



# Washington University in St. Louis

---

## OLIN BUSINESS SCHOOL

用于预测业务成果的深度学习

最终项目

**Lending Club** 贷款违约检测项目第 28 组第

22 条 小组成员:

孙繁迪、陈宇健、周宇凡 2024 年 5 月

7 日

## 摘要

在数字市场中，在线评论的完整性因虚假评论的盛行而受到损害。该项目提出了一种结合 LSTM 和 CNN 的混合深度学习模型，以增强对虚假评论的检测。该模型利用 Kaggle 包含超过 40,000 条记录的数据集，旨在通过准确识别欺诈内容来超越传统的检测限制，目标准确率超过 90%。这种方法有望通过提高在线评论的真实性来增强消费者的信任并确保公平的商业行为。

## 1 介绍

在在线商务领域，客户评论在影响购买决策和建立企业声誉方面发挥着至关重要的作用。因此，这些评论的真实性对消费者和企业都至关重要。然而，在数字领域，虚假评论（操纵反馈）的盛行率有所上升，这扭曲了消费者的看法并提供了不公平的竞争优势。

传统的虚假评论检测方法，例如简单的算法分析和手动验证，由于准确性低且无法应对在线评论数据的指数级增长，已被证明是不够的。这些方法通常难以区分细微差别的真实内容和虚假内容，从而导致大量误报和漏报。虚假评论中使用的语言非常复杂，通常模仿真实用户的情绪，这对现有模型提出了重大挑战。

该项目的目标是开发一个能够更可靠地检测虚假评论的系统，从而恢复对在线评论系统的信任并保护电子商务交互的完整性。通过实现超过 90% 的准确率，该项目旨在为审核验证技术设立新标准。

## 2 文献综述

## 3 问题描述

目前识别虚假评论的方法面临多重挑战：

不准确：许多现有算法难以区分由复杂方法生成的复杂类人文本和真实评论，导致误报和漏报率很高。

可扩展性：随着在线内容容量的增长，传统的检测方法通常涉及手动验证或简单的自动化技术，无法有效处理每天生成的大量数据。

适应性：虚假评论策略不断演变，使用更先进的技术

逃避检测的技术和微妙的语言风格。现有系统通常无法快速适应这些变化，因此随着时间的推移，效率会降低。

此外，用于检测这些评论的数据集虽然涵盖多个产品类别并包含超过 40,000 个条目，但也提出了其自身的挑战。评论质量的多样性、欺诈策略的微妙性以及类别之间语言表达的差异都增加了开发有效检测模型的复杂性。

该项目旨在通过开发一个强大、可扩展且适应性强的系统来解决这些问题，该系统能够准确识别欺诈性评论。该系统采用混合模型，将 LSTM 的顺序数据处理能力与 CNN 的模式识别优势相结合，旨在显著改善虚假评论的检测并恢复在线评论平台的完整性。

## 4 数据库背景及数据预处理

数据集来自Kaggle，有4列，四万多行，真假评论比例约为50比50，包含2万条虚假评论和2万条真实产品评论。对于四个栏目，我们有类别、评级、标签和评论栏目，我们将主要关注标签和评论栏目。在标签列中，有两个唯一值：**OR** 和 **CG**。**OR** = 原始评论（大概是人类创建的且真实的）；**CG** = 计算机生成的虚假评论。计算机生成的评论由 GPT 2 生成，原始评论来自亚马逊。

## 方法

### 循环神经网络

循环神经网络 (RNN) 是一种深度学习模型，特别适合序列数据分析。与单向前馈神经网络相反，它是一种双向人工神经网络，这意味着它允许某些节点的输出影响相同节点的后续输入。这使得它们对于先前输入的上下文显著影响当前预测的任务非常有效。

- 嵌入层：

将评论中的每个单词转换为固定长度的单词向量。词向量可以捕获单词的语义信息并将其映射到连续的向量空间。通过指定 `max_features` 参数，模型可以处理预定义数量的项。

- RNN 层 (SimpleRNN)：

该模型使用三个堆叠的 RNN 层，每个层包含多个循环单元。循环单元能够记住先前输入的信息并用它来处理当前的信息

输入，捕获文本序列中的上下文信息和潜在模式。RNN 的最后一层将输出传递到密集层进行最终分类。

- 致密层：  
最后一层使用 sigmoid 激活函数，它输出一个介于 0 和 1 之间的概率值，表示该评论是由计算机生成的概率。
- 执行部分：  
sigmoid 激活函数用于二元分类，以确定评论是否是机器生成的。优化器是“nadam”，它结合了 Adam 和 NAG 的优点：Adam 的自适应学习率与 NAG 的动量校正相结合，可实现更快的收敛和更好的泛化性能。评估方法设置为“准确度”，即真阳性和真阴性占整个结果的百分比。
- 型号限制：  
在这一部分中，我们面临着过拟合问题，训练数据集和测试数据集的损失和准确度值之间存在巨大差距。我们将添加惩罚部分来处理这个问题。同时，我们还需要继续尝试更好的模型来不断提高准确率。

