

基于CNN与Bi LSTM融合的用户行为预测方法

于璐娜, 刘国奇*, 崔文超

沈阳建筑大学计算机科学与工程学院, 辽宁 沈阳

收稿日期: 2024年7月1日; 录用日期: 2024年7月28日; 发布日期: 2024年8月6日

摘要

随着个性化服务需求的增长, 通过分析人类行为模式来预测人们的生活习惯、需求和偏好变得尤为重要。本文探讨了基于生活日志数据进行用户行为分类预测的重要性, 提出一种融合卷积神经网络(CNN)和双向长短期记忆网络(Bi LSTM)的用户行为预测模型TS-CNN-Bi LSTM。使用Liu-Life Log收集十二年的用户生活日志数据作为样本数据。考虑生活日志数据存在不公开位置信息的用户, 基于同一数据集构建了两种不同特征组合的数据集DCP-Life Log和TS-Life Log。使用不同数据集评估模型预测性能。实验结果表明, 所提模型在两个数据集上的宏平均精确率、召回率和F1值均优于基准模型, 分别提升2.23%、1.30%和1.27%。

关键词

生活日志, 神经网络, 行为预测, 组合模型

User Behavior Prediction Method Based on Fusion of CNN and Bi LSTM

Luna Yu, Guoqi Liu*, Wenchao Cui

School of Computer Science and Engineering, Shenyang Jianzhu University, Shenyang Liaoning

Received: Jul. 1st, 2024; accepted: Jul. 28th, 2024; published: Aug. 6th, 2024

Abstract

With the growing demand for personalized services, it has become essential to predict people's habits, needs, and preferences by analyzing human behavior patterns. In this paper, we explore the importance of classifying and predicting user behavior based on life log data and propose a user behavior prediction model TS-CNN-Bi LSTM that integrates a convolutional neural network (CNN) and bidirectional long short-term memory network (Bi LSTM). Twelve years of user life log

*通讯作者。

data collected by Liu-Life Log is used as sample data. Considering the existence of users who do not disclose their location information in the life log data, two datasets with different feature combinations, DCP-Life Log and TS-Life Log, are constructed based on the same dataset. The model prediction performance is evaluated using different datasets. The experimental results show that the proposed model outperforms the benchmark model on both datasets regarding the macro-mean precision, recall, and F1 value, with an improvement of 2.23%, 1.30%, and 1.27%, respectively.

Keywords

Life Logs, Neural Networks, Behavioral Prediction, Combinatorial Models

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着数字技术的进步，生活记录变得更加可行。生活日志的可能来源包括可穿戴相机、可穿戴传感器、智能手机和电脑。例如，生活日志可以是与他人交流的信息、GPS 捕捉到的位置线索、生物传感器捕捉到的生理状态、数字相册、带有照片的博客文章和运动相机捕捉到的视频片段[1] [2]。这些个性化多媒体数据呈现了个人日常体验的各个方面，为各种个人生活记录应用提供了丰富的存储库信息，包括生活方式理解、饮食监测、人类记忆回忆辅助和视觉生活日志检索[3]。因此，生活日志作为通过多媒体传感器被动采集和处理的反映生活经验的数据，受到了学术界和工业界的关注。生活日志为人类行为活动预测提供了一种新的可能性，因为已经证明日常活动反映了人们的主观行为[4]。

通过生活日志对人类行为预测的意义是深化对个体生活方式和偏好的理解，从而为个人提供更加个性化、精准的服务和建议。目前的生活日志分析主要侧重于特定的短期目标，比如场景搜索、饮食监测、项目推荐等，这些功能对于改善生活质量有一定的帮助。然而，如果能够将生活日志分析延伸至对个人行为的长期趋势和心理特征的研究，其意义将更为深远。通过生活日志的持续记录和分析，发现个体的行为模式和习惯，揭示出隐藏在日常活动背后的心理动机和情感需求[5]。例如，一个人的睡眠时间、饮食偏好、运动习惯等数据，可以反映出他们的生活节奏、健康状况以及情绪波动。基于这些数据，我们可以预测个体行为趋势，比如可能的购物偏好、社交活动倾向、甚至是情绪状态的变化。在实际应用中，通过对生活日志的预测分析，可以为个人定制个性化的健康管理方案、时间安排建议、甚至是情绪调节策略[6]。比如，针对一个习惯晚睡的人，可以推荐调整作息时间来改善睡眠质量；对于一个饮食习惯不健康的人，可以提供营养均衡的饮食方案。此外，通过对个体情绪波动的预测，还可以提供情感支持和心理疏导，帮助个人更好地应对生活中的挑战和压力[7]。

虽然通过对生活日志进行行为预测具有很大的潜力，但也面临诸多挑战[8]-[11]。数据多样性与复杂性：生活日志记录了个体多个方面的信息，包括文字、图片、音频、视频等多种模态，这些数据涵盖了丰富的生活细节和情境信息。然而，这种多样性和复杂性使得数据的处理和分析变得更加困难，需要综合利用多种技术手段进行有效的信息提取和特征表征。数据量和质量：生活日志数据量大、且质量参差不齐，其中可能包含了大量的噪音和无关信息，这会影响到对用户行为的准确分类和预测。如何从海量数据中筛选出有效的特征，并处理好数据中的噪声和异常值，是一个具有挑战性的问题。

与传统的基于生活日志数据集的文本分类方法不同，本文基于 CNN 与 Bi LSTM 融合的用户行为预

测方法。首先, 考虑到生活日志中的时间节点信息对于用户行为分类具有重要的影响。将时间因素考虑到特征选择中, 使得模型更好地捕捉用户行为与时间的关系。其次, 为克服不公开位置信息的用户可能带来的影响。构建了两种特征组合的数据集(包含位置信息和不包含位置信息), 并用这两个数据集来评估模型性能。通过结合 CNN 和 Bi LSTM 模型, 充分利用语义感知处理得到的特征, 实现对用户行为活动的准确分类预测。

