

**本科生毕业论文**

****

**题 目 基于多粒度残差学习和置信度估计的时序预测研究**

**学 院 计算机学院**

**专 业 计算金融交叉试验班**

**学生姓名 任凯**

**学 号 2020141460080 年级 2020级**

**指导教师 陈彦如**

**教务处制表**

**2024年4月26日**

# 声 明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得四川大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

本学位论文成果是本人在四川大学读书期间在导师指导下取得的，论文成果归四川大学所有，特此声明。

作者签名： 导师签名：

年 月 日

# 学位论文使用授权书

本学位论文作者完全了解四川大学有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或相关机构送交论文的原件、复印件和电子版，允许论文被查阅和借阅。本人授权四川大学将本学位论文的全部或部分内容编入有关数据库进行信息技术服务，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文，并用于学术活动。

（涉密学位论文在解密后适用于本授权书）

作者签名： 导师签名：

年 月 日

# 摘 要

时间序列预测在各领域中的广泛应用具有很高的实用价值，例如辅助量化投资策略的构建，以及协助电力供应方案的制定等。现有工作大多忽略了时序频率模式的多样性和动态性，导致无法充分提取多粒度特征信息，进而造成预测精度受限。然而由于跨粒度特征融合往往具有严重的信息冗余，并且不同粒度数据对预测目标的有效性总在动态变化，使用多粒度数据对时序预测目标进行建模，仍然面临着较大挑战而未被充分探索。为了解决这些问题，本研究提出了一种基于多粒度残差学习和置信度估计的神经网络，能够更为有效地发掘与融合多粒度信息，用以提高时序预测的准确性。一方面，由于信息冗余，直接使用诸如拼接或集成等简单的特征融合方式，会造成粗粒度数据中的趋势特征被不断重复，导致模型被冗余的趋势特征支配，而无法充分利用细粒度信息。因此，本研究设计了一种新颖的多粒度残差学习网络，基于粗粒度数据对细粒度数据分布的先验知识进行特征重建，并通过多粒度特征间的残差运算消除冗余信息。另一方面，为了更为准确地把握多粒度特征有效性的动态变化，本研究在多粒度残差学习网络的基础上，又引入了一个自监督目标来进行置信度估计。在电力数据、股指数据和期货数据三个真实世界数据集上所进行的大量实验表明，多粒度信息的确能显著提高时序预测性能，同时本研究所提出的模型较比其他方法具有更强的多粒度信息捕捉能力，在电力数据、股指数据和期货数据上分别比当下最优方法提升 0.5%、1% 和 1.5% 的预测精度。

**关键词**：时序预测；多粒度数据；残差学习；置信度估计

# Abstract

Time series prediction has high practical value in a wide range of applications in various fields, such as assisting the construction of quantitative investment strategies, and helping the formulation of power supply programs. Most of the existing work ignores the diversity and dynamics of time series frequency patterns, resulting in the inability to adequately extract multi-granularity feature information, which in turn leads to limited prediction accuracy. However, since cross-granularity feature fusion tends to have serious information redundancy and the effectiveness of different granularity data for prediction targets always changes dynamically, modeling time-series prediction targets using multi-granularity data still faces large challenges and is not fully explored. To address these issues, this study proposes a neural network based on multi-granularity residual learning and confidence estimation, which is able to explore and fuse multi-granularity information more effectively to improve the accuracy of time-series prediction. On the one hand, due to the redundancy of information, the direct use of simple feature fusion methods, such as splicing or integration, can result in the constant repetition of trend features in coarse-grained data, leading to the model being dominated by the redundant trend features and failing to fully utilize the fine-grained information. Therefore, in this study, a novel multi-granularity residual learning network is designed to reconstruct features based on the a prior knowledge of fine-grained data distributions from coarse-grained data and eliminate redundant information through residual operations among multi-granularity features. On the other hand, in order to more accurately capture the dynamic changes in the validity of multi-granularity features, this study introduces a self-supervised objective for confidence estimation on the basis of multi-granularity residual learning network. Extensive experiments on three real-world datasets, i.e., electricity data, stock index data and futures data, show that multi-granularity information can indeed significantly improve the performance of time-series prediction. The proposed model in this study has a stronger ability to capture multi-granularity information, and improves the prediction accuracy by 0.5%, 1%, and 1.5% compared with the SOTA methods in electricity data, stock index data, and futures data, respectively.