## 美国有线电视新闻网

步骤与 RNN 模型非常相似，只是 RNN 层改为一维 CNN 和池化层。

嵌入和卷积神经网络 (CNN) 的结合是处理顺序数据的强大技术，特别是对于自然语言处理任务。词嵌入通过将词表示为密集向量来捕获词的语义信息，而 CNN 可以通过卷积滤波器检测序列数据中的模式。

## LSTM 与 CNN

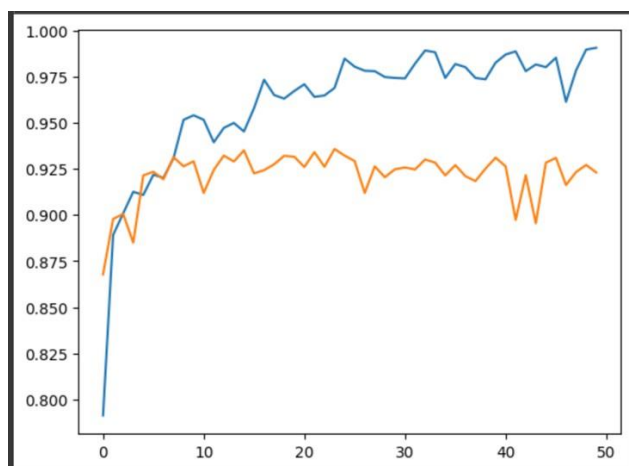
长短期记忆 (LSTM) 是一种循环神经网络 (RNN) 架构，旨在解决传统 RNN 中存在的梯度消失问题。与其他 RNN 相比，它对间隙长度相对不敏感是其优势。

## 带有 CNN 的双向 LSTM

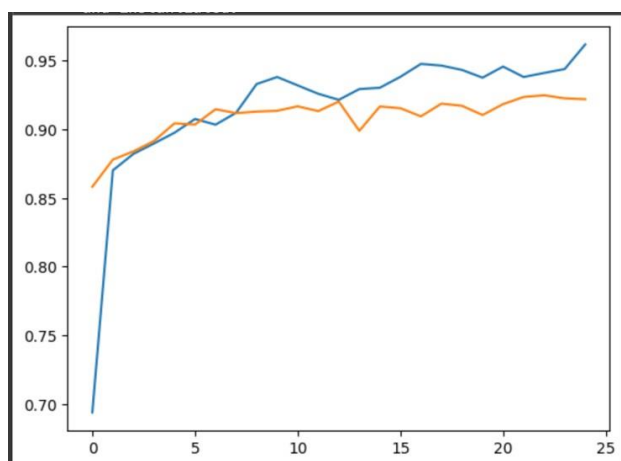
Bi-LSTM（双向长短期记忆）是一种循环神经网络 (RNN)，可以向前和向后处理顺序数据。将双向长短期记忆 (LSTM) 网络与卷积神经网络 (CNN) 相结合可以增强模型学习空间和时间依赖性的性能。这种组合对于理解顺序关系并同时识别数据中的模式很有用。

## 结果

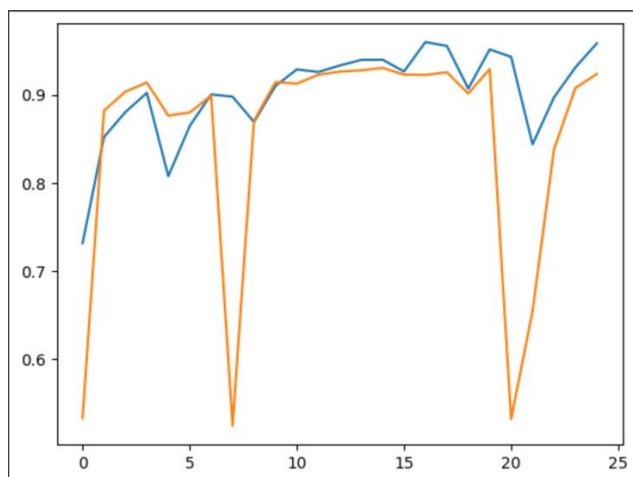
### 表现



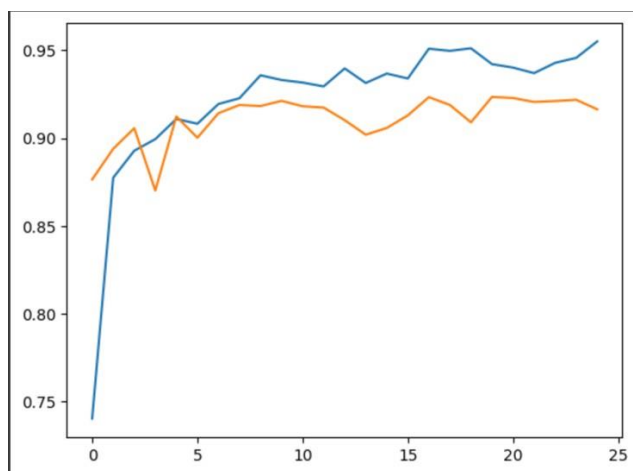
在RNN模型中，训练数据集的准确率已经提升到97.5%左右，测试数据集的准确率稳定在92.5%左右，无法再提升。训练数据集和测试数据集之间存在差距，因此我们面临过拟合问题，我们需要添加 L2 或/和从每一层中删除一些神经元。



