

本科毕业设计[论文]

基于多任务学习的多用户负荷联合预测研究

院 系_____

专业班级_____

姓 名_____

学 号_____

指导教师_____

2023 年 5 月 5 日

摘 要

电力负荷预测是电力系统调度和管理的关键问题之一。电力负荷预测是电力系统规划的重要组成部分，也是电力系统经济运行的基础，其对电力系统规划和运行都极其重要。多个用户需要联合预测多个目标的负荷情况，以实现合理的资源分配和优化的协同感知。然而，由于不同用户之间的数据特征和预测目标的差异性，传统的单任务学习方法往往难以满足这种需求。为此，本文提出了一种基于多任务学习的多用户电力负荷联合预测方法。主要工作如下：

首先是用户负荷数据相关度分析。用户负荷数据相关度分析是电力负荷预测中的重要环节，能够帮助分析用户负荷之间的相互影响和相关性，为准确预测用户负荷提供依据和支持。对用户负荷数据进行了数据预处理和清洗，保证数据的质量和可用性。并且采用 Pearson 相关系数和 Spearman 秩相关系数等方法对用户负荷数据的相关性进行了分析和计算。

其次是多任务学习模型构建。本文提出了一种基于多任务学习的多用户负荷联合预测模型，其中多任务学习部分采用双向长短时记忆网络作为基础模型，实现了对多个任务的联合学习。该模型包括共享层和专用层两部分，多个用户的任务共享共同的参数，并且每个用户拥有自己单独的参数，以提高模型的泛化能力和对每个用户更佳的预测能力。

最后是对上述数据相关度分析方法和多任务学习模型进行实验分析，在 AMPds 数据集上进行实验验证。与单任务方法和硬参数共享方法相比，本文所设计模型具有更好的表现。

关键词：多用户负荷预测；相关度分析；多任务学习；双向长短时记忆网络

Abstract

Power load forecasting is one of the key issues in power system dispatch and management. It is an important component of power system planning and the foundation of economic operation. It plays a crucial role in power system planning and operation. Multiple users need to jointly forecast the load status for multiple targets to achieve rational resource allocation and optimized collaborative perception. However, due to the differences in data characteristics and prediction objectives among different users, traditional single-task learning methods often struggle to meet these requirements. Therefore, this thesis proposes a multi-user power load joint forecasting method based on multi-task learning. The main contributions are as follows:

Firstly, correlation analysis of user load data. Correlation analysis of user load data is a critical step in power load forecasting. It helps analyze the mutual influence and correlation among user loads, providing a basis and support for accurate load prediction. The user load data is preprocessed and cleaned to ensure data quality and usability. Pearson correlation coefficient and Spearman rank correlation coefficient methods are employed to analyze and compute the correlation of user load data.

Secondly, construction of the multi-task learning model. This thesis proposes a multi-user load joint forecasting model based on multi-task learning, where the multi-task learning part utilizes bidirectional long short-term memory networks as the foundational model to achieve joint learning of multiple tasks. The model consists of shared layers and task-specific layers, where multiple users share common parameters while each user has its own individual parameters, aiming to enhance the model's generalization ability and better prediction capability for each user.

Lastly, experimental analysis of the aforementioned data correlation analysis methods and multi-task learning model is conducted on the AMPds dataset. The

designed model outperforms single-task methods and hard parameter sharing methods, demonstrating better performance in the experiments.

Key Words: Multi-user load prediction; correlation analysis; multi-task learning; Bidirectional Long Short-Term Memory Network

目 录

摘 要	I
Abstract	II
1 绪论	1
1.1 研究背景与意义	1
1.2 研究现状	1
1.3 现有方法存在的问题	3
1.4 论文的研究内容和结构安排	4
2 用户负荷数据相关性分析	6
2.1 数据预处理	6
2.2 相关性分析	6
2.3 相关性分析方法	7
2.4 相关性分析结果	8
3 多任务学习时序预测模型	11
3.1 多任务学习时序预测理论基础	11
3.2 多任务学习预测模型结构设计	17
4 模型实验及结果分析	19
4.1 数据集预处理	19
4.2 评价标准	19
4.3 预测模型实验	20
4.4 算法结果演示	22
4.5 实验结果总结	26
5 总结与展望	27
5.1 本文内容总结	27
5.2 研究发展展望	28
致谢	错误！未定义书签。
参考文献	29

1 绪论

1.1 研究背景与意义

随着能源消费模式和需求的不断变化,电力系统的负荷管理和调度日益复杂。电力负荷预测作为电力系统调度和管理的关键问题之一,成为实现资源优化分配和协同感知的重要手段。传统的电力负荷预测方法通常采用单任务学习模型进行预测,然而,由于多用户之间负荷数据的差异性和预测目标的不同,这些方法往往难以满足联合预测的需求。因此,基于多任务学习的多用户负荷联合预测研究显得尤为重要。

多任务学习是一种机器学习技术,能够将多个任务组合成一个模型进行学习和预测。多任务学习能够有效减少模型训练和预测的时间和成本,提高模型的效率和可扩展性。相比于传统的单任务学习方法,多任务学习能够同时学习多个任务之间的关联性,提高模型的泛化能力和预测效果。在电力负荷预测中,多任务学习能够充分利用多个用户之间的相关性,提高负荷预测的准确度和稳定性,同时避免了传统方法中存在的重复数据浪费的问题。因此,多任务学习在电力负荷联合预测领域有着广阔的应用前景和重要意义。

本文所提出的基于多任务学习的多用户负荷联合预测研究,不仅能够提高电力负荷预测的准确度和效率,还具有广泛的应用前景和实际意义。例如,在智能电网和能源互联网中,多任务学习能够更好地支持多种能源的联合管理和优化调度;在工业生产中,多任务学习能够更好地支持多种负荷的协同控制和调度,实现更加高效的生产和节能减排。

1.2 研究现状

目前,电力负荷预测按预测周期的区间可分为长、中、短和超短期预测,其中短期预测对于电力负荷预测研究具有重要的意义。短期预测的时间跨度为一小时到一星期,短期预测的准确性能够指导竞争市场参与者制定短期竞价合理的方案,确保在竞争激烈的市场中获得利益。同时,短期预测也可以指导电力生产调度者进行电力调度,保障电力系统的稳定运行,并给新能源的并网提供良好的环境。除了短期预测,长期预测 also 具有重要意义。长期预测可以为电

力系统规划提供参考,为电力行业的发展提供战略支持。长期预测的时间跨度通常为数年甚至十年以上,对于电力市场的投资、规划、建设等方面有着重要的指导作用。此外,超短期预测也逐渐受到重视,它的时间跨度通常在数分钟到一小时之间,能够为电力系统提供更加精细化的电力调度和控制,提高电力系统的安全性和可靠性。综上所述,不同时间跨度的电力负荷预测都有其重要性和应用场景,需要根据具体情况进行选择和应用。

在早期的电力负荷预测领域,研究人员使用数学、统计学等方法来预测电力负荷。例如, Moghaddas-Tafreshi[1]等人使用线性回归方法进行电力负荷预测, Zheng[2]等人使用卡尔曼滤波算法进行负荷预测, Souza[3]使用自回归综合移动平均算法进行电力负荷预测。

随着机器学习技术的发展,越来越多的研究人员开始尝试使用基于机器学习的方法来进行电力负荷预测。其中,最常用的机器学习方法包括神经网络、决策树、支持向量机等。

Bhatia[4]等人基于 XGBoost 的模型,设计了滚动预测方法进行负荷预测。该模型在德国电力市场的实验中表现出良好的精度。而 Yang[5]等人采用人为选取最优输入特征的方法,结合自相关函数和最小二乘支持向量机,建立了电力负荷预测的混合模型 AS-GCLSSVM。该模型用于预测未来一周和半小时的电力负荷,作者将该模型应用于三个地区,并与其他 9 个对比模型进行了比较,结果表明 AS-GCLSSVM 显著提高了短期电力负荷预测的准确性。不过,该模型也存在一些缺点,例如耗时高、算法复杂等。

Ahmad[6]等人提出了一种基于机器学习和启发式算法的电力负荷预测模型来进行短期预测。他们使用 XGBoost 模型和决策树进行特征选择。他们将模型应用于新英格兰独立电力系统的实时电力数据集。Toubeau[7]等人基于相关策略对预测的多元分布进行抽样,改善了时间序列在分类概率预测中的依赖性。

随着人工神经网络的发展,人们发现神经网络具有很强的非线性拟合能力,因此许多研究人员将神经网络视为电力负荷预测的不错选择。Keles[8]等人使用人工神经网络来获得第二天电价的预测值。Zahedi[9]等人使用自适应模糊神经网络对加拿大安大略省的电力需求建立数学模型,并使用皮尔逊相关等方法对输入进行过滤。

目前使用最广泛的方法是深度学习。Chang[10]等人将 Adam 优化后的长短记忆网络(Long Short-Term Memory, LSTM) 应用于电价预测。实验结果表明, Adam 优化后的 LSTM 模型优于未优化的 LSTM、BP (Back Propagation, BP) 等模型。