**Key words**: Time series prediction; Multi-granularity data; Residual learning; Confidence estimation

# 目 录

[声 明 II](#_Toc164688604)

[学位论文使用授权书 II](#_Toc164688605)

[摘 要 III](#_Toc164688606)

[Abstract IV](#_Toc164688607)

[目 录 IV](#_Toc164688608)

[第1章 绪论 1](#_Toc164688609)

[1.1 选题背景及意义 1](#_Toc164688610)

[1.2 国内外研究现状 3](#_Toc164688611)

[1.2.1 单粒度时序预测研究现状 3](#_Toc164688612)

[1.2.2 多粒度时序预测研究现状 4](#_Toc164688613)

[1.3 研究目的与创新点 5](#_Toc164688614)

[1.4 论文组织结构 6](#_Toc164688615)

[第2章 基于多粒度残差学习和置信度估计的神经网络 7](#_Toc164688616)

[2.1 引言 7](#_Toc164688617)

[2.2 单粒度和多粒度时序预测问题建模 7](#_Toc164688618)

[2.3 多粒度残差学习神经网络 8](#_Toc164688619)

[2.3.1 粒度对齐模块 9](#_Toc164688620)

[2.3.2 特征提取模块 9](#_Toc164688621)

[2.3.3 跨粒度残差堆积 10](#_Toc164688622)

[2.3.4 多粒度残差学习网络的预测结果与优化过程 10](#_Toc164688623)

[2.4 自监督的置信度估计 11](#_Toc164688624)

[2.4.1 置信度估计模块 11](#_Toc164688625)

[2.4.2 加入置信度估计后的神经网络预测结果与优化过程 13](#_Toc164688626)

[第3章 实验结果与分析 14](#_Toc164688627)

[3.1 实验设置 14](#_Toc164688628)

[3.1.1 数据集 14](#_Toc164688629)

[3.1.2 比较方法 15](#_Toc164688630)

[3.1.3 实验细节 15](#_Toc164688631)

[3.2 实验结果 16](#_Toc164688632)

[3.2.1 与各类基准方法的比较结果 16](#_Toc164688633)

[3.2.2 多粒度残差学习的有效性 17](#_Toc164688634)

[3.2.3 特征粒度选择对预测精度的影响 17](#_Toc164688635)

[第4章 总结与展望 20](#_Toc164688636)

[引用文献 21](#_Toc164688637)

# 第1章 绪论

## 1.1 选题背景及意义

基于特征数据历史观测值的时间序列预测任务，有助于探索时序未来的变化趋势，进而对当下的决策做出有利指导。对时序未来趋势的精准预测，在当代社会中的各个领域一直都发挥着广泛且深远的作用，例如对金融市场变化趋势进行预测[11][54]从而辅助微观投资决策或宏观经济政策制定、气象预测[38]辅助各行业确定生产计划、电力需求预测[48]帮助构建恰当的电力供应和销售体系，以及健康情况监测[49]协助医疗诊断等。

提高时间序列预测的准确性和稳定性，一直是学术界和工业界重点关注的问题。现有研究大多围绕基于单粒度数据进行时序预测展开[12][43][56]，也即通过提取单个特定频率的时序数据特征对具有相同频率的趋势目标进行预测，例如使用日频用电量数据预测用户未来的日用电量，或是基于股票市场中的日频“价格-交易量”数据预测每日的股价变动，又或者通过股指期货市场中分钟频的限价订单簿（Limit Order Book，LOB）数据预测分钟级的股指期货价格走势。然而需要注意的是：相同时间区间内的多粒度数据，通常能够补充原始单一粒度数据所不具备的更加丰富且详细的特征信息[44]，这些信息对于进行合理且准确的时序预测至关重要。例如在股票市场中做出投资决策时，经验丰富的股票分析师往往需要利用不同粒度（如周频、日频和分钟频）的数据，全面综合地分析股票的历史趋势和当前状态。这主要是因为长期（粗粒度）特征反映了时序低频变化的总体趋势，而短期（细粒度）特征具体刻画了时间序列在高频时间窗口内的细微变化，粗、细粒度数据特征对特定频率任务的高质量时序预测都十分关键。尽管理论上使用多粒度特征对时间序列进行建模能够提高预测性能，但是在实际中设计有效的多粒度数据时序建模方案时，仍存在两个较大的困难。