由于我们在模型中添加了L2，过拟合问题得到了进一步解决，差距仍然存在，但距离变小了。由于我们只是将RNN改为CNN，所以我们在模型选择上并没有太大的改进，整体准确率没有提高。训练集准确率仍在95%左右，测试集准确率在90%左右。



我们在模型中将 LSTM 与 CNN 结合起来。同时，我们在该部分还使用了dropout和recurrent\_dropout函数，我们设置为20%。因此，过拟合问题也不明显。测试和训练数据集都在 90% 左右或以上。



我们在模型中将两个双向 LSTM 与一个 CNN 结合起来。我们还在模型中添加了dropout和L2函数，因此训练和测试数据集的准确率非常接近。测试和训练数据集都超过 90%，接近 92% 和 95%。

## 结论

总的来说，我们实施了 4 种方法来改进我们的模型，并尝试解决我们面临的潜在问题，特别是欠拟合和过拟合问题。最终，我们将测试数据集的准确率稳定在90%以上。我们认为它处于一个相对较高的水平。通过模型的准确率超过 90%，该项目提供了一个强大的解决方案，有可能对在线市场产生重大影响。恢复对在线评论平台的信任，使消费者能够做出明智的决定。最终，为更加透明、更加公平的在线生态系统做出贡献。

附录

日期说明:

变量	描述
类别	掉落
评分	掉落
标签	CG/OR（OR = 原始评论（大概是人类创建的且真实的）；CG = 计算机生成的虚假评论）
文本	用户对产品的评论（字符串）

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 50, 64)	640000
simple_rnn (SimpleRNN)	(None, 50, 32)	3104
simple_rnn_1 (SimpleRNN)	(None, 50, 16)	784
simple_rnn_2 (SimpleRNN)	(None, 32)	1568
dense (Dense)	(None, 1)	33

=====  
Total params: 645,489  
Trainable params: 645,489  
Non-trainable params: 0

RNN 对每一层进行建模并输出形状细节

Layer (type)	Output Shape	Param #
embedding_3 (Embedding)	(None, 50, 64)	640000
conv1d_2 (Conv1D)	(None, 46, 64)	20544
max_pooling1d_1 (MaxPooling1D)	(None, 15, 64)	0
conv1d_3 (Conv1D)	(None, 11, 32)	10272
global_max_pooling1d_1 (GlobalMaxPooling1D)	(None, 32)	0
dense_1 (Dense)	(None, 1)	33

=====  
Total params: 670,849  
Trainable params: 670,849  
Non-trainable params: 0

CNN 对每一层进行建模并输出形状细节

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 50, 64)	640000
conv1d (Conv1D)	(None, 46, 64)	20544
lstm (LSTM)	(None, 32)	12416
dense (Dense)	(None, 1)	33
Total params: 672,993		
Trainable params: 672,993		
Non-trainable params: 0		

具有 LSTM 的 CNN 每层和形状细节

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 50, 64)	640000
conv1d_1 (Conv1D)	(None, 46, 64)	20544
bidirectional_2 (Bidirectional)	(None, 46, 64)	24832
bidirectional_3 (Bidirectional)	(None, 64)	24832
dense_1 (Dense)	(None, 1)	65
Total params: 710,273		
Trainable params: 710,273		
Non-trainable params: 0		

每层有两个双向的 CNN 和形状细节

## 参考

Joni Salminen、Chandrashekhar Kandpal、Ahmed Mohamed Kamel、Soon-gyo Jung、Bernard

J. Jansen, 创建和检测在线产品的虚假评论, 《零售和消费者服务杂志》, 第 64 卷, 2022,102771, ISSN 0969-6989, <https://doi.org/10.1016/j.jretconser.2021.102771> (<https://www.sciencedirect.com/science/article/pii/S0969698921003374>)

N. A. Patel 和 R. Patel, “利用机器学习技术检测虚假评论的调查”, 2018 年第四届国际计算通信与自动化会议 (ICCCA), 印度大诺伊达, 2018 年, 第 1-6 页, doi: 10.1109/ICCA.2018.8777594. 关键词: {特征提取; 支持向量机; 语言学; 机器学习; 分类算法; 决策树; 虚假评论; 情感分析; 意见垃圾邮件; 虚假评论检测技术; 机器学习},