2. 相关工作

生活日志作为通过多媒体传感器被动采集和处理的反映生活经验的数据, 受到了学术界和工业界的关注。生活日志是一种现象, 人们可以用数字方式记录自己日常生活的不同细节, 用于各种目的[12]。Ziaei 等[13]利用大量的音频、语音和词汇特征来描述日常音频流中的事件。采用语音活动检测和说话人拨号系统对音频文件进行高层次的语义分割, 并提出一种分析和分类个人录音中日常活动的新方法。Tong 等[14]使用收集的包含生活日志和音乐信息的数据集。将生活日志信息与音频和音乐元数据结合起来, 提出一种基于二维 Thayer 情绪检测的模型。实验验证基于用户信息的分类方法可以有效地识别音乐情绪。Lee 等[15]从生活日志中提取特定的特征, 将这些特征用于基于人工神经网络的分类。本文的目的是探讨人类行为的潜在特征, 以促进快速诊断和治疗轻度认知障碍的早期诊断。Li 等[16]利用生活日志进行自我意识的新方向, 揭示了日常活动记录与身体和心理自我意识之间的重要关系。通过对个体的生理和心理状态进行分析、预测和干预, 更准确地了解个体状况, 并获得有效的健康建议, 甚至在早期发现身体和精神的异常。

人类行为的预测一直是一个特别活跃的研究课题。因为它在安全检测、健康分析、娱乐、个人大数据和人机交互等领域越来越重要[17]。早期的研究主要采用决策树、支持向量机(SVM)、贝叶斯等传统机器学习方法进行分类。Jain 等[18]提出了一种基于描述符的活动分类方法。采集加速度计和陀螺仪传感器信号来识别用户执行的活动。此外, 使用收集到的信号导出时域和频域信号。Gupta 等[19]提出使用朴素贝叶斯和 k 最近邻(k-NN)进行活动分类, 使用 Relief-F 和顺序前向浮动搜索(SFFS)从一系列先前发布的特征以及本文中引入的新特征中执行特征选择。建立一种利用穿戴式无线加速度计的精确活动识别系统, 用于患者监测的实际应用。Jalloul 等[20]使用随机森林(RF)分类器对活动进行分类。采用 6 个惯性测量单元组成监测系统。选择一些满足统计检验的网络测度组成特征集, 有效提高活动分类总体准确度。Ahmed 等[21]出了一种基于方差的活动分类方法。其中引入 Poincare Plot 来分析根据原始加速度计数据计算出的角度方差, 对躺着、坐着、站立和行走活动进行了分类。

近年来学术界的研究人员发现深度学习算法能够更高效地开展分类工作, 通过构建更深层次的网络结构, 可以更好地提高分类的准确度。Liu 等[22]提出了一种 DCP-Bi LSTM 行为分类模型, 将包含描述信息和位置信息的双向语义变量组合到 DCP-Bi LSTM 模型中识别用户的日常活动。Ordóñez 等[23]提出了一种结合深度 CNN 和 LSTM 活动识别分类器, 对 27 种手势和 5 种动作进行分类。该框架可以应用于同质传感器模式, 也可以融合多模式传感器以提高性能。Shi 等[24]提出一种可以考虑上下文信息的特定模型 C-LSTM, 模型在词嵌层后加入了 CNN 和 LSTM 网络, 充分利用了 CNN 和 LSTM 各自的优点。Lin 等[25]提出了一种随机森林迭代卷积神经网络, 对雷达签名的活动模式进行准确分类, 取代传统的微多普勒图像预处理。所提出的策略使用迭代深度学习框架来自动定义和提取特征, 以准确地对活动或主题进行分类。

3. 数据集描述

3.1. Liu-Life Log

本文利用 Liu-Life Log 数据集进行日志分类预测的方法与应用。Liu-Life Log 项目始于 2011 年, 由我们

的团队在过去 12 年中自主收集[26]。涵盖了 26 种不同的日常行为类别，如工作、休息、学习和体育活动等，体现了教学工作者在多样化环境中的行为模式。现在任何人都可以在我们的网站(<http://www.lifelog.vip/>)上免费下载数据集和移动应用程序，并获得我们公开的生活日志数据来参与我们的项目。表 1 给出了部分数据集。此研究不仅为 Life Logging 数据的分类提供了新的视角，也为相关领域的研究者和实践者提供了实用的工具和方法。

Table 1. Example of Liu-Life Log

表 1. Liu-Life Log 示例

| Cluster | Date | Describe | Address | Behavior |
|---------|------------------------------|--|---|----------------|
| 27 | 2011/12/18 Sunday 9:53:54 | On the way to the new campus | Suizhong County, Huludao City, Liaoning Province, China | On road |
| 7 | 2012/2/5 Sunday 19:28:06 | Wash the car outside. It's super fine today. | Binhe Road, Tiedong District, Siping City, Jilin Province, China | Rest |
| 1 | 2012/5/1 Tuesday 17:51:03 | The new president came in and appointed the conference to be held. | Dongling District, Shenyang City, Liaoning Province, China | Work in school |
| 62 | 2012/5/18 Friday 18:05:35 | Company Meetings. | Shenyang Railway Station | Work outside |
| 54 | 2012/5/6 Sunday 18:30:49 | The computer in the laboratory has been stripped of its shell. | 108 National Highway, Haidian District, Beijing, China | Work in school |

