation. All rights reserved.  
Tel: +86-021-62233586 Fax: +86-021-62606775  
E-mail: [ecnu@ecnu.com.cn](mailto:ecnu@ecnu.com.cn) Http: [//www.ecnu.edu.cn](http://www.ecnu.edu.cn)