Luo[12]等人将卷积神经网络与支持向量机进行了结合, 实验发现所提出的 RCNN-SVR 模型在电力负荷的预测中具有良好的性能。Chen[13]等人将卷积神经网络与门控循环单元 (Gated Recurrent Unit, GRU)进行混合用于电力负荷的预测, 实验发现作者提出的模型预测性能远优于传统的 BP, LSTM 等网络。Park[14]使用了强化学习模型来进行相似日的选择, 并提出了基于相似日的反向传播神经网络来进行负荷预测。实验结果表明, 作者所提出的相似日选择模型确定相似负荷日的精度可达 97.19%, 同时也大幅度提升了负荷预测的精准度。Atef[15]等人使用了深层单向 LSTM 网络和双向 LSTM 网络来预测电力负荷消耗的影响。他们使用两种叠加方法比较了不同深度的 LSTMs, 并使用优化算法对每个模型进行优化。Kumar[16]等人提出了一种基于 LSTM 和 GRU 的模型来解决电力负荷数据预测中的非线性和季节性问题。

1.3 现有方法存在的问题

以下是现有负荷预测方法存在的一些问题:

传统的数学和统计方法在处理线性预测问题时是有效的, 但在处理复杂的非线性时间序列和数据时是无效的。

预测准确度难以满足需求。传统的电力负荷预测方法主要采用统计分析和机器学习等方法, 存在模型精度低、模型拟合不足等问题, 导致预测准确度难以满足实际需求。

对数据质量要求高。电力负荷预测方法对数据质量要求较高, 而现实中电力负荷数据质量不一, 存在缺失、异常等情况, 对预测结果会产生影响。

预测周期有限。现有的电力负荷预测方法主要针对短期预测, 预测周期通常为几小时到几天, 长期预测的准确度有限。

单一模型难以满足不同负荷类型的需求。不同类型的电力负荷有着不同的特征和需求,而传统的电力负荷预测方法通常采用单一模型,难以满足多样化的需求。

另外单任务学习相对于多任务学习,存在以下缺点:

难以充分利用任务之间的相关性。单任务学习通常只针对单个任务进行优化,难以利用不同任务之间的相似性和差异性进行联合学习,导致模型的泛化能力和预测效果有限。

数据利用率低。在单任务学习中,如果训练数据只能用于一个任务,那么其他任务就不能从这些数据中获得任何信息。这会导致数据的浪费和利用率降低。

参数量大。当需要处理多个任务时,单任务学习需要建立多个独立的模型,导致参数量的增加和计算复杂度的提高。需要针对每个任务单独进行调参。在单任务学习中,每个任务需要单独进行调参以优化模型的性能,这会增加调参的难度和时间成本。

难以应对新任务。当需要处理新的任务时,单任务学习需要重新训练模型,增加了模型更新的成本和时间。

1.4 论文的研究内容和结构安排

本课题的研究目标是解决用户间负荷数据预测的问题,为了达到这个目标,本文采用了两个步骤。首先,本文对用户间负荷数据进行相关度分析,选择相关度强的用户数据用于构建多任务数据集。这个数据集包含了多个任务,每个任务都是针对不同用户的负荷预测。然后,本文采用基于多任务学习和 BiLSTM 的时序预测方法,使用 pytorch 编写深度学习模型来应用于第一部分构建的多任务数据集上进行时序预测。

在实验阶段,本文首先使用单任务学习对各个用户负荷数据进行预测,然后得出实验结果。接着,本文使用多任务学习模型进行预测,对比两者实验结果,验证多任务学习的优势。实验结果表明,基于多任务学习和 BiLSTM 的时序预测方法相对于单任务学习方法具有更好的预测性能。这是因为多任务学习可以同时学习多个任务,从而更好地利用了不同任务之间的相关性和相互影响

关系。同时,多任务学习还可以有效地缓解数据稀疏和数据不平衡的问题,提高预测的准确性和鲁棒性。

论文的结构安排如下:

第一章简述了电力负荷预测的相关背景以及研究意义,总结了现有方法中存在的一些值得改进的地方。

第二章首先进行数据预处理,之后介绍相关性分析的意义及作用,然后介绍本文章使用的三种相关性分析方法并给出数学公式。并编写相应程序进行计算得出各个用户负荷数据之间的相关度数值。

第三章首先介绍了目前常用的神经网络和时序预测模型,然后介绍多任务学习模型。并根据本文研究内容设计合适的多任务学习和时序预测模型。

第四章首先在数据集上应用单任务学习方法进行时序预测得出实验结果,然后应用硬参数共享多任务学习模型进行实验得出预测数据,最后使用硬参数共享改进多任务学习模型进行实验得出结果。将三种实验结果进行对比,并以表格的形式展示出来。

第五章分析并总结了全文的研究内容,并对展望了未来可能继续推进工作的一些方向和思路。

2 用户负荷数据相关性分析

本章将数据集进行数据预处理, 阐释数据预处理的原因与意义, 介绍相关性分析的意义, 对本文使用的两种相关性分析方法进行说明以及数学公式表达。并可视化展示相关性分析结果。

2.1 数据预处理

数据预处理是数据挖掘、机器学习、深度学习等领域中不可或缺的重要步骤。通常, 原始数据需要经过一系列的处理步骤, 包括数据清洗、缺失值处理、数据变换、特征选择、数据归一化和标准化等, 才能用于建模和分析。数据预处理的主要目的是为了提高模型的性能和效果, 同时还可以减少噪声和误差对模型的影响, 提高模型的可靠性和稳定性。

在数据预处理中, 数据清洗是最基本也是最重要的一步。由于原始数据来源的多样性和不确定性, 数据中可能存在缺失值、异常值、重复值、错误值等问题, 这些问题会严重影响到模型的准确性和稳定性, 因此需要进行数据清洗。缺失值处理是数据清洗中的一个关键问题, 缺失值可能由于数据采集的问题、数据传输的问题或者数据存储的问题而产生。缺失值处理的方法包括删除、插补和回归等。

本文所使用的 AMPds 数据集是一个公开的高精度电力数据集, 由加拿大不列颠哥伦比亚大学的电力组提供。该数据集包含来自全球不同地区的家庭的电力负荷数据, 每小时纪录一次。本文使用的共有 28 个用户的负荷数据。

此数据集不同用户数据的起止时间不同, 需要寻找具有相同时间区间的用户数据, 并进行对齐处理。经过处理后, 最终选择出 14 个用户的数据可供使用。

用户数据中存在一些缺失值问题需要进行处理, 本文采用向前插值的方式进行处理。

2.2 相关性分析

相关性分析是指通过对两个或多个变量之间的关系进行量化分析, 来确定它们之间的相关性质和程度[23]。在数据分析和数据挖掘中, 相关性分析是一

种常见的分析方法,可以帮助人们了解和发现变量之间的关系,以便更好地理解数据和进行预测分析。

在多任务学习中,任务之间的相关性分析是一个重要的研究方向。该分析旨在识别不同任务之间的相关性,以帮助优化多任务学习模型的性能。在多任务学习中,不同任务之间的相关性可以指任务之间存在的相互依赖性或共享性。这种相关性分析可以帮助我们更好地理解不同任务之间的关系,并为设计更好的多任务学习模型提供指导。

进行相关性分析可以改进多任务学习的性能:在多任务学习中,通过对任务之间的相关性进行分析,我们可以识别任务之间的相互依赖性或共享性,并针对这些依赖性或共享性进行优化。这样可以使得多任务学习模型更加高效、准确,提高模型的性能。

相关性分析可以帮助我们更好地理解不同任务之间的关系,了解它们之间的相互作用,从而更好地设计多任务学习模型。例如,我们可以通过相关性分析识别出一些任务之间的相互依赖性,然后针对这些依赖性设计特定的模型结构和算法。

相关性分析是一种重要的数据分析方法,它不仅可以帮助人们发现变量之间的关系,还可以帮助人们对数据进行清理和转换,为后续的分析和预测提供基础。相关性分析可以应用于各种领域,例如金融、医疗、工业制造等,可以帮助人们解决实际问题 and 做出决策。

在实际应用中,相关性分析可以结合数据可视化技术进行分析,例如散点图、热力图、线性回归图等,这些图表可以直观地展示变量之间的关系,并且能够发现隐藏在数据中的模式和趋势。此外,相关性分析也可以结合机器学习算法进行预测分析,例如基于回归模型的预测、基于聚类分析的数据分类等。

2.3 相关性分析方法

2.3.1 皮尔逊系数

皮尔逊系数(Pearson correlation coefficient),也称为皮尔逊相关系数,是用来衡量两个变量之间线性关系强度的统计量。其值介于-1和1之间,接近1

时表示两个变量呈正相关关系, 接近-1时表示两个变量呈负相关关系, 接近 0 时表示两个变量之间不存在线性关系。

皮尔逊系数的计算方法是将每个变量减去其均值, 然后计算两个变量之间的协方差除以它们各自的标准差的乘积。在多任务学习中, 可以利用皮尔逊系数来衡量不同任务之间的相关性, 从而选择合适的多任务学习模型和训练策略, 提高多任务学习的效果。