3.2. 数据预处理

为了给所提出的神经网络提供一定的数据维度，提高模型的精度，需要对生活日志系统采集的原始数据进行如下预处理。

3.2.1. 数据过滤和清理

在采集生活日志的过程中，由于系统故障或者用户提交原因，数据集中存在大量的空数据。此外，用户在使用它的时候不会坚持一个设定的时间来发布动态，而且它是不连续的。为了完成分类和预测任务，需要对来自 Liu-Life Log 的所有噪声数据进行清理。识别并处理数据集中的缺失空数据，采取删除策略删除特定格式和无意义的的数据，去除重复数据、错误数据，保证数据的一致性和准确性。

3.2.2. 构建特征集

考虑到用户隐私安全以及大多数用户并不愿意在生活日志中公布自己的位置信息。对 Liu-Life Log 数据集中提取分离出两种不同类型的数据集，分别是包含位置信息的 DCP-Life Log 和不包含位置信息的 TS-Life Log。具体描述如表 2 所示，这两种数据集来源一个数据集，由不同的特征组成。并且我们发现，用户的行为活动与特定时间有一定关系。例如，我们针对生活日志数据集中大量数据分析出，用户在 11:00~13:00 上传的生活日志普遍是吃饭的日常，对应的用户活动行为是“在餐厅”。因此，在模型的构建中加入时间节点，有利于网络模型捕获基于时间节点的信息特征，强化模型预测用户行为活动能力。为保证实验的合理性，将两个数据集同样的各取 9485 条作为训练集，验证集和测试集各 2000 条。

通过将图片和视频通过技术转成本数据，与原本的文本信息结合作为网络模型特征，可以为网络模型提供更多样化、更丰富的特征表达。利用 Youcai Zhang 等开发的预训练深度学习模型 RAM [27] 对图片进行特征提取，该模型表现出卓越的图像标记能力和强大的零样本泛化能力；并且 RAM 自动识别 6400+ 个常用标签，比 Open Images V6 涵盖更多有价值的类别。通过使用该技术将生活日志中的图片内容转化为文字标签[28]。Recognize -anything 工作代码可在 Git Hub 中找到[29]。对于视频格式数据，通过

对视频进行预处理，调整大小、剪裁、帧率控制，并通过贝叶斯决策方法与帧间差分法相结合，划分视频图像，对视频进行选帧，选出最具有代表的图片[30]。这里只对该技术进行应用。

Table 2. Data set details

表 2. 数据集详情

| | 特征组成 | 训练集/条 | 测试集/条 | 验证集/条 | 类别 |
|--------------|--|-------|-------|-------|----|
| DCP-Life log | city, position, Describe, image_labels, video_labels, date | 9485 | 2000 | 2000 | 26 |
| TS-Life Log | Describe, image_labels, video_labels, date | | | | |

4. TS-CNN-Bi LSTM 模型

本文构建所提出的模型，能够对多模态生活日志中的用户行为进行活动分类。首先，使用 **Tokenizer** 分词器对文本进行分词和标记化，将文本转换为序列，每个单词对应一个标记。使用 **Embedding** 层将标记化的文本转换为词嵌入表示。然后经过处理的词向量通过 **CNN** 层，利用卷积层和池化层提取文本中的局部特征。这有助于模型捕捉文本中的重要信息和特征。接着提取的局部特征经过 **Bi LSTM** 层，用于建模文本中的上下文联系。**Bi LSTM** 能够有效地处理时序数据，并捕捉长距离的依赖关系。通过 **CNN** 与 **Bi LSTM** 的结合，模型能够捕捉整个句子的全局关系，从而更好地理解文本的语义信息。最后，将得到的文本表示传入全连接层进行活动分类。通过 **Soft Max** 函数将分类结果映射到(0, 1)概率区间，输出对应的活动类别。

4.1. 词嵌入

词嵌入是将单词或子词转换为向量表示的过程。在自然语言处理中，单词或子词通常被表示为一个高维度的稀疏向量，其中每个维度对应一个单词或子词的特征。例如，在一个包含 100 个单词的词表中，每个单词可以表示为一个大小为 100 的稀疏向量，其中只有一个维度为 1，其余维度都为 0。这种高维度的稀疏表示不仅浪费存储空间，而且难以计算单词之间的相似度。因此，词嵌入技术被用来将单词或子词映射到一个低维度的稠密向量空间中，从而可以更有效地表示和计算单词之间的相似度。

生活日志通常是由一系列单词或子词组成的，而分词器的任务就是将这些单词或子词从文本中分离出来，并将它们转换为计算机可以处理的数字表示。**Tokenizer** 分词算法是将文本转换为单词或子词序列的过程。基于 **Tokenizer** 可以将文本转换成独立的 **token** 列表，进而转换成输入的向量成为计算机可以理解的输入形式。在分类任务中分词器和词嵌入技术通常是联合使用的，以将文本转换为计算机可以处理的数字表示。使用分词器将文本分成单词序列，然后使用词嵌入技术将每个单词表示为一个向量，最后将这些向量输入到神经网络中进行分类。

4.2. TS-CNN-Bi LSTM 模型框架

卷积神经网络 - 双向长短期记忆网络模型(**TS-CNN-Bi LSTM**)是一种融合了卷积神经网络和双向长短期记忆网络的深度学习模型。如图 1 所示，在这个模型中，**CNN** 负责提取高维数据的特征，如图像或者其他类型的数据，而 **Bi LSTM** 则专门用于处理时序特征，例如文本或者时间序列数据。这两个模型都具有强大的非线性拟合能力，能够自动地提取数据中的关键特征。通过将 **CNN** 和 **Bi LSTM** 进行融合，可以充分发挥深度学习模型在特征提取方面的优势。