**一方面，不同粒度的数据之间存在严重的信息冗余，导致某些特定粒度数据中的有效信息被忽略，进而难以充分融合多种粒度的特征。**因为粗粒度数据通常是由对应时间段内的细粒度数据降频聚合得到的，所以细粒度数据一般会包含粗粒度数据的信息。以股票交易数据为例，大部分日频特征都是根据同一天中所有分钟频特征计算出来的，比如日频开盘价是当天第一分钟的开盘价，日频最高价是当天所有分钟频最高价的最大值，日频交易量是当天所有分钟频交易量的总和。信息冗余广泛存在于各种粒度数据之间，因此直接使用诸如拼接（Concatenation）、集成（Ensemble）等简单的特征融合方式，会造成粗粒度数据中的趋势特征被不断重复，容易导致模型被冗余的趋势特征支配，而无法充分利用细粒度信息。然而，受限于多粒度数据之间的异构性，在细粒度数据中去除冗余的粗粒度信息仍面临着无法对准等较大的困难。如何在充分提取每种粒度数据特征的同时，**尽量避免不同粒度数据之间的语义信息重叠，从而充分融合隐藏在各粒度数据中的独特信息，仍是一个较大的挑战**。



**图1.1 不同粒度数据之间存在信息冗余，细粒度数据在去除的趋势特征后其特殊信息更显著**

**另一方面，不同粒度数据对时序预测目标的有效性并非一成不变的，而是会随着时间的推移不断变化。**直观来说，各种粒度的特征数据对目标粒度的预测结果，在不同的时间段上会有截然不同的影响。例如，在预测用户日频用电量的任务中，用户的历史日频用电量数据在大多数时间段内都起着十分重要的作用。然而在夏季的用电高峰期时，用户的细粒度用电特征，比如是否使用诸如空调之类的大功率电器，可能会对耗电量产生重要影响。此时在预测日频用电量时，细粒度的时频数据可能比粗粒度的日频数据重要得多。再比如，为了预测股指期货的分钟频收益率，一般核心关注对应的分钟频 LOB 特征数据。但在刚开盘的几分钟里，由于前一天订单簿的大量堆积，秒频的 LOB 数据可能会决定期货价格的分钟频的波动。考虑到不同粒度数据对预测结果的有效性是随时间变化的，**动态判断各粒度特征对预测目标是否有充足的置信度有着重要意义**。



**图1.2 不同粒度数据对预测目标的有效性在动态变化**

综上所述，本文将从时序预测任务出发，针对使用多粒度数据所面临的两大核心困难，设计并实现能够尽量消除冗余信息，同时动态估计各粒度数据置信度的时序建模方法，解决众多时序预测场景下由于特征提取、融合不充分所导致的精度受限问题。这将有助于更准确地把握未来的时序变动趋势，进而辅助当下做出更优的生产、生活决策。

## 1.2 国内外研究现状

时序预测做为一个经典的研究问题，在过去的几十年中一直受到学术界和工业界众多研究者的重点关注。在工业数据预测[48]、气象学建模[38][41][58]、生物科学[27]、医学监测[49]、零售业商业决策[7][34][55]和金融[11][23][57][59][25]等方面，精准的时序预测都有着重要的应用和巨大的价值。根据所用特征数据频率的不同，可以将时序预测粗略地分为**单粒度时序预测**和**多粒度时序预测**：单粒度时序预测是指通过提取单个特定频率的时序特征，对具有相同频率的趋势目标进行预测；多粒度时序预测则是基于多个频率的时序特征对某一特定频率的目标进行建模。