皮尔逊系数计算公式:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (2-1)$$

其中 x_i 代表用户一的数据, \bar{x} 代表用户一数据的平均值, y_i 代表用户二的数据, \bar{y} 代表用户二数据的平均值。

2.3.2 斯皮尔曼系数

斯皮尔曼系数 (Spearman correlation coefficient) 是一种用于衡量两个变量之间相关性的非参数方法。它基于每个变量的排名而不是原始数据值来计算相关系数。斯皮尔曼系数可以衡量两个变量是否具有单调关系 (即变量随着另一个变量的增加或减少而增加或减少, 但不一定是线性关系), 并且可以在数据集中存在异常值或偏差时提供更稳健的结果。

斯皮尔曼系数的取值范围在-1 到 1 之间, 其中-1 表示完全负相关, 0 表示无相关性, 1 表示完全正相关。计算斯皮尔曼系数的步骤如下:

(1)对每个变量的数据进行排名 (从小到大排名), 如果出现相同值, 则取平均排名。

(2)计算每个观测值的排名差 (d_i), 即第一个变量排名减去第二个变量排名。

(3)计算排名差的平方和 ($\sum d_i^2$)

(4)根据公式(2-2)计算斯皮尔曼系数 ρ , 其中 n 是数据点的数量。

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2-2)$$

2.4 相关性分析结果

根据经过编程计算得出的 14 个用户间的两个相关系数矩阵，本文选择了四组具有较强相关性的用户数据用于构建多任务数据集，分别是用户 4 和用户 19，用户 4 和用户 20，以及用户 19 和用户 20，用户 5 和用户 13。

2.4.1 皮尔逊系数结果

经过计算，用户 4 和用户 19 负荷数据之间的皮尔逊系数为 0.5，用户 19 和用户 20 之间的皮尔逊系数为 0.41，两组用户负荷数据之间具有线性相关性，且相关性比较强，因此可以用来构建多任务学习。

以下是用户负荷数据之间皮尔逊系数相关矩阵，绿色部分表示相关性较弱，黄色部分表示相关性一般，红色表示相关性较强。

	2	3	4	5	7	8	9	10	11	12	13	14	19	20
2	1	0.14	0.16	0.16	0.16	0.14	0.17	0.17	0.19	0.19	0.22	0.23	0.16	0.21
3	0.14	1	0.32	0.23	0.19	0.19	0.18	0.21	0.16	0.27	0.27	0.22	0.24	0.24
4	0.16	0.32	1	0.3	0.25	0.28	0.32	0.31	0.25	0.3	0.37	0.31	0.5	0.38
5	0.16	0.23	0.3	1	0.18	0.26	0.25	0.24	0.3	0.27	0.34	0.3	0.26	0.25
7	0.16	0.19	0.25	0.18	1	0.19	0.21	0.21	0.22	0.2	0.27	0.25	0.23	0.23
8	0.14	0.19	0.28	0.26	0.19	1	0.2	0.25	0.24	0.19	0.29	0.28	0.24	0.21
9	0.17	0.18	0.32	0.25	0.21	0.2	1	0.25	0.3	0.29	0.32	0.29	0.31	0.28
10	0.17	0.21	0.31	0.24	0.21	0.25	0.25	1	0.25	0.28	0.31	0.32	0.24	0.29
11	0.19	0.16	0.25	0.3	0.22	0.24	0.3	0.25	1	0.3	0.36	0.33	0.32	0.29
12	0.19	0.27	0.3	0.27	0.2	0.19	0.29	0.28	0.3	1	0.32	0.29	0.23	0.26
13	0.22	0.27	0.37	0.34	0.27	0.29	0.32	0.31	0.36	0.32	1	0.39	0.38	0.37
14	0.23	0.22	0.31	0.3	0.25	0.28	0.29	0.32	0.33	0.29	0.39	1	0.22	0.28
19	0.16	0.24	0.5	0.26	0.23	0.24	0.31	0.24	0.32	0.23	0.38	0.22	1	0.41
20	0.21	0.24	0.38	0.25	0.23	0.21	0.28	0.29	0.29	0.26	0.37	0.28	0.41	1

图 2.1 用户负荷数据之间皮尔逊系数相关矩阵

2.4.2 斯皮尔曼系数结果

经过计算,用户4和用户19负荷数据之间的斯皮尔曼系数为0.56,用户19和用户20之间的皮尔逊系数为0.55,用户5和用户13负荷数据之间的斯皮尔曼系数为0.49,用户4和用户20负荷数据之间的斯皮尔曼系数为0.49,四组用户负荷数据之间具有相关性,且相关性比较强,可以用来构建多任务学习。

以下是用户负荷数据之间斯皮尔曼系数相关矩阵,绿色部分表示相关性较弱,黄色部分表示相关性一般,红色表示相关性较强。

	2	3	4	5	7	8	9	10	11	12	13	14	19	20
2	1	0.34	0.37	0.38	0.34	0.37	0.29	0.37	0.31	0.32	0.47	0.42	0.32	0.4
3	0.34	1	0.37	0.36	0.28	0.29	0.28	0.27	0.27	0.37	0.37	0.29	0.29	0.33
4	0.37	0.37	1	0.38	0.39	0.38	0.28	0.33	0.27	0.26	0.46	0.33	0.56	0.49
5	0.38	0.36	0.38	1	0.34	0.42	0.36	0.39	0.43	0.35	0.49	0.4	0.33	0.37
7	0.34	0.28	0.39	0.34	1	0.37	0.32	0.35	0.34	0.24	0.43	0.34	0.37	0.38
8	0.37	0.29	0.38	0.42	0.37	1	0.3	0.44	0.38	0.27	0.48	0.41	0.31	0.34
9	0.29	0.28	0.28	0.36	0.32	0.3	1	0.28	0.44	0.27	0.37	0.27	0.34	0.32
10	0.37	0.27	0.33	0.39	0.35	0.44	0.28	1	0.34	0.32	0.46	0.43	0.26	0.36
11	0.31	0.27	0.27	0.43	0.34	0.38	0.44	0.34	1	0.29	0.46	0.34	0.3	0.33
12	0.32	0.37	0.26	0.35	0.24	0.27	0.27	0.32	0.29	1	0.35	0.32	0.13	0.24
13	0.47	0.37	0.46	0.49	0.43	0.48	0.37	0.46	0.46	0.35	1	0.47	0.43	0.49
14	0.42	0.29	0.33	0.4	0.34	0.41	0.27	0.43	0.34	0.32	0.47	1	0.18	0.32
19	0.32	0.29	0.56	0.33	0.37	0.31	0.34	0.26	0.3	0.13	0.43	0.18	1	0.55
20	0.4	0.33	0.49	0.37	0.38	0.34	0.32	0.36	0.33	0.24	0.49	0.32	0.55	1

图 2.2 用户负荷数据之间斯皮尔曼系数相关矩阵

3 多任务学习时序预测模型

多任务学习预测模型是指使用共享模型来同时预测多个任务的模型。这种模型可以将不同任务之间的相互关联性考虑在内,从而提高整体预测效果[24]。在时序预测领域中,多任务学习预测模型可以通过将多个时间序列数据作为输入,学习多个相关联的预测任务来提高预测精度。相比单任务模型,多任务学习预测模型可以通过共享参数来减少训练时间和模型参数,同时利用任务之间的相关性来提高预测精度。近年来,多任务学习预测模型在能源负荷预测、股票价格预测等领域得到了广泛应用,展现出了较好的预测效果和泛化能力。本章将构建多任务学习时序预测模型进行时序预测,介绍多任务学习理论基础和具体多任务学习模型结构设计。

3.1 多任务学习时序预测理论基础

3.1.1 神经网络

神经网络[25]是一种基于生物神经系统的结构和功能来设计的人工神经网络,也被称为人工神经元网络或者连接主义系统。它由一个或多个神经元(或称为节点或单元)组成,每个神经元都接收来自其他神经元的输入,并且通过一些计算后将结果传递给下一个神经元。神经网络的学习过程是通过对大量数据进行反复训练,使得神经网络能够从数据中提取出相应的规律和特征,从而实现分类、回归、聚类等任务。

神经网络可以分为许多不同的类型,如前馈神经网络、循环神经网络、卷积神经网络等。前馈神经网络是最基本的神经网络,也是最常用的神经网络类型,它的输入只能从输入层到输出层的单向流动,每个节点都只连接上一层和下一层的节点。循环神经网络与前馈神经网络不同,它的节点之间可以相互连接,形成环状结构,这种结构能够处理序列数据,例如时间序列和自然语言处理。卷积神经网络主要应用于图像处理和计算机视觉领域,它通过卷积运算来提取图像的特征,可以进行图像分类、目标检测等任务。

神经网络的优点是可以自动学习和提取特征，能够处理高维数据，且具有较高的准确性和泛化能力。但也存在一些缺点，如对于大规模复杂的问题，需要大量的计算资源和时间进行训练，模型的可解释性较差等[26]。

通常情况下，神经网络被分成三部分：输入层、隐藏层和输出层。每层由多个神经元组成，神经元之间通过权重相连，如下图 3.1 所示。

各层神经元之间通过加权连接进行信息传递，权值 w 用于处理信息。每个神经元通过偏置 b 来确定阈值的大小，并通过激活函数来模拟神经元的判断行为，即选择接受或拒绝该信号。其计算过程可以总结为：

$$z = f(wx + b) \quad (3-1)$$

神经网络的计算过程可以概括为：输入信号经过输入层传递给下一层，经过隐藏层的处理后最终传递到输出层输出结果。在神经网络的训练过程中，通过不断调整权值和偏置的大小以及选择合适的激活函数，以最小化误差来提高神经网络的预测能力。神经元之间的连接和权重在训练过程中被优化，以最小化误差函数。常见的优化算法包括梯度下降、随机梯度下降、自适应学习率和动量法等。