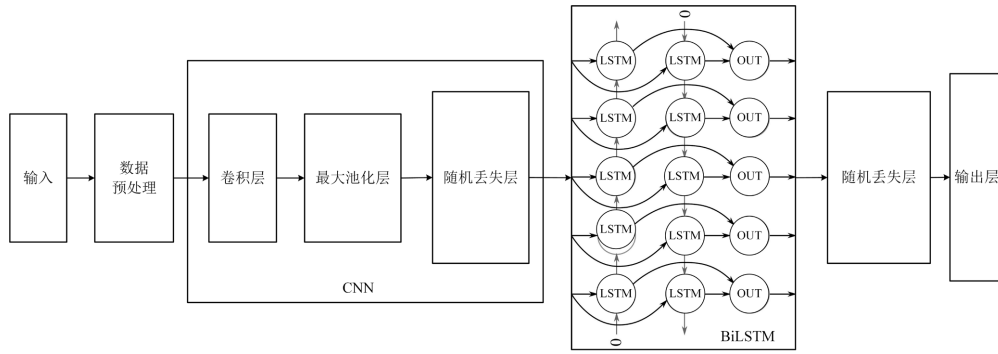


Figure 1. TS-CNN-Bi LSTM model structure

图 1. TS-CNN-Bi LSTM 模型结构

4.2.1. CNN 模型

卷积层：由于其学习文本、图像和语音等数据中独特表征的能力，CNN 越来越受欢迎。CNN 中最重要的单元是卷积层，它利用卷积核对输入进行卷积操作。这种操作可以理解为一种过滤器的工作，它能够捕捉输入数据中的局部特征。随后，经过非线性激活函数的激活，使得卷积层能够更好地学习数据的抽象特征表达。

$$a_{i,j} = f \left(\sum_{m=1}^M \sum_{n=1}^N w_{m,n} \cdot x_{i+m,j+n} + b \right) \quad (1)$$

其中 $a_{i,j}$ 为对应的激活， $w_{m,n}$ 为卷积核的 $m \times n$ 矩阵， $x_{i+m,j+n}$ 表示与神经元 (i,j) 。 b 为偏置值， f 为非线性函数。

在本文中，卷积层采用整流线性单元(ReLU)来计算特征映射，其非线性函数定义为

$$\sigma(x) = \max(0, x) \quad (2)$$

池化层：卷积层输出的特征向量通常具有较高的维度，会给模型造成较大的运算压力。池化层主要负责对卷积层提取到的特征数据进行选择 and 过滤，可以降低卷积层获取的信息维度，这不仅可以缓解模型运算压力，而且可以防止模型出现过拟合现象。

常见的池化操作有三种，最大池化(Max Pooling)、最小池化(Minimum Pooling)和平均池化操作(Average Pooling)，本文模型采用了最大池化。最大池化的原理是在输入数据的每个子区域中选择最大的特征值作为输出。具体而言，最大池化将输入数据分割成不重叠的区域(通常是矩形区域)，然后在每个区域中选择最大的特征值。这样就可以减少数据的维度，保留最显著的特征。最大池化的公式可以表示为

$$\max \text{ Pooling}(X_{i,j}) = \max(X_{i,j}[p,q]) \quad (3)$$

其中 $X_{i,j}$ 为输入数据的第 i 行、第 j 列的子区域， $X_{i,j}[p,q]$ 表示子区域中的第 p 行、第 q 列的元素。

4.2.2. Bi LSTM 模型

在 Bi LSTM 模型中，输入序列被拆分为两个方向，即正向和反向，然后在每个方向上构建一个 LSTM 单元。这种结构允许模型同时考虑到上下文信息，从而提高了模型的性能。将 CNN 的输出特征拼接在一起，形成一个更长的特征向量，然后将其传递给 Bi LSTM。Bi LSTM 模型结构可以通过以下公式表示

$$i_t = \sigma(W_{x_i} x_t + W_{h_i} h_{t-1} + b_i) \quad (4)$$

$$f_t = \sigma(W_{x_f} x_t + W_{h_f} h_{t-1} + b_f) \quad (5)$$

$$c_t = f \odot c_{t-1} \odot \tanh(\mathbf{W}_{x_c} x_t + \mathbf{W}_{h_c} h_{t-1} + b_c) \quad (6)$$

$$o_t = \sigma(\mathbf{W}_{x_o} x_t + \mathbf{W}_{h_o} h_{t-1} + b_o) \quad (7)$$

$$h_t = o_t \odot \tanh(c_t) \quad (8)$$

其中 i_t 、 f_t 、 o_t 分别为输入门、遗忘门、输出门的输出。 c_t 为细胞状态的更新， h_t 为隐藏状态的更新。 \tanh 和 σ 为激活函数， \odot 为逐个元素相乘运算。 x_t 为当前时间步的输入， h_{t-1} 为上一个时间步的隐藏状态。 \mathbf{W}_{x_i} 、 \mathbf{W}_{x_f} 、 \mathbf{W}_{x_c} 、 \mathbf{W}_{x_o} 和 \mathbf{W}_{h_i} 、 \mathbf{W}_{h_f} 、 \mathbf{W}_{h_c} 、 \mathbf{W}_{h_o} 分别为输入/输出和隐藏状态的权重矩阵。 b_i 、 b_f 、 b_c 、 b_o 为偏执项。

在 TS-CNN-Bi LSTM 模型中，输出层由一个全连接层和一个 Soft Max 分类器组成。全连通层的每个节点都与上层的节点相连，以便从上层提取的特征可以合并。在全连接层的后面是 Soft Max 分类器，它将上层的输出转换为概率向量，其值表示当前样本所属类的概率。在多分类任务中，使用交叉熵损失函数作为模型的损失函数。如果标签是独热编码形式，通常使用 Categorical Crossentropy。本文中使用的分类标签为整数形式(非独热编码)，使用 Sparse Categorical Crossentropy 作为损失函数，计算公式如下

$$H(y, \hat{y}) = -\sum_i y_i \log(\hat{y}_i) \quad (9)$$

其中， y 是真实标签的整数形式， \hat{y} 是模型的预测概率分布向量， i 表示类别的索引。

5. 实验结果

5.1. 参数设置

Table 3. Experimental parameter setting

表 3. 实验参数设置

| 层类型 | 参数名称 | 参数值 |
|--------------------|--------|----------|
| Embedding | 词向量维度 | 800 |
| | 卷积核数量 | 128 |
| Conv1D | 卷积核大小 | 4 |
| | 激活函数 | ReLU |
| | 池化窗口大小 | 4 |
| MaxPooling1D | 池化窗口大小 | 4 |
| Dropout | 丢弃比率 | 0.1 |
| Bidirectional LSTM | 隐藏单元个数 | 20 |
| | 丢弃比率 | 0.1 |
| Dense | 输出类别数 | 26 |
| | 激活函数 | Soft Max |
| | 正则化器 | L2 |
| | 学习率 | 1e-3 |