### 1.2.1 单粒度时序预测研究现状

在与预测目标粒度相同的数据中提取时序特征，能够让模型天然地捕捉到标签的频率信息。**传统的统计学方法**和**基于深度学习的方法**都在单粒度数据预测上取得了一定的成绩，国内外学者针对这两类方法有着细致且深入的研究。

**（1）传统的统计学方法**。传统的统计学方法侧重于使用特定领域的先验知识，构建具有独特表征能力的参数模型，如自回归[5]（Auto Regression，AR），自回归积分移动平均[20]（Auto Regressive Integrated Moving Average，ARIMA)，指数平滑法[26]以及一些结构化的时间序列模型[17]。值得注意的是，传统的统计学方法大多只能刻画不同时间步特征之间的线性关系，这就导致其在拟合具有高度非线性特征的真实时间序列数据上存在固有缺陷。为了解决这一问题，部分研究使用自回归模型的变体来模拟非线性关系，如岭回归[22]（Ridge Regression）、最小二乘支持向量回归[53]（LSSVR）以及高斯过程[45]（Gaussian Process）。然而，这些模型对时间序列的特征分布或函数形式都有较为严格的假设，难以捕捉到不同形式的非线性信息[24]，在实际应用中存在较大的局限性。

**（2）基于深度学习的方法**。基于神经网络对时间序列进行建模，能够灵活地表达多种非线性关系，进而更加全面地捕捉各时间步之间的关联特征，以实现精准的预测。这些方法在大量的数据驱动下，能够动态地学习时序特征。循环神经网络（Recurrent Neural Networks，RNNs）及其变种长期短期记忆网络[21]（Long Short-Term Memory。LSTM）、门控循环单元[9]（Gated Recurrent Unit，GRU）在时序预测任务中被广泛使用，具有较强的特征提取能力、复杂的模式表征特性以及准确的长期依赖关系建模效果。在这些网络的基础上，为得到更为精确的预测，循环跳过层（LSTNet-S）和时间注意力层（LSTNet-A）[33]，以及一种新的时间模式注意力机制[47]（TPA）等更复杂模型结构都被陆续提出。部分研究[4][10]使用卷积神经网络对时序特征进行编码以增强特征提取能力，或是使用生成对抗网络学习数据的生成分布并据此计算时序的概率预测 [31]。基于自注意力机制的 Transformer 模型[52]也被广泛地应用于时序建模[30][60]，在一定程度上突破了特征输入的时间步限制，增强了信息提取能力，目前实现了较好的预测效果。

### 1.2.2 多粒度时序预测研究现状

因为单粒度数据具有天然的时序建模优势，所以以往大多数研究都是围绕单粒度数据预测而展开的。然而考虑到粗粒度数据的来源方式以及多种粒度数据的广泛存在，仅使用特定粒度的特征数据必定会导致信息的丢失。事实上，一个稳健的时序预测模型应该能够提取并融合多种粒度数据的特定信息，以全面捕获时间序列模式。为了完成使用多粒度观测数据的预测任务，研究者提出了一些线性方法。按照粗细粒度模块的处理方式，可以将这些方法大致分为以下三类：单变量响应回归模型、向量自回归模型和动态因子模型。

**（1）单变量响应回归模型**。混合数据抽样回归模型[15]是这类方法中的一个典型代表，该模型将粗粒度目标预测作为其自身滞后值和较细粒度因子对应时间段滞后值的函数。此类模型基本上与特征粒度的种类和数量无关。但需要注意的是，细粒度变量较比粗粒度而言拥有大量的滞后特征值，这会导致对应回归系数激增。因此这类研究在基础模型上还开发了各种系数约束和加权方案，以平衡模型对各粒度数据的信息提取能力。

**（2）向量自回归模型**。基于低频向量自回归模型演化出的多频率预测框架，一般需要将细粒度特征对齐到粗粒度特征上[14]。具体来说，每一种细粒度时间序列都会被“扩展”成 个新的时间序列（ 是粒度对应实际时间间隔的比值），其中第 个新的时间序列是由从索引 开始提取的每 个原始观测值经过稀疏处理得到的。显然这种方式所进行的扩展会导致模型参数的激增。