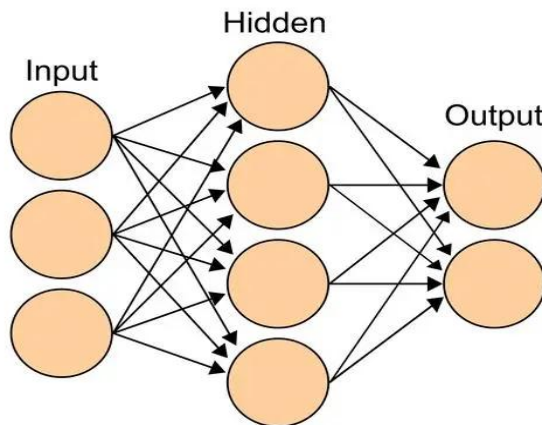


图 3.1 基础神经网络示意图

在神经网络中，激活函数[27]是将输入信号转换为输出信号的一种非线性映射。常见的激活函数有 sigmoid 函数、ReLU 函数、Leaky ReLU 函数、tanh 函数等。

sigmoid 函数是最早被使用的激活函数之一，它的形式如公式 3-2 所示，它的输出值范围在 0 和 1 之间，可以将其看做是一个概率值。但是，由于 sigmoid

函数在输入值很大或很小时会出现梯度消失的问题，而且计算过程比较复杂，会影响神经网络的运算速度，因此在深度神经网络中不太常用。

$$f(x) = \frac{1}{1 + e^{-z}} \quad (3-2)$$

ReLU 函数是目前最常用的激活函数之一，其形式如公式 3-3 所示。它的输出值范围为 0 到正无穷，计算简单，能够在保持精度的前提下大幅提高神经网络的训练速度。但是，当输入值小于 0 时，ReLU 函数的导数为 0，导致神经元无法更新，称为“死亡 ReLU”问题。

$$ReLU = \max(0, x) \quad (3-3)$$

tanh 函数也是一种常用的激活函数，其形式如公式 3-4 所示，它的输出值范围在-1 和 1 之间。与 sigmoid 函数类似，tanh 函数的梯度在输入值很大或很小时会出现梯度消失的问题，因此在深度神经网络中也不太常用。

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3-4)$$

除了以上常见的激活函数，还有一些变种如 ELU 函数、Softplus 函数、Swish 函数等，在不同的场景下选择合适的激活函数可以提高神经网络的性能。

3.1.2 多任务学习

多任务学习 (multi-task learning) [28] 是一种机器学习方法，它可以同时处理多个相关的任务，并共享特征或模型参数来提高模型的泛化能力和学习效率。与单任务学习不同，多任务学习不仅能够对一个目标任务进行建模，还可以对多个任务进行建模。

在多任务学习中，每个任务都有其独特的标签和损失函数。同时，多个任务可能共享一些相同的特征或模型层，以便更好地处理数据中的共性和异质性。共享特征的多任务学习方法可以通过将所有任务的数据馈送到一个神经网络中来实现，而多任务学习的目标是最小化所有任务的损失函数之和。

多任务学习的应用十分广泛[29]。例如，在计算机视觉中，一个图像可能涉及到目标检测、分类和分割等多个任务。在自然语言处理中，语言模型可以同时执行文本分类、命名实体识别和情感分析等多个任务。此外，多任务学习还可以应用于医疗诊断、金融风险评估、推荐系统等领域。

多任务学习的优势在于它可以提高模型的泛化能力,因为多个相关任务的学习可以通过共享模型参数来捕捉数据的共性和异质性,从而在新的数据上更加鲁棒。此外,多任务学习可以加速模型的训练过程,因为共享层可以从多个任务中共同学习,从而减少数据的需求,提高训练效率。

然而,多任务学习也面临一些挑战。例如,任务之间的关联性和相对重要性可能不太清楚,如何有效地设计多任务学习模型是一个开放性问题。此外,共享模型参数的决策可能会影响不同任务之间的性能差异,这就需要谨慎设计任务损失和模型结构。

在多任务学习中,模型通常包含多个任务模块,每个模块对应一个特定的任务。这些任务可能涉及不同的输入空间、输出空间和损失函数。任务模块之间共享一些层或参数,这些共享的层或参数被称为“多任务层”或“共享层”。

在训练过程中,多个任务的样本同时被输入到模型中,每个任务的样本都会通过各自的模块进行处理,然后在共享层处进行信息的交互和传递。共享层的存在可以使得每个任务模块得到更多的信息和知识,从而提高模型的泛化能力和效果。

具体地说,多任务学习的工作方式可以分为以下几个步骤:

(1)设计多任务模型结构:根据多个相关任务的特点和需求,设计多任务模型的结构和架构,包括任务模块和共享层。

(2)定义损失函数:对于每个任务,定义相应的损失函数,并将所有任务的损失函数加权组合为一个总的损失函数。

(3)进行训练:将多个任务的样本同时输入到模型中,通过反向传播算法进行模型参数的更新和优化。

(4)进行预测:使用已训练好的模型对新的数据进行预测,并根据每个任务的需求输出相应的结果。

需要注意的是,多任务学习的效果和性能很大程度上取决于任务之间的相关性和相互影响,以及共享层的设计和使用方式。因此,在设计多任务模型时需要充分考虑任务之间的关系,并对共享层的参数进行合理的初始化和调节。

3.1.3 多任务学习参数共享方式

多任务学习目前存在四种参数共享方式，分别是硬共享，软共享，分层共享，稀疏共享：

(1)硬共享是一种简单的共享参数策略，适合处理有较强相关性的任务，可以有效地降低过拟合的风险，但对于弱相关性的任务表现不佳。在硬共享中，多个任务共享同一组参数，即多个任务共享相同的神经网络结构和权重参数，这使得模型的参数更少，训练速度更快，但缺点是不适用于任务之间差异较大的情况。

(2)软共享是一种每个任务都有自己的模型和参数，但不同任务网络中的参数可以相互访问，并通过正则化参数之间的距离使得参数相似。这种策略需要引进更多的参数，但适用于任务之间差异较大的情况。

(3)分层共享[29]将相对较困难的任务放置在网络的高层处理，而相对较简单的任务放置在网络低层处理，可以更灵活地适应多任务学习的不同需求。然而，为多个任务设计高效的分层结构是比较困难的，需要进行复杂的网络架构设计和超参数调整。

(4)稀疏共享[30]是一种特殊的共享策略，它旨在通过选择性地共享权重来减少模型参数的数量。在这种策略中，模型的不同部分被分配不同的稀疏性级别，例如，某些部分可以被设置为高度共享，而其他部分可以被设置为低度共享或不共享。在稀疏共享中，每个任务都有自己的模型和参数，但是模型中的某些参数可以被多个任务共享，而其他参数则只属于单个任务。这样，模型可以根据任务之间的相似性自适应地调整参数共享的程度。稀疏共享的优点是可以减少模型参数数量，从而降低过拟合的风险，并提高模型的泛化能力。此外，稀疏共享还可以提高模型的可解释性，因为共享的权重可以用于解释任务之间的相似性和差异性。

以下是四种参数共享模式的示意图：

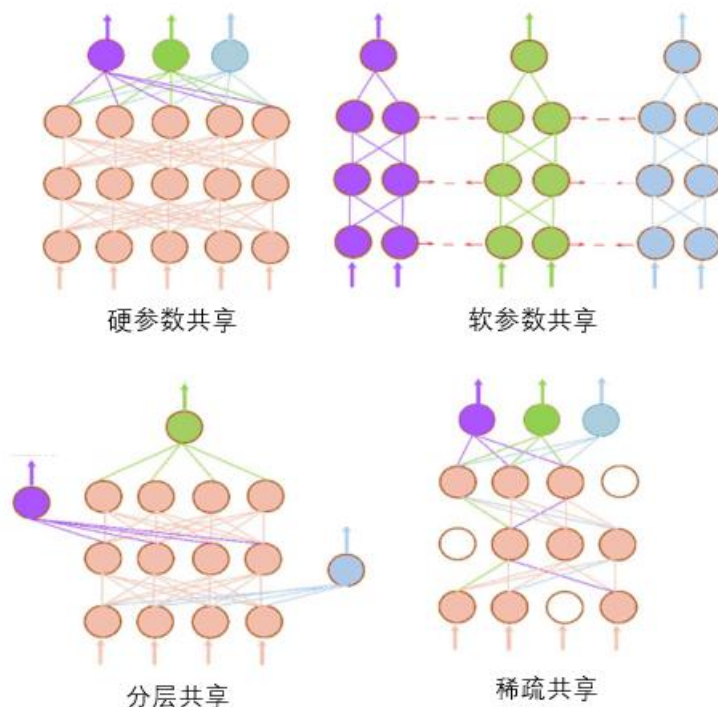


图 3.2 四种参数共享方法

3.1.4 LSTM

LSTM (Long Short-Term Memory) [31]是一种常用的循环神经网络(RNN)架构,用于处理序列数据,如语音识别、自然语言处理等领域。LSTM 由 Hochreiter 和 Schmidhuber 于 1997 年提出,是对传统 RNN 的改进。