TS-CNN-Bi LSTM 模型首先使用了一个 Embedding 层，将输入的整数序列(每个整数代表一个词)映射成密集向量，其中输入维度为 8000，输出维度为 800，输入长度为 80。接下来是一个 Conv1D 层，它包含了 128 个卷积核，每个卷积核的大小为 2，激活函数为 ReLU。该卷积层用于在序列数据上进行卷积操作。MaxPooling1D 层对卷积层的输出进行下采样，池化窗口大小为 4。Dropout 层丢弃比率为 0.1，用

于在训练过程中随机丢弃部分神经元，以防止过拟合。双向 LSTM 层隐藏单元个数为 20，它可以在两个方向上处理序列数据。最后一个 Dropout 层丢弃比率同样为 0.1，用于防止过拟合。模型的最后一层是一个 Dense 层，输出类别数为训练集类别数量，激活函数为 Soft Max，同时应用 L2 正则化器。这个全连接层用于输出模型的预测结果。学习率初始为 $1e - 3$ ，优化方法选择使用 Adam 加快模型训练速度。具体参数如表 3 所示。

5.2. 模型实现

在实验中，我们使用 Keras 来构建提议的网络结构，Keras 是一个用 Python 编写的高级神经网络 API，能够在 Tensor Flow、CNTK 或 Theano 上运行。实验选择了 Tensor Flow 作为后端。模型的训练和分类都是在一台配置为 64 位 Windows 操作系统的 PC 上进行的，该 PC 拥有 2.20 GHz 的 CPU Inter Core i7-8750H、64GB RAM 和 16GB 内存的 NVIDIA QUADRO P5000 显卡。

模型的训练采用了全监督的方式，使用交叉熵损失函数来衡量预测值与真实值之间的误差。梯度从 Soft Max 层反向传播到 LSTM 层。每一层的权重和偏差都是通过随机选择的值进行初始化的。为优化模型选用 Adam 优化器作为优化算法。在模型训练过程中使用回调操作提高训练效率。1) 早停法是一种常用的防止过拟合的方法，它会在模型的性能不再提升时停止训练，从而避免了模型在训练集上的过度拟合。设置监视的指标为验证集上的损失值，当验证集损失值不再下降时，训练将被提前停止，并且在停止训练时恢复模型的最佳权重。2) 学习率调整可以自动地降低学习率，以便更好地适应数据的变化。训练过程设置了监视的指标为验证集上的准确率，当验证集准确率不再提升时，学习率将按照给定的因子进行衰减，以帮助模型更好地收敛。

5.3. 性能评估

在收集人类活动日志数据时，经常会遇到不平衡的情况，数据集中各个类别的样本数量差异较大，其中某些类别的样本数量远远多于其他类别。举例来说，假设我们收集了一份人类活动日志数据集，用于识别人类的运动行为，包括走路、跑步、骑车和驾驶等。在这个数据集中，可能会发现走路这一类别的样本数量远远大于其他类别，因为走路是人们日常生活中最常见的活动之一。这种情况导致数据集呈现不平衡的状态。如果分类器将每个实例都预测为多数类，并使用总体分类精度来评估模型性能，可能会得到较高的准确率。但是，它不能很好地反映出模型在少数类上的表现。

为了解决这个问题，可以使用 F-measure (F1 score)，它同时考虑了假阳性和假阴性。F1 Score 结合了两种测度，即“Precision”和“Recall”，它们都是基于正确识别样本总数来定义的。在信息检索领域，精确率指的是模型预测为正例的样本中有多少是真正的正例，而召回率指的是所有真正正例中有多少被模型正确地预测为正例。F1 Score 综合考虑了这两个因素，因此是一种更全面、更平衡的性能度量指标。定义 Precision 和 Recall 为

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

其中 TP 、 FP 分别为真阳性和假阳性的数量， FN 为假阴性的数量。F1 Score 通过基于样本比例对类别进行加权来抵消类别中的不平衡。F1 Score 公式如下

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

宏平均是对每个类别的性能指标进行平均，然后再求平均值。宏平均对每个类别的性能给予了相等的权重，不受样本数量不平衡的影响。在不平衡数据集中，使用宏平均可以更客观地评估模型对各个类别的性能表现。本文定义精确率的宏平均 Macro-P、召回率的宏平均 Macro-R 以及 F1 score 的宏平均 Macro-F1 对模型进行评估

$$\text{Macro-P} = \frac{1}{N} \sum_{i=1}^N \text{Precision} \quad (13)$$

$$\text{Macro-R} = \frac{1}{N} \sum_{i=1}^N \text{Recall} \quad (14)$$

$$\text{Macro-F1} = \frac{1}{N} \sum_{i=1}^N \text{F1} \quad (15)$$

其中 N 表示类别数， i 表示第 i 个类别。

5.4. 实验结果与对比分析

5.4.1. 卷积核大小对实验结果影响

为验证多卷积核尺寸的选取对活动分类最终效果的影响，将模型中的相关参数按照表 3 中设置，通过改变卷积核大小，观察三种评价指标的变化趋势。实验中选取卷积核大小分别是 1、2、3、4、5。每组卷积核数目均设为 128。实验结果如图 2 所示。