**（3）动态因子模型**。一些研究针对细粒度预测目标而搭建，对 VAR 模型进行改进，将细粒度和粗粒度数据作为一个联合系统进行演化，把对应时间段内粗粒度特征的未观测值视为缺失值并进行估算[1]。还有一些研究假设系统的动态变化被少数因子支配，因此可以先从细粒度数据中提取因子，并根据状态空间估算时序的动态变化。随后，在以粗粒度变化为响应的回归模型中，把这些因子作为特征输入。为实现粒度对齐，可以将细粒度因子在低频处汇总[13]，也可以使用与低频时间戳相对应的值[16]。

仍需强调的是，这些线性方法在对具有复杂非线性关系的多粒度时间序列进行建模时，有明显的缺陷。注意到，目前已经有一些研究[59][36]尝试构建神经网络模型，对时间序列进行“多尺度”的特征提取，然而本质上依然是针对单粒度数据进行建模，并未使用到多粒度特征。的确有一些模型采用集成[58]或者线性转换后拼接特征[20][35]的方法，将多粒度信息引入到最终的趋势预测中。但这些方法仅仅是对不同粒度的信息进行简单的整合，没有解决粒度间的信息冗余问题，也无法捕获每种粒度特征有效性的动态变化，因此在能融合的数据粒度种类上存在限制，预测精度上仍存在一定不足。

## 1.3 研究目的与创新点

前面的两节说明了多粒度数据特征在时序建模中拥有信息更充足、表征更全面的优势，但同时也面临着严重的信息冗余干扰以及置信度难以被动态衡量的两大难题。针对这些问题，本论文从工业、金融等现实场景下的多粒度时序预测出发，着重突破冗余信息消除和置信度估计两大关键技术，创新性地设计了一种基于多粒度残差学习和置信度估计（Multi-granularity Residual Learning with Confidence Estimation，MgRL\_CE）的神经网络模型，以实现对多粒度特征模式的有效探索，提高时序预测准确性。具体而言：

**（1）针对信息冗余问题，以残差运算和细粒度特征重建为切入点，提出了一种多粒度残差学习（MgRL）神经网络，尽量消除跨粒度信息融合时的冗余特征。**该网络的具体结构如图1.3所示，其包含多个结构相似的特征提取模块，每个模块负责学习与之对应的粒度数据信息。为了消除各模块之间的冗余特征，确保每个模块的输入都是对应粒度的独特信息，本网络在各模块间引入了一种新颖的残差设计。此处以粗、细两种粒度的数据为例，具体阐述残差设计的工作方式。请注意，粗粒度特征和细粒度特征都是由原始特征数据经过降频算法聚合而来的，细粒度数据的变化模式总是会受粗粒度趋势信息支配，所蕴藏的特殊的变化模式难以被直接捕捉。但也正因此，给定粗粒度数据后，可以在一定程度上获得细粒度数据分布的先验知识，并基于此重建出给定粗粒度数据下可能出现的细粒度特征。然后，通过在原始细粒度特征与重构特征之间进行残差运算，就能去除多粒度数据之间的冗余信息，进而轻松地捕捉到独特的细粒度变化模式。



**图1.3 多粒度残差学习神经网络框架图（加大一些字体）**

**（2）针对多粒度数据置信度不断变化而难以被衡量的问题，以自监督的对比学习为抓手，在 MgRL 的每个特征提取模块中都引入了一个置信度估计（CE）模块，动态估计特定粒度数据对最终预测结果的置信度权重。**引入 CE 后的特征提取模块如图 1.4 所示。具体来说，在基础优化目标之上加入了自监督的对比损失来训练一个识别器。该识别器能够通过历史趋势与当前状态之间的“时序趋势一致性”衡量各种粒度数据对预测目标的有效性。



**图1.4含有自监督置信度估计的特征提取模块**

## 1.4 论文组织结构

本论文的组织结构及各章节之间的逻辑关系具体为（**加一张框架图**）：

第一章，绪论。首先阐述了多粒度时序预测的研究背景与意义；其次，对单粒度时序预测和多粒度时序预测方法进行了调研与综述；进一步，总结了本文的研究目的与实施方案；最后，对本论文的组织结构进行了介绍。