LSTM 解决了 RNN 的“梯度消失”问题,也就是在反向传播时,误差信号随着时间步长的增加而指数级地减小,导致无法学习长期依赖关系的问题。LSTM 使用了一个称为“门”的机制,可以选择性地记忆和遗忘信息。

LSTM 的基本结构由一个单元组成,包括输入门、遗忘门、输出门和记忆单元。其中,输入门控制输入数据对记忆单元的影响,遗忘门控制前一状态的遗忘,输出门控制记忆单元对输出的影响。

LSTM 的记忆单元有一条额外的信息通道,可以保留长期的状态信息,以解决 RNN 的梯度消失问题。这条信息通道是由遗忘门和输入门控制的,当遗忘门决定将之前的状态遗忘时,输入门决定将当前状态的信息加入到记忆单元中。LSTM 的输出可以通过输出门进行调节,从而提取有用的信息。

总体来说, LSTM 通过在记忆单元中引入门控制机制, 可以选择性地保留和遗忘信息, 从而有效地学习长期依赖关系, 是一种非常有效的序列模型。

3.1.5 BiLSTM

BiLSTM (Bidirectional Long Short-Term Memory) [32] 是一种双向 LSTM 网络, 可以同时考虑过去和未来的上下文信息。在自然语言处理领域中, BiLSTM 经常用于序列标注任务, 如命名实体识别和词性标注。

BiLSTM 的结构类似于 LSTM, 但是它有两个 LSTM 层, 一个正向 LSTM 和一个反向 LSTM。正向 LSTM 处理正向序列, 即从左到右的顺序, 而反向 LSTM 处理反向序列, 即从右到左的顺序。正向 LSTM 和反向 LSTM 的输出在每个时间步被连接在一起, 形成 BiLSTM 的最终输出。

BiLSTM 的优点在于它可以考虑到整个序列的上下文信息, 而不仅仅是当前时间步的信息。它能够有效地解决一些自然语言处理中的问题, 如序列标注任务中的标注歧义和语言模型中的长期依赖问题。

它能够在时序预测任务中更好地捕捉序列中的上下文信息。相比于传统的单向循环神经网络, BiLSTM 在输入序列的正向和反向两个方向上都有一个 LSTM 单元, 因此可以同时考虑过去和未来的信息。

总之, BiLSTM 是一种非常有用的模型, 可以用于多种时序预测任务, 并且可以帮助模型更好地捕捉序列中的上下文信息, 提高预测准确度。

3.1.6 GRU

Gated Recurrent Unit (GRU)[33] 是一种类似于 LSTM 的循环神经网络结构, 它也能够解决循环神经网络在处理长序列时出现的梯度消失和梯度爆炸的问题, 同时比 LSTM 更为简单, 训练速度更快。

GRU 和 LSTM 一样, 也是由一系列门控机制来控制信息的流动, 它包含了一个重置门 (reset gate) 和一个更新门 (update gate) 来控制循环神经网络的记忆状态。与 LSTM 不同的是, GRU 只有一个记忆单元, 称为隐状态 (hidden state), 而没有 LSTM 中的细胞状态。

3.2 多任务学习预测模型结构设计

以下为本文研究所设计的多任务学习预测模型结构:

首先输入两个任务用户的负荷数据，分别是 x_1, x_2 。

每个任务具有自己独立特征提取层，特征提取层 f_1, f_2 是由全连接层和激活层构成，激活层的激活函数采用 ReLU 函数。输入的数据 x_1, x_2 经过特征提取层 f_1, f_2 计算得到 c_1, c_2 值用于后续的计算。

两个任务拥有一个共享参数的时序预测层， c_1, c_2 作为共享时序预测层的输入，经过共享预测层计算后得到结果 m 。

同时每个任务拥有自己独立的时序预测层， c_1 经过任务一的时序预测层后得到结果 m_1 ， c_2 经过任务二的时序预测层后得到结果 m_2 。

两个任务的单独预测结果 m_1, m_2 和共享任务的结果 m 分别经过各自的全连接层后得到结果，再通过加权平均计算得到最终结果。

模型结构框架图如下图 3.2 所示。

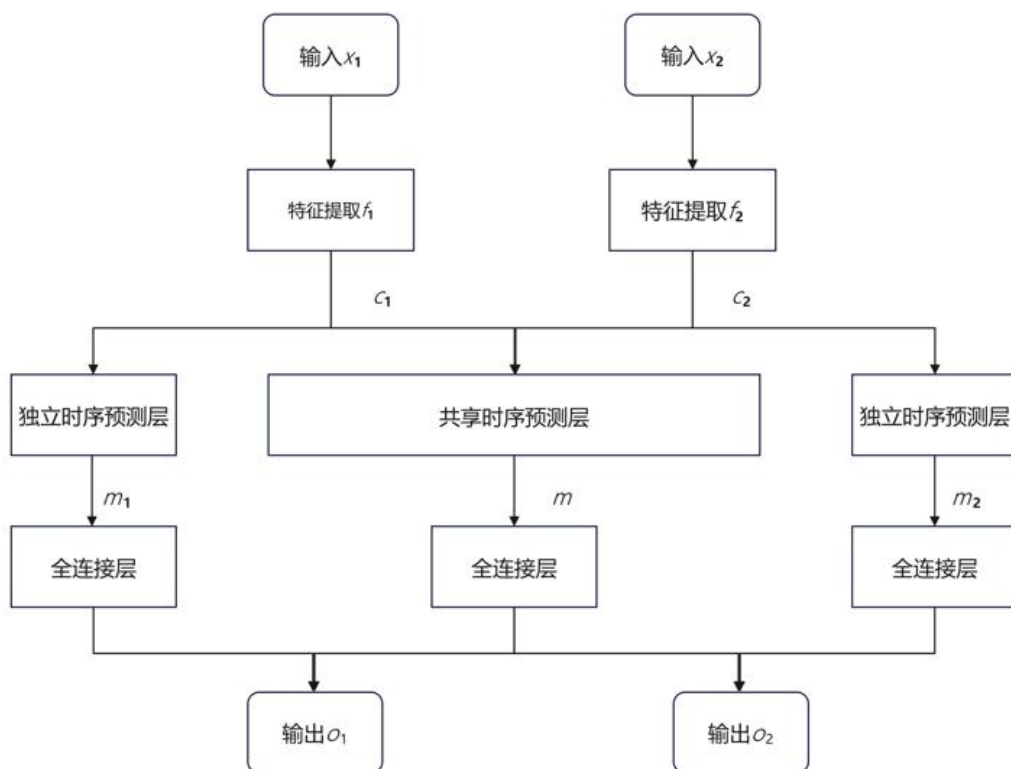


图 3.3 模型结构框架

4 模型实验及结果分析

4.1 数据集预处理

在本研究中，为了进行实验分析，我们使用了 AMPds 数据集。这个数据集包含了加拿大一户住宅的能源使用数据，涵盖了多种不同类型的电器和能源设备。然而，由于数据集时间跨度不统一，为了避免时间上的偏差，我们选择了时间跨度较为连续的 2016 年 6 月 9 日 0 时到 2018 年 1 月 29 日 23 时这段时间内的数据，共计 14401 条数据。在进行实验之前，我们还需要对所选用的数据进行预处理，以保证数据的可用性和准确性。其中，缺失值的处理是一个重要的环节。由于数据集中存在缺失值，我们采用了前向填充的方式进行处理，即使用前一天同一时间点的数据来填补缺失值，以尽可能地保留原始数据的完整性。这样处理后的数据可以更好地反映实际情况，也为后续的实验分析打下了基础。

4.2 评价标准

4.2.1 平均绝对百分比误差

本文采用平均绝对百分比误差（Mean Absolute Percentage Error, MAPE）作为预测精准度评价标准，计算公式如下所示：

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{|y_t|} \times 100\%$$

其中， y_t 是真实值， \hat{y}_t 是预测值， n 是样本数量。MAPE 的取值范围在 0 和 100% 之间。MAPE 值越小，表示模型的预测结果越准确。MAPE 衡量了预测值与真实值之间的平均百分比误差。MAPE 越小表示模型的预测结果越准确。MAPE 可以很好地反映预测误差的实际情况。因此，MAPE 的使用可以帮助我们更加客观地评估模型的预测能力，从而为实际应用提供更加准确和可靠的预测结果。

4.2.2 平均绝对误差

平均绝对误差（Mean Absolute Error, MAE）：MAE 衡量了预测值与真实值之间的平均绝对误差。MAE 越小表示模型的预测结果越准确。计算公式如下：

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t|$$

其中, y_t 是真实值, \hat{y}_t 是预测值, n 是样本数量。MAE 可以帮助我们直观的观测预测值和真实值之间的误差情况。

4.2.3 均方根误差

均方根误差 (Root Mean Squared Error, RMSE): RMSE 是 MSE 的平方根, 它也是衡量预测值与真实值之间的误差的一种指标。与 MSE 一样, RMSE 越小表示模型的预测结果越准确。