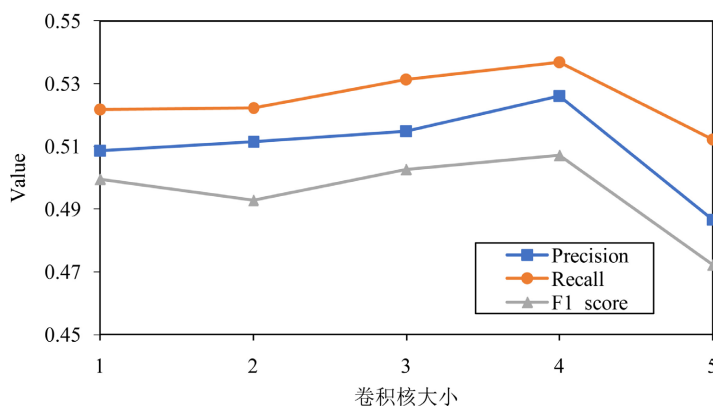


Figure 2. Experimental results for different convolutional kernel sizes
图 2. 不同卷积核大小的实验结果

使用不同尺寸卷积核的实验结果如图 2 所示，当卷积核大小为 4 时，分类效果明显优于其他数值，精确率、召回率和 F1 值分别达到了 52.61%、53.67% 和 50.72%，分类效果最佳。其它 4 组数据，在准确率和 F1 值上效果均不理想。因此 CNN 卷积核的大小取为 4。

5.4.2. Bi LSTM 隐藏单元个数对实验结果的影响

在实验 1 中，保持 CNN 通道部分的参数不变，改变 Bi LSTM 隐藏单元个数，起始值选为 10，调节幅度大小为 10，实验中选取 10、20、30、40、50 这 5 个数值，根据实验结果从中选取出 Bi LSTM 最优的隐藏单元个数。图 3 中给出了使用不同隐藏单元个数时，模型分类效果在 Precision、Recall 和 F1 值上的对比情况。在隐藏单元个数达到 30 之前，随着隐层值的增加，该模型文本分类效果越来越好，当达到 30 时，分类效果达到最好峰值，隐藏单元个数设置大于 30 分类效果明显下降。从 3 个评估标准和模型计算量等方面综合考虑，在之后的实验中 Bi LSTM 的隐层值大小统一设置为 30。

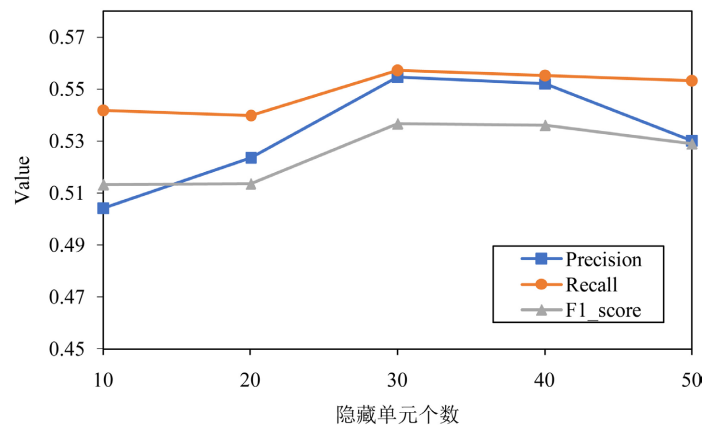


Figure 3. Experimental results for different number of hidden units
图 3. 不同隐藏单元个数的实验结果

5.4.3. 对比实验结果与分析

当涉及到用户行为分类时，以下是几种常用的分类方法介绍。本文将 TS-CNN-Bi LSTM 模型与以下方法在同一数据集上进行了对比，对比方法包括：

1) CNN (卷积神经网络): CNN 主要用于图像处理，可以用于处理时序数据，如时间序列或序列数据。它通过卷积操作在输入数据的不同区域上提取特征，并使用池化层对提取的特征进行下采样。

2) LSTM (长短期记忆网络): LSTM 是一种适用于序列数据的循环神经网络的变体，特别适用于处理长序列和捕捉序列中的长期依赖关系。它通过门控机制来控制信息的流动，包括遗忘门、输入门和输出门。LSTM 在处理文本、语音等序列数据时表现出色。

3) CNN-LSTM: CNN-LSTM 结合了卷积神经网络和长短期记忆网络的特点。通常情况下，CNN 用于提取序列中的局部特征，然后将这些特征序列输入到 LSTM 中进行处理。这种结合可以有效地捕捉序列数据中的空间和时间信息，适用于例如视频分类等任务。

4) Bi LSTM (双向长短期记忆网络): Bi LSTM 是一种具有双向结构的 LSTM 网络，它能够同时从过去和未来的信息中学习特征表示。通过在输入序列的正向和反向上构建两个独立的 LSTM 层，Bi LSTM 能够更好地捕捉到序列数据中的上下文信息，提高了模型对序列数据的理解能力。

Table 4. Results of experiments on categorization of different model activities (%)

表 4. 不同模型活动分类实验结果(%)

| 模型 | DCP-Life Log | | | TS-Life Log | | |
|----------------|--------------|---------|----------|-------------|---------|----------|
| | Macro-P | Macro-R | Macro-F1 | Macro-P | Macro-R | Macro-F1 |
| CNN | 52.72 | 52.92 | 50.76 | 49.99 | 52.72 | 49.72 |
| LSTM | 45.13 | 50.22 | 46.56 | 44.34 | 46.93 | 49.72 |
| CNN-LSTM | 48.11 | 51.77 | 49.36 | 44.83 | 48.28 | 45.14 |
| Bi LSTM | 52.73 | 56.02 | 53.00 | 50.39 | 52.37 | 49.46 |
| TS-CNN-Bi LSTM | 54.30 | 56.12 | 54.00 | 52.61 | 53.67 | 50.72 |