第二章，研究方法。首先使用公式化语言对单粒度和多粒度时序预测问题进行数学建模；然后介绍了多粒度残差学习神经网络的整体结构，并详细说明了其中每个模块的具体设计方式；最后阐述了针对多粒度数据的置信度估计模块，并将其应用于多粒度残差学习框架中。

第三章，实验结果与分析。首先说明了实验设置包括数据集、对比方法和实验细节；然后展示了各对照实验的定量结果；最后对实验结果进行分析说明。

第四章，总结与展望。对全文所有的工作进行总结，并提出本文的工作不足和未来的研究方向。

# 第2章 基于多粒度残差学习和置信度估计的神经网络

## 2.1 引言

时序预测在各领域中的广泛应用具有较高的价值，众多学者针对该问题展开了深入的研究。在以往大多数工作中[9] [30][60]，使用与预测目标粒度相同的数据对未来趋势进行建模，能够让模型天然地捕捉到时间序列的频率信息。专注于提升模型对单粒度数据的特征提取能力，能够在一定程度上提高时序预测精度。

然而，考虑到常用的粗粒度数据来源以及数据粒度的多样性，仅使用特定粒度的特征数据必定会导致信息丢失。这也就意味着，只使用单粒度进行建模存在精度限制。目前有部分研究尝试挖掘多粒度特征，采用集成[58]或者线性转换后拼接特征[20][35]的方法，将多粒度信息引入到最终的趋势预测中。但是这些简单多粒度建模方式，总会或多或少地受到不同粒度间信息冗余的干扰，同时难以有效地衡量各粒度数据对预测目标的有效性，从而导致无法充分提取较细粒度的特征并且在信息的动态融合上存在较大缺陷。

因此，为了解决信息冗余干扰和各粒度有效性无法估计两大问题，本研究设计并实现了一种**基于多粒度残差学习和置信度估计的神经网络模型**。其由两个核心创新点构成：一是**多粒度残差学习神经网络**，使用跨粒度残差堆积的方式将与各粒度特征相对应的粒度对齐模块与特征提取模块进行组合，以充分提取各粒度的特殊信息；二是**自监督的置信度估计**，将估计不同粒度数据对未来趋势的置信度作为一个可学习的过程，并将其引入到多粒度残差学习神经网络的特征提取模块中，使模型能够动态估计不同粒度数据对预测目标有效性的动态变化。

本章将详细阐述基于多粒度残差学习和置信度估计的神经网络模型。首先，使用公式化的语言对单粒度和多粒度时序预测问题进行建模，以构建系统性的表达；其次，介绍神经网络模型的基础构成，即多粒度残差学习神经网络，并详细说明其中粒度对齐模块和特征提取模块的具体设计细节；最后，引出自监督的置信度估计，具体展示了置信度估计模块的理论基础，以及将其融入到基础特征提取模块中的具体设计，同时还介绍了加入置信度估计模块后的神经网络优化过程。

## 2.2 单粒度和多粒度时序预测问题建模

现有的时间序列预测研究大多基于**单粒度时序预测**展开，即使用单一粒度的历史特征来预测对应标签的未来趋势，同时特征输入和预测输出的粒度完全相同。具体而言，在单粒度时序预测过程中，预测模型学习了一个函数：

式中： 表示滞后 个时间步的 维特征。特征输入和预测输出在相邻时间步之间的实际时间间隔均为 。

事实上，针对同一时序预测问题，较细粒度数据往往包含原始粒度时序数据中所不具备的更加丰富且详细的特征信息。然而当特征输入的粒度与预测输出的粒度不一致时，模型一般难以直接捕捉到标签粒度的规律，造成预测精度受限，因此很少有研究只向模型输入最细粒度的数据。为了既能保持原始粒度的规律又能提取更为详细的特征，本文提出了一种同时使用与预测标签粒度相同的数据以及多个更细粒度数据的时序预测建模方法。