以下是均方根误差的计算公式:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2}$$

其中, y_t 是真实值, \hat{y}_t 是预测值, n 是样本数量。

4.3 预测模型实验

4.3.1 时序预测网络选择

使用三种时序预测网络进行实验, 分别是 LSTM, GRU, BiLSTM。

设置两层全连接层进行特征提取, 两层时序预测层进行时序预测, 最后通过两层全连接层输出结果。超参数 epochs 设置为 20, batch size 设置为 16, learning rate 设置为 0.01。分别对用户 4, 用户 5, 用户 13, 用户 19, 用户 20 进行预测, 并计算平均绝对百分比误差, 平均绝对误差和均方根误差。

以下是三种时序预测网络的实验结果:

表 4-1 用户 4 预测网络 MAPE 实验结果

MAPE(%)	LSTM	GRU	BiLSTM
User 4	0.2079	0.2048	0.2040
User 5	0.4486	0.4936	0.4467
User 13	0.3592	0.3058	0.2978
User 19	0.2869	0.2514	0.2476
User 20	0.3294	0.3176	0.3031

BiLSTM 在 MAPE 评价标准下每个任务都达到了最优效果, 说明在本文研究所选用的数据集上 MAPE 标准下 BiLSTM 的预测精准度优于其他两种预测方法。

表 4-2 预测网络 MAE 实验结果

MAE	LSTM	GRU	BiLSTM
User 4	0.0204	0.0199	0.0182
User 5	0.0373	0.0372	0.0364
User 13	0.0254	0.0230	0.0218
User 19	0.0019	0.0029	0.0006
User 20	0.0016	0.0170	0.0159

BiLSTM 在 MAE 评价标准下的每个任务都达到了最优效果,说明在本文研究所选用的数据集上,在 MAE 标准下 BiLSTM 预测网络的预测精准度优于其他两种预测方法。

表 4-3 预测网络 RMSE 实验结果

RMSE	LSTM	GRU	BiLSTM
User 4	0.0999	0.0975	0.0890
User 5	0.1827	0.1821	0.1781
User 13	0.1242	0.1128	0.1270
User 19	0.0093	0.0141	0.0031
User 20	0.0782	0.0832	0.0778

BiLSTM 在 RMSE 评价标准下除了用户十三的任务以外,其他的任务都达到了最优效果,说明在本文研究所选用的数据集上,在 MAE 标准下 BiLSTM 预测网络的预测精准度优于其他两种预测方法。

预测网络实验结果表明 BiLSTM 预测网络在本文研究所选用数据集上三种评价标准下效果优于其他两种预测网络,所以后续多任务学习采用 BiLSTM 网络作为时序预测网络部分构建多任务学习模型。

4.3.2 Hard Sharing-BiLSTM 多任务学习实验

首先对两个用户的数据通过各自的全连接层进行特征提取,将提取出的特征输入到一个共享的 BiLSTM 层进行时序预测。在共享层输出后,又经过了两个全连接层的处理,得到了各自的预测结果。模型中采用了硬参数共享的方式,即共享的 BiLSTM 层在两个任务中使用相同的权重,使得模型可以同时预测两个任务,减少了模型参数量,提高了模型的泛化能力。为了避免过拟合,采用了 Dropout 层对共享层和全连接层进行正则化。超参数 epochs 设置为 20, batch size 设置为 16, learning rate 设置为 0.01,根据实验结果选择了这些超参数。最

后, 分别对用户 4 和用户 19、用户 4 和用户 20、用户 5 和用户 13、用户 19 和用户 20 进行分组预测, 通过计算平均绝对百分比误差来评估模型的性能。

以下是硬参数共享实验的实验结果:

表 4-4 硬参数模型共享预测实验结果

	User4	User5	User13	User19	User20
MAPE(%)	0.2079	0.4977	0.2855	0.2124	0.3567
MAE	0.0191	0.0371	0.0219	0.0001	0.0198
RMSE	0.0935	0.182	0.1072	0.0005	0.0964

4.3.3 本文模型结构实验

同样采用两层全连接层分别对两个用户数据进行特征提取, 然后使用 BiLSTM 对每个用户进行单独预测, 再经过两层全连接层得到两个任务单独的结果; 同时共享 BiLSTM 层对两个用户进行预测, 经过全连接层得到共享预测结果。最终将共享结果与单独结果进行加权平均的得到最终结果。

超参数 epochs 设置为 20, batch size 设置为 16, learning rate 设置为 0.01。对用户 4 和用户 9, 用户 2 和用户 20, 用户 5 和用户 12 进行分组预测, 并计算平均绝对百分比误差。

表 4-5 本文模型共享预测实验结果

	User4	User5	User13	User19	User20
MAPE(%)	0.2001	0.4225	0.2814	0.2399	0.2976
MAE	0.0179	0.0366	0.0210	0.0003	0.0156
RMSE	0.0826	0.1711	0.1061	0.0017	0.0848

4.4 算法结果演示

表 4-6 预测模型 MAPE 标准实验结果对比

MAPE(%)	LSTM	GRU	BiLSTM	Hard Sharing	This Model
User 4	0.2079	0.2048	0.2040	0.2079	0.2001
User 5	0.4483	0.4936	0.4467	0.4977	0.4225
User 13	0.3592	0.3058	0.2978	0.2855	0.2814
User 19	0.2869	0.2514	0.2476	0.2124	0.2399
User 20	0.3294	0.3176	0.3031	0.3567	0.2976

在 MAPE 标准下, 本文所提出模型除了用户 19 的预测任务, 其他任务都取得了最优效果。用户 19 任务中硬参数共享的多任务学习取得了最优效果。说明在 MAPE 标准下多任务模型具有更优的表现。

表 4-7 预测模型 MAE 标准实验结果对比

MAE	LSTM	GRU	BiLSTM	Hard Sharing	This Model
User 4	0.0204	0.0199	0.0182	0.0191	0.0179
User 5	0.0373	0.0372	0.0364	0.0371	0.0366
User 13	0.0254	0.0230	0.0218	0.0219	0.0210
User 19	0.0019	0.0029	0.0006	0.0001	0.0003
User 20	0.0160	0.0170	0.0159	0.0198	0.0156

在 MAE 标准下, 用户 5 任务中 BiLSTM 模型取得最好效果, 用户 19 任务中硬参数共享模型取得最优效果, 其他用户任务本文模型都获得了最优效果。

表 4-8 预测模型 RMSE 标准实验结果对比

RMSE	LSTM	GRU	BiLSTM	Hard Sharing	This Model
User 4	0.0999	0.0975	0.0890	0.0935	0.0826
User 5	0.1827	0.1821	0.1781	0.1820	0.1711
User 13	0.1242	0.1128	0.1270	0.1072	0.1061
User 19	0.0093	0.0141	0.0031	0.0003	0.0017
User 20	0.0782	0.0832	0.0778	0.0964	0.0748

在 RMSE 标准下, 本文所提出模型除了用户 19 的预测任务, 其他任务都取得了最优效果。用户 19 任务中硬参数共享的多任务学习取得了最优效果。说明在 RMSE 标准下多任务模型具有更优的表现。