分析表 4 中数据可知，我们发现在相同的 DCP-Life Log 或 TS-Life Log 数据集上，本文提出的 TS-CNN-Bi LSTM 模型的分类效果较其他模型更好。具体而言，CNN 模型在两个数据集上的性能相对较为一致，Macro-F1 分别为 50.76% 和 49.72%，但并未达到理想水平。这可能是由于 CNN 模型在处理文本

数据时难以捕捉到文本的长期依赖关系和上下文信息,导致性能不佳。而 LSTM 模型在 DCP-Life Log 数据集上的性能表现也较差, Macro-F1 为 46.56%, 其低精确率(45.13%)和召回率(50.22%)反映了模型在分类过程中存在一定程度的偏差或遗漏。TS-CNN-Bi LSTM 模型的表现也不尽如人意, Macro-F1 为 49.36%, 这可能是由于两者结合的方式并未很好地利用彼此的优势。

相比之下, Bi LSTM 模型表现出更好的平衡, 其 Macro-F1 在 DCP-Life Log 数据集和 TS-Life Log 数据集分别为 53.00% 和 49.46%, 其能够利用双向上下文信息, 更好地捕获文本特征。然而, TS-CNN-Bi LSTM 模型在两个数据集上的表现最佳, 其 Macro-F1 分别达到了 54.00% 和 50.72%。这表明将 CNN 和 Bi LSTM 结合起来能够更好地利用文本的局部和全局信息, 从而取得更好的分类效果。

在 TS-Life Log 数据集上, 模型的性能普遍略低于 DCP-Life Log 数据集。这是由于 TS-Life Log 数据集中的文本特征信息较少, 缺少位置信息导致模型在分类任务上面临更大的挑战。尽管 TS-Life Log 数据集缺少位置信息, 但结合 CNN 和 Bi LSTM 的模型依然能够在这种情况下有效地提高分类性能。相比之下, 单独使用 CNN 或 LSTM 的模型在处理具有位置信息的 DCP-Life Log 数据集时性能可能会受到影响, 因为它们可能无法很好地利用位置信息。

综上所述, 结合 CNN 和 Bi LSTM 的模型能够更好地适应不同类型的文本数据集, 并取得较好的分类效果。在处理具有位置信息的 DCP-Life Log 数据集和缺少位置信息的 TS-Life Log 数据集时, TS-CNN-Bi LSTM 模型表现出色。

6. 结论和展望

我们提出了基于深度学习的 TS-CNN-Bi LSTM 组合模型, 用于生活日志数据的行为分类预测。通过利用文本、图片和视频数据, 全面地理解用户的行为和偏好。同时将时间因素纳入行为分类模型中, 以更好地捕捉用户行为随时间的变化趋势。构建两种包含不同信息的数据集, 观察用户位置信息的对分类性能的影响。实验结果表明, 与其他常用的分类方法进行了对比, TS-CNN-Bi LSTM 模型结合了卷积神经网络和双向长短期记忆网络, 综合利用局部和全局信息, 提高特征表达能力。该模型在不涉及用户位置信息, 保障用户数据隐私安全, 仍然能更全面地理解用户行为和偏好, 提高行为分类的准确性。最后, 我们希望继续利用 Liu-Life Log 数据集进行日志分类预测, 为个性化服务和行为预测提供了新的方法和思路, 提高日志数据的理解和应用价值, 促进个人生活质量的提升和更精准的行为干预策略的制定。

基金项目

辽宁省教育厅自然基金项目(LJKZ0595)。

参考文献

- [1] Cho, Y., Seo, J., Lee, H., Choi, S., Choi, A., Sung, M., *et al.* (2020) Platform Design for Lifelog-Based Smart Lighting Control. *Building and Environment*, **185**, Article 107267. <https://doi.org/10.1016/j.buildenv.2020.107267>
- [2] Ribeiro, R., Trifan, A. and Neves, A.J.R. (2022) Lifelog Retrieval from Daily Digital Data: Narrative Review. *JMIR mHealth and uHealth*, **10**, e30517. <https://doi.org/10.2196/30517>
- [3] Chakraborty, P. and Sanyal, D.K. (2023) A Comprehensive Survey of Personal Knowledge Graphs. *WIREs Data Mining and Knowledge Discovery*, **13**, e1513. <https://doi.org/10.1002/widm.1513>
- [4] Yen, A., Huang, H. and Chen, H. (2019) Personal Knowledge Base Construction from Text-Based Lifelogs. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Paris, 21-25 July 2019, 185-194. <https://doi.org/10.1145/3331184.3331209>
- [5] Tam, S., Said, R.B. and Tanriover, O.O. (2021) A ConvBilstm Deep Learning Model-Based Approach for Twitter Sentiment Classification. *IEEE Access*, **9**, 41283-41293. <https://doi.org/10.1109/access.2021.3064830>
- [6] Salido Ortega, M.G., Rodríguez, L. and Gutierrez-Garcia, J.O. (2019) Towards Emotion Recognition from Contextual