本论文的核心在于使用多粒度特征数据进行时间序列预测，即**多粒度时序预测**。具体而言，假设最粗粒度的特征与预测标签的粒度一致，多粒度预测模型将学习函数：

式中： 表示数据粒度的种类数。该函数能够将历史中的多粒度数据特征映射到反映未来趋势的标签空间里。对于每个粒度 ，数据 表示对应粒度滞后 个时间步的特征。在每一个时间步 上， 是由一个最粗粒度时间步内 个均分时间段的特征共同组成的，粒度 特征数据的实际时间间隔为 。每一种粒度数据在每个均分时间段上的特征维度均为 ，与时序预测目标粒度相同的最粗粒度数据的 一定是1。

以一个日频多粒度时序预测任务为例，该任务的 为1天，假设使用日频（最粗粒度）、12小时频以及1小时频（最细粒度）共三种粒度（）的特征数据进行建模，那么对于日频粒度数据而言： ，时间间隔为 1 天；12 小时频数据： ，时间间隔为12小时；1小时频数据： ，时间间隔为1小时。

## 2.3 多粒度残差学习神经网络

给定多粒度数据 ，其中 至 分别代表从粗粒度到细粒度的特征，多粒度残差学习就是发掘隐藏在不同粒度特征中与时间序列未来趋势相关信息的过程。多粒度残差学习神经网络框架如图1.3所示。一方面，为了充分学习各粒度数据中的信息，多粒度残差学习神经网络引入了多个结构相似的特征提取模块（图1.3中的橙色矩形），每个模块只负责发掘对应特定粒度数据中的信息。另一方面，由于不同粒度数据之间存在严重的信息冗余，这可能会导致模型被重复的粗粒度趋势信息支配，因此网络以从粗到细的级联方式堆叠特征提取模块，构建一种跨粒度残差堆积的方法，以确保每个模块的输入只包含对应特定粒度数据中的独特信息。本小节首先将详细介绍多粒度残差学习神经网络的两大基本模块；然后将阐述如何通过跨粒度残差堆积的方法将这些模块进行组合。

### 2.3.1 粒度对齐模块

由于不同粒度原始数据的维度不一致，网络首先需要通过粒度对齐模块（图 1.2 中的蓝色矩形）将它们对齐到同一维度空间上，以方便后续的残差学习运算。具体来说，本文针对滞后 个时间步粒度为 的原始特征输入 ，在网络中构造了一个简单的线性变换 进行粒度对齐并得到对齐后的特征 。粒度对齐模块的整体过程可以表述为：

显然，经过粒度对齐模块后，不同粒度原始数据的维度都被统一了。

### 2.3.2 特征提取模块

特征提取模块的作用是发掘每一种粒度数据中的特定信息，其具体结构如图 1.3 左侧所示。针对各粒度数据的特征提取模块有着完全相同的结构，本小节将以与粒度 对应的第 个特征提取模块为例，详细介绍该模块的具体运算方式。总的来说，第 个模块输入其对应的粒度特征 并产生两个输出：和。**输入** 表示粒度为 的数据在完成粒度对齐和冗余消除后的隐特征。粒度对齐过程已在 2.2.1 小节中被详细说明，3.2.3 小节将详细阐述冗余消除过程。**输出** 表示基于对应粒度 的特征所完成的时序预测结果。**输出** 表示该粒度模块对下一粒度 （更细粒度）数据对齐特征 的最优估计，它代表着模型在粒度 的特征数据中已经提取到的信息，该输出的最终目的是帮助后续更细粒度的特征提取模块去除输入中对预测没有帮助的冗余信息。

具体来说，特征提取模块内部由三个部分网络组成。第一部分是时序特征编码网络 ，它将对输入 进行进一步的时序特征提取并得到隐特征 ：

时序特征编码网络 的具体网络架构有着较多的选择，一些常用的时序特征提取模型架构均可以作为 ，本文在此处采用了双层 GRU[8]。

第二部分是**预测网络**  ，它将隐特征 的最后一时间步特征 作为输入，并输出对应特定粒度的时序预测结果 ：

第三部分是**细粒度特征重建网络** ，其将隐特征 作为输入，并输出对下一粒度数据对齐特征 的最优估计 ：

本文此处的预测网络 和细粒度特征重建网络 均为双层感知机。

### 2.3.3 跨粒度残差堆积

本研究搭建了一种新颖的残差学习方法，来解决各粒度数据之间信息冗余的问题。以往研究中所用到的残差学习方法[18][19][40]能够显式地让堆叠较深的神经网络层拟合出残差映射 而非直接拟合所需的底层映射 ，以尽量避免梯度消失的问题同时提高信息提取能力。受到这些研究的启发，本文提出基于残差学习的方式从粗粒度到细粒度堆叠 3.2.2 中介绍的特征提取模块，通过在细粒度对齐特征和使用较粗粒度数据重建出细粒度特征之间进行残差运算，以实现消除冗余信息的目标。