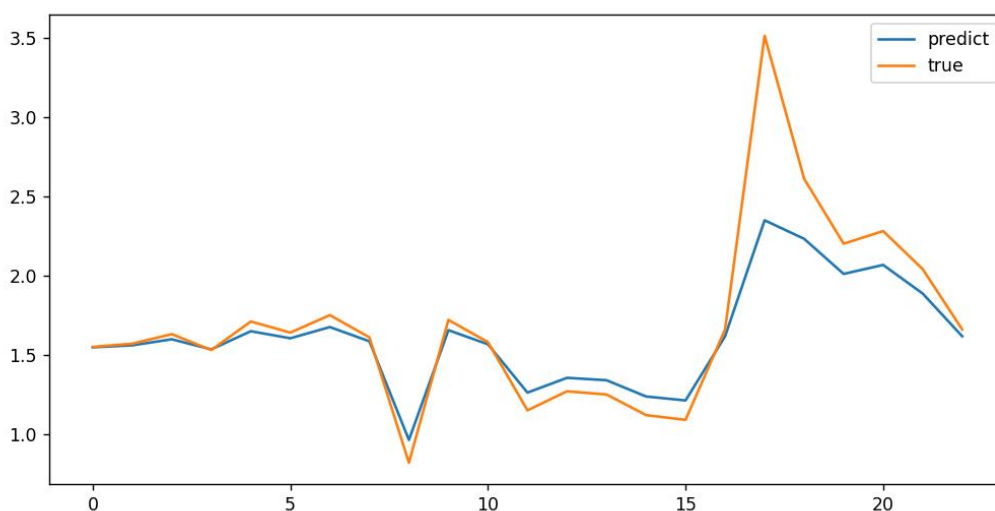


图 4.1 用户 4 预测值与真实值对比

图 4.1 是用户 4 在本文模型预测任务中 24 小时电力负荷的预测值和真实值对比, 其中黄线代表真实值, 蓝线代表预测值。

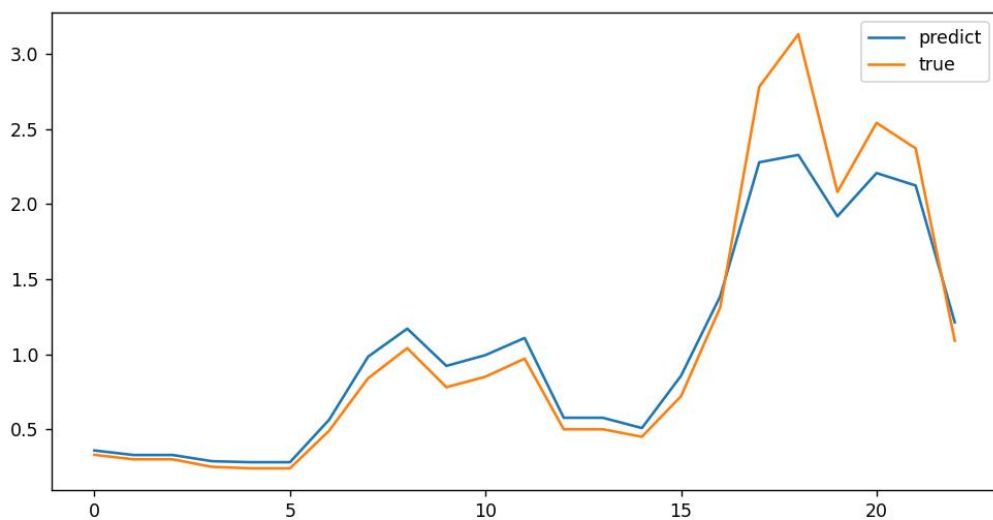


图 4.2 用户 5 预测值与真实值对比

图 4.2 是用户 5 在本文模型预测任务中 24 小时电力负荷的预测值和真实值对比，其中黄线代表真实值，蓝线代表预测值。

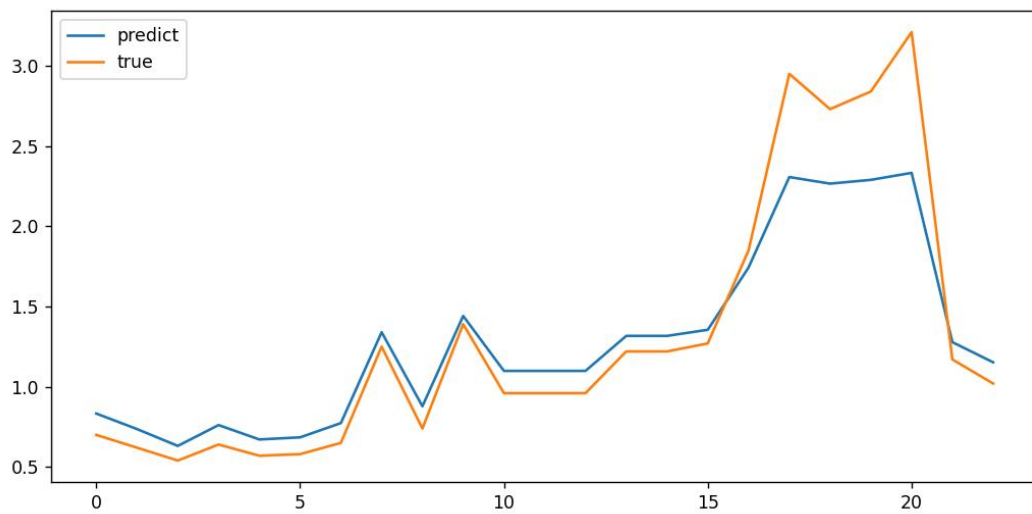


图 4.3 用户 13 预测值与真实值对比

图 4.3 是用户 13 在本文模型预测任务中 24 小时电力负荷的预测值和真实值对比，其中黄线代表真实值，蓝线代表预测值。

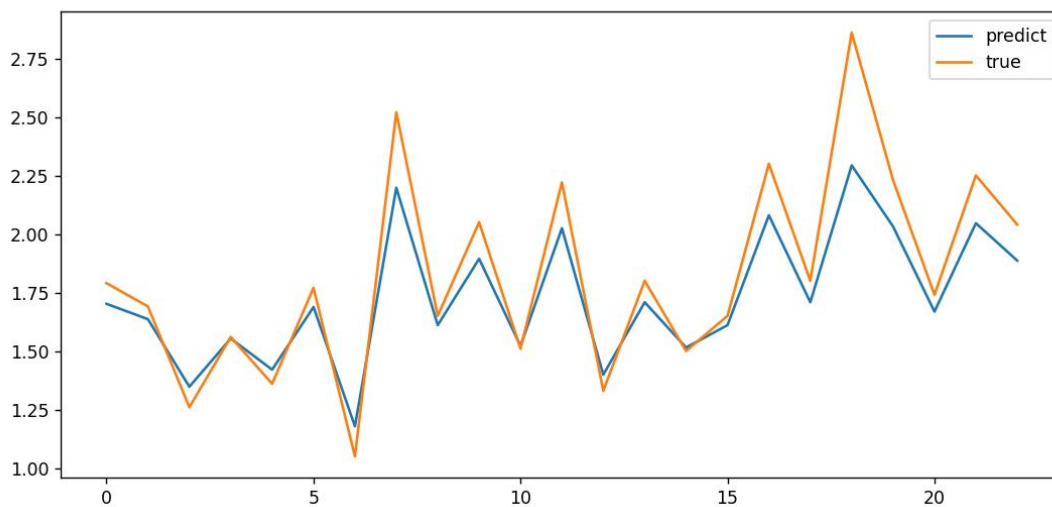
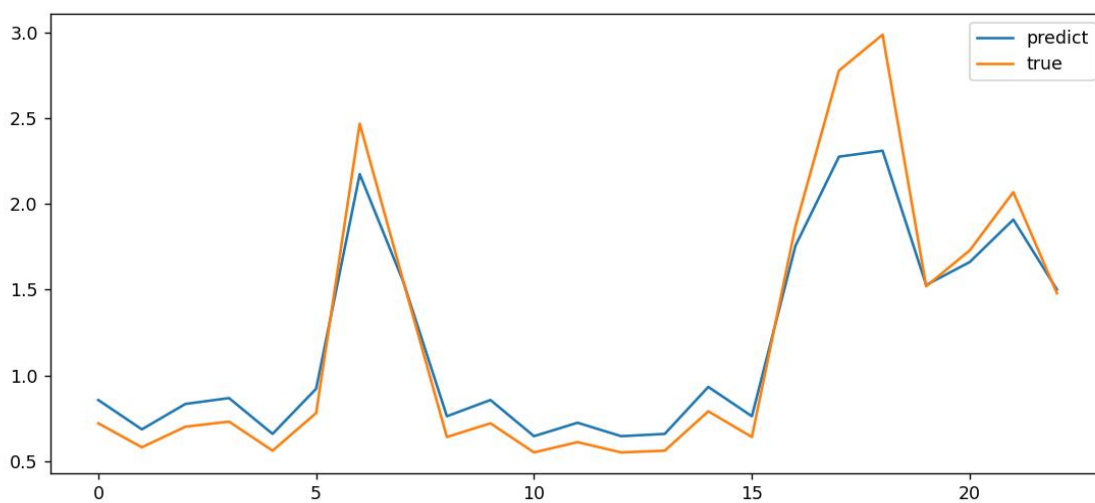


图 4.4 用户 19 预测值与真实值对比

图 4.4 是用户 19 在本文模型预测任务中 24 小时电力负荷的预测值和真实值



对比，其中黄线代表真实值，蓝线代表预测值。

图 4.5 用户 20 预测值与真实值对比

图 4.5 是用户 20 在本文模型预测任务中 24 小时电力负荷的预测值和真实值对比，其中黄线代表真实值，蓝线代表预测值。

根据图 4.1 至图 4.5 所示, 本文所构建的多任务预测模型能够实现对不同用户的联合预测。各住宅用户的用电模式存在较大差异, 但本文所提出的联合预测方法能够精准地捕捉潜在时序特征, 实现高精度地预测, 具有很好的泛化性。

4.5 实验结果总结

在本文所提出的用户数据预测任务中, 通过比较多任务方法和单独任务方法的预测结果, 可以发现多任务方法具有更好的性能。同时, 通过比较不同的多任务学习方法, 本文所提出的模型结构在大多数任务上取得了最优的性能。

具体来说, 本文所提出的模型结构采用共享 BiLSTM 层和两个任务单独的全连接层, 通过加权平均的方式得到最终结果。该模型结构可以在两个任务之间共享特征, 同时保持任务单独的输出结果。实验结果表明, 该模型在大多数任务上取得了最优的性能, 这表明通过设计合适的模型结构, 多任务学习可以在用户数据预测任务中取得较好的性能提升。

因此, 本文所提出的多任务模型可以为用户数据预测任务提供一个有效的解决方案。同时, 本研究的实验结果也说明了多任务学习在其他领域中也具有广泛的应用前景。

5 总结与展望

5.1 本文内容总结

本文主要探讨了基于多任务学习的多用户负荷联合预测问题，该问题是负荷预测问题，目前存在的方法大多为单任务负荷预测方法，但是由于不同用户之间的数据特征和预测目标的差异性，传统的单任务学习方法往往需要耗费更多时间和参数进行预测，无法很好的满足目前的需求。为此，本文研究了基于多任务学习的负荷预测方法，在基础的硬参数共享方法基础上进行了多任务模型的改进，并通过计算平均绝对百分比误差进行比较来评估模型预测效果。