- Information Using Machine Learning. *Journal of Ambient Intelligence and Humanized Computing*, **11**, 3187-3207. <https://doi.org/10.1007/s12652-019-01485-x>
- [7] Soleimaninejadian, P., Zhang, M., Liu, Y. and Ma, S. (2018) Mood Detection and Prediction Based on User Daily Activities. 2018 *First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*, Beijing, 20-22 May 2018, 1-6. <https://doi.org/10.1109/aciiasia.2018.8470365>
- [8] Ksibi, A., Alluhaidan, A.S.D., Salhi, A. and El-Rahman, S.A. (2021) Overview of Lifelogging: Current Challenges and Advances. *IEEE Access*, **9**, 62630-62641. <https://doi.org/10.1109/access.2021.3073469>
- [9] Martín, H., Bernardos, A.M., Iglesias, J. and Casar, J.R. (2012) Activity Logging Using Lightweight Classification Techniques in Mobile Devices. *Personal and Ubiquitous Computing*, **17**, 675-695. <https://doi.org/10.1007/s00779-012-0515-4>
- [10] Akhavian, R. and Behzadan, A.H. (2016) Smartphone-Based Construction Workers' Activity Recognition and Classification. *Automation in Construction*, **71**, 198-209. <https://doi.org/10.1016/j.autcon.2016.08.015>
- [11] Chung, S., Jeong, C.Y., Lim, J.M., Lim, J., Noh, K.J., Kim, G., et al. (2021) Real-World Multimodal Lifelog Dataset for Human Behavior Study. *ETRI Journal*, **44**, 426-437. <https://doi.org/10.4218/etrij.2020-0446>
- [12] Gurrin, C., Smeaton, A.F. and Doherty, A.R. (2014) LifeLogging: Personal Big Data. Now Publishers Inc. <https://doi.org/10.1561/9781601988034>
- [13] Ziaei, A., Sangwan, A., Kaushik, L. and Hansen, J.H.L. (2015) Prof-Life-Log: Analysis and Classification of Activities in Daily Audio Streams. 2015 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, 19-24 April 2015, 4719-4723. <https://doi.org/10.1109/icassp.2015.7178866>
- [14] Tong, H., Zhang, M., Soleimaninejadian, P., Zhang, Q., Wu, K., Liu, Y., et al. (2018) Music Mood Classification Based on Lifelog. *Information Retrieval*, Guilin, 27-29 September 2018, 55-66. https://doi.org/10.1007/978-3-030-01012-6_5
- [15] Lee, S., Kang, W. and Moon, C. (2018) Lifelog-Based Classification of Mild Cognitive Impairment Using Artificial Neural Networks. 2018 *International Conference on Electronics, Information, and Communication (ICEIC)*, Honolulu, 24-27 January 2018, 1-2. <https://doi.org/10.23919/elinfocom.2018.8330611>
- [16] Li, J., Ma, W., Zhang, M., Wang, P., Liu, Y. and Ma, S. (2021) Know Yourself: Physical and Psychological Self-Awareness with Lifelog. *Frontiers in Digital Health*, **3**, Article 676824. <https://doi.org/10.3389/fdgh.2021.676824>
- [17] Nair, N., Thomas, C. and Jayagopi, D.B. (2018) Human Activity Recognition Using Temporal Convolutional Network. *Proceedings of the 5th International Workshop on Sensor-Based Activity Recognition and Interaction*, Berlin, 20-21 September 2018, 1-8. <https://doi.org/10.1145/3266157.3266221>
- [18] Jain, A. and Kanhangad, V. (2018) Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors. *IEEE Sensors Journal*, **18**, 1169-1177. <https://doi.org/10.1109/jsen.2017.2782492>
- [19] Gupta, P. and Dallas, T. (2014) Feature Selection and Activity Recognition System Using a Single Triaxial Accelerometer. *IEEE Transactions on Biomedical Engineering*, **61**, 1780-1786. <https://doi.org/10.1109/tbme.2014.2307069>
- [20] Jalloul, N., Poree, F., Viardot, G., L' Hostis, P. and Carrault, G. (2018) Activity Recognition Using Complex Network Analysis. *IEEE Journal of Biomedical and Health Informatics*, **22**, 989-1000. <https://doi.org/10.1109/jbhi.2017.2762404>
- [21] Ahmed, S., Bhuiyan, T.A., Kishi, T., Nii, M. and Kobashi, S. (2021) Human Activity Classification Based on Angle Variance Analysis Utilizing the Poincare Plot. *Applied Sciences*, **11**, Article 7230. <https://doi.org/10.3390/app11167230>
- [22] Liu, G., Cui, W., Niu, S., Ma, J. and Wang, Z. (2023) Classification of Various Daily Behaviors in Lifelog Using Deep Learning. *Procedia Computer Science*, **224**, 98-105. <https://doi.org/10.1016/j.procs.2023.09.016>
- [23] Ordóñez, F. and Roggen, D. (2016) Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors*, **16**, Article 115. <https://doi.org/10.3390/s16010115>
- [24] Shi, M., Wang, K. and Li, C. (2019) A C-LSTM with Word Embedding Model for News Text Classification. 2019 *IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*, Beijing, 17-19 June 2019, 253-257. <https://doi.org/10.1109/icis46139.2019.8940289>
- [25] Lin, Y., Le Kernec, J., Yang, S., Fioranelli, F., Romain, O. and Zhao, Z. (2018) Human Activity Classification with Radar: Optimization and Noise Robustness with Iterative Convolutional Neural Networks Followed with Random Forests. *IEEE Sensors Journal*, **18**, 9669-9681. <https://doi.org/10.1109/jsen.2018.2872849>
- [26] 陈亮, 杨建常, 刘国奇, 张立波. 一种基于事件的 Lifelog 管理模型[J]. *计算机科学与应用*, 2024, 14(1): 29-40.
- [27] <https://github.com/xinyu1205/recognize-anything>.

- [28] Zhang, Y., Huang, X., Ma, J., *et al.* (2024) Recognize Anything: A Strong Image Tagging Model. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024*, Seattle, 17-21 June, 2024, 1724-1732.
- [29] Recognize Anything: A Strong Image Tagging Model. <https://recognize-anything.github.io/>
- [30] 杨军. 基于视频处理技术的运动目标姿态检测与控制[J]. 沈阳工业大学学报, 2023, 45(1): 90-96.