第二章介绍了本文所使用的数据集，并讨论了数据预处理的方法，阐释数据预处理的原因与意义，介绍相关性分析的意义，对本文使用的三种相关性分析方法进行说明以及数学公式表达，并可视化展示相关性分析结果。

第三章介绍了多任务学习预测模型的理论基础，包含神经网络，多任务学习，LSTM，BiLSTM，GRU，BiGRU 等内容。其中详细介绍了多任务学习部分，介绍了多任务学习的应用，基本构成，参数共享模式，以及模型计算基本流程。同时本章还介绍了时序预测任务中的几种常见的评价指标，并给出了具体的评价方式。本章也详细介绍了本文所设计的多任务学习预测模型结构，该方法采用共享层和任务特定层相结合的模型架构，可以充分利用多个用户之间的相似性和差异性，并通过共享参数来降低模型的复杂度和训练时间。同时，该方法还可以提高负荷预测的准确性和泛化能力，因为它可以学习到不同用户之间的关系和共同特征，从而更好地进行预测。

在第四章实验部分，本研究对选定的用户数据集进行了三个模型的实验比较。首先，进行了基本的 BiLSTM 单任务预测实验，以便获得单任务预测的基准。然后，进行了硬参数共享的多任务学习预测实验，以比较共享参数对多任务预测的影响。最后，本研究采用自己提出的多任务学习预测模型进行实验，通过计算平均绝对百分比误差（MAPE），得出实验结果表格进行对比，实验结果表明，本文提出的多任务学习预测模型具有显著优势，多任务学习模型能够更好地利用任务间的关联性，从而提高负荷预测的准确性。

5.2 研究发展展望

本文虽然取得了一定的成果，但还有一些问题值得改进和进一步探究。首先是模型结构部分，虽然本文模型在大多数任务中取得较好结果但仍不能保证其通用性，在一些任务上的表现不如硬参数共享网络，可以后续设计更优的网络结构。其次是预测部分所选择的网络，本文选用的是 BiLSTM 网络进行预测，还可以选用其他预测效果更优的网络进行预测，也许会得到更优的结果。另外，本文在数据预处理过程中采用了前向填充的方法处理缺失值，这种方法可能会引入一定的噪声影响预测结果的准确性，因此可以考虑使用更高级的数据插值方法来处理缺失值。此外，本文所选用的数据集仅包含少量用户数据，未来可以考虑扩展数据集范围并进行更全面的实验和验证。最后，本文的研究仅限于用户负荷联合预测问题，还可以考虑将多任务学习应用于其他相关领域的问题中进行探究。

参考文献

- [1] Moghaddas-Tafreshi S M , Farhadi M . A linear regression-based study for temperature sensitivity analysis of Iran electrical load.Industrial Technology, 2008. ICIT 2008. IEEE International Conference on. IEEE, 2008.
- [2] Zheng Z, Chen H, Luo X. A Kalman filter-based bottom-up approach for household short-term load forecast. Applied Energy, 2019, 250: 882–894.
- [3] Hippert H S , Pedreira C E . Combining Neural Networks and ARIMA Models for Hourly Temperature Forecast.IEEE-INNS-ENNS International Joint Conference on Neural Networks. IEEE, 2000.
- [4] Bhatia K, Mittal R, Nisha, et al. A Multi-Phase Ensemble Model for Long Term Hourly Load Forecasting. 2020 IEEE 7th International Conference on Industrial Engineering and Applications, ICIEA 2020, 2020: 592–598.
- [5] Yang A, Li W, Yang X. Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines. Knowledge-Based Systems, 2019, 163: 159–173.
- [6] Ahmad W, Ayub N, Ali T, et al. Towards short term electricity load forecasting using improved support vector machine and extreme learning machine. Energies, 2020, 13(11): 1–17.
- [7] Toubeau J F, Bottieau J, Vallee F, et al. Deep Learning-Based Multivariate Probabilistic Forecasting for Short-Term Scheduling in Power Markets. IEEE Transactions on Power Systems, 2019, 34(2): 1203– 1215.
- [8] Agrawal R K, Muchahary F, Tripathi M M. Ensemble of relevance vector machines and boosted trees for electricity price forecasting. Applied Energy, 2019, 250: 540–548.
- [9] Zahedi G, Azizi S, Bahadori A, et al. Electricity demand estimation using an adaptive neuro-fuzzy network: A case study from the Ontario province - Canada. Energy, 2013, 49(1): 323– 328.

- [10] Chang Z, Zhang Y, Chen W. Effective Adam-Optimized LSTM Neural Network for Electricity Price Forecasting. Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS. IEEE Computer Society, 2019, 2018(11): 245–248.
- [11] Kuo P H, Huang C J. An electricity price forecasting model by hybrid structured deep neural networks. Sustainability (Switzerland), MDPI AG, 2018, 10(4):1280.
- [12] Luo F, Zhang X, Yang X, et al. Load Analysis and Prediction of Integrated Energy Distribution System Based on Deep Learning. Gaodianya Jishu/High Voltage Engineering, 2021, 47(1): 23–32.
- [13] 陈国涛, 滕欢. 基于混合神经网络深度学习的短期负荷预测. 水电能源科学, 2020, 38(4):4.
- [14] Park R J, Song K Bin, Kwon B S. Short-term load forecasting algorithm using a similar day selection method based on reinforcement learning. Energies, 2020, 13(10):2640.
- [15] Atef S, Eltawil A B. Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting. Electric Power Systems Research, 2020, 187: 106489.
- [16] Kumar S, Hussain L, Banarjee S, et al. Energy Load Forecasting using Deep Learning Approach-LSTM and GRU in Spark Cluster. Proceedings of 5th International Conference on Emerging Applications of Information Technology, EAIT 2018, IEEE, 2018: 2–5.
- [17] Wang Y F, Xiao C B, Chen Y, et al. 一种利用多任务学习的短期住宅负荷预测方案. Beijing Youdian Xueyuan Xuebao/Journal of Beijing University of Posts And Telecommunications, 2021, 44(3): 47-52.
- [18] Guo Y, Li Y, Qiao X, et al. BiLSTM Multi-Task Learning Based Combined Load Forecasting Considering the Loads Coupling Relationship for Multi-Energy System. IEEE Transactions on Smart Grid, 2022.

- [19] Zhang K, Liu Z, Zheng L. Short-term prediction of passenger demand in multi-zone level: Temporal convolutional neural network with multi-task learning. IEEE transactions on intelligent transportation systems, 2019, 21(4): 1480-1490.
- [20] Qin J, Zhang Y, Fan S, et al. Multi-task short-term reactive and active load forecasting method based on attention-LSTM model. International Journal of Electrical Power & Energy Systems, 2022, 135: 107517.
- [21] 岳伟民,刘青荣,阮应君,钱凡悦,孟华.基于 MTL-GRU-Attention 的综合能源系统多元负荷预测.电力系统及其自动化学报,2022,10:1-8.
- [22] 沈亚敏. 基于多任务深度学习的电价与负荷预测模型研究.江西理工大学,2021.
- [23] 赵沫. 基于机器学习的中国汽车销量预测研究.安徽师范大学,2021.
- [24] 林希程.基于组合预测模型电力负荷优化算法研究.电子技术与软件工程,2023,No.247(05):110-113.
- [25] 冯述虎, 侯运炳. 基于时序分析与神经网络的能源产量预测模型. 辽宁工程技术大学学报, 2003, 22(2):168-171.
- [26] 卜钰家.基于组合神经网络模型的金融时序预测.现代计算机,2022,28(22):37-41+46.
- [27] 倪志伟. BP 网络中激活函数的深入研究. 安徽大学学报: 自然科学版, 1997, 21(3): 48-51.
- [28] 曹益飞,徐焕良,吴玉强等.基于时序高光谱和多任务学习的水稻病害早期预测研究.农业机械学报,2022,53(11):288-298.
- [29] Ding K , Dong X , He Y , et al. MSSM: A Multiple-level Sparse Sharing Model for Efficient Multi-Task Learning.SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2021.
- [30] Sun T , Shao Y , Li X , et al. Learning Sparse Sharing Architectures for Multiple Tasks.AAAI Conference on Artificial Intelligence. 2020.
- [31] Wu S , Peng D . Pre-SMATS: A multi-task learning based prediction model for small multi-stage seasonal time series. Expert Systems with Application, 2022, 201: 117-121.

- [32] Poostchi H, Piccardi M. BiLSTM-SSVM: Training the BiLSTM with a structured hinge loss for named-entity recognition. *IEEE Transactions on Big Data*, 2019, 8(1): 203-212.
- [33] Qu Z, Bo X, Yu T, et al. Active and passive hybrid detection method for power CPS false data injection attacks with improved AKF and GRU - CNN. *IET Renewable Power Generation*, 2022, 16(7): 1490-1508.
- [34] 邵蔚元, 郭跃飞. 多任务学习及卷积神经网络在人脸识别中的应用. *计算机工程与应用*, 2016(52):32-37.
- [35] 李智勇, 苏寅生, 李斌,等. 基于 MT-BSGP 的电力负荷预测方法. *哈尔滨理工大学学报*, 2021, 26(4):46-55.