跨粒度残差堆积的具体过程如图 1.3 中的绿色矩形所示。与最粗粒度数据所对应的第一个特征提取模块较为特殊，由于不存在更粗粒度的特征数据，因此其输入就是粒度对齐后的特征，即 。因为粗粒度数据具有和细粒度数据分布相关的先验信息（如趋势信息等），所以本文在每个特征提取模块中都引入了细粒度特征重建网络 ，使用粗粒度数据特征重建出细粒度数据特征以模拟不同粒度之间的冗余信息，进而通过在特征输入处利用残差运算达到消除冗余的目的。具体而言，首先与粒度 对应的特征提取模块通过 会重建出 ，表示粒度 数据中所存在的和粒度 数据之间的冗余信息；然后，与粒度 对应的特征提取模块在特征输入处通过残差运算去除冗余的先验信息，从而使该模块专注于学习更细粒度数据中的特定知识。跨粒度残差堆积整体过程为：

式中： 为与粒度 对应的第 个特征提取模块的**去冗余特征输入**。为了既能从较粗粒度的数据中学习到趋势信息，又能让较细粒度的信息在残差运算后保持特异性，有必要让基于较粗粒度特征重构出的细粒度信息尽可能符合原始的细粒度信息。本文专门为此设计了一种**细粒度信息重构损失函数**：

式中： 表示向量的弗罗贝尼乌斯范数（Frobenius Norm）。

### 2.3.4 多粒度残差学习网络的预测结果与优化过程

每一个特征提取模块通过预测网络 所得到的结果 表示的是对应粒度数据的特定时序预测输出。为了综合所有粒度数据的信息进而得到最终的时序预测输出，本文参考集成学习的思想，简单地将所有粒度数据的结果平均，得到最终的预测输出 ：

针对时序预测任务，本文选取 MSE 损失作为核心优化目标，同时融合上述细粒度信息重构损失函数以及网络参数正则项，构造出了**多粒度残差学习网络的损失函数**：

式中： 是预测时间步的总数，本文所有任务的 均为 1，即仅对未来一个时间步进行预测； 和 是为了平衡各损失函数的超参数； 是用于降低网络过拟合风险的 L2 正则项。使用 Adma 算法[28]，以小批量反向传播的方式更新神经网络参数，公式（2-10）中的参数 也在此算法中进行设置。

## 2.4 自监督的置信度估计

由于时间序列中的潜在时序特征模式经常随着时间变化而变化，因此不同粒度数据对未来趋势预测的有效性也是动态变化的。所以，本文提出将估计不同粒度数据对未来趋势的置信度作为一个可学习的过程，并将其引入到 2.2 节所介绍的多粒度残差学习神经网络中，构造含有置信度估计的多粒度残差学习框架，其能够动态估计不同粒度数据的置信度从而提高时序预测能力。直观地说，某一粒度数据在近期的时序趋势越连续，使用该粒度数据预测未来的置信度就越高，因此可以通过不同粒度数据自身随时间变化的一致性来估计这种动态变化的有效性。考虑到无法直接获得衡量数据“时序趋势一致性”的真实标签，本研究依靠时间序列数据的固有特性来补充构建自监督信号。本小节首先将详细介绍置信度估计模块；然后将说明加入置信度估计后的神经网络优化过程。

### 2.4.1 置信度估计模块

如图 1.4 所示，置信度估计模块被引入到 2.2.2 小节介绍的特征提取模块中。与每一种粒度数据相对应的置信度估计模块都会输出一个权重 ，用以估算特定粒度时序预测结果 的置信度。因此，基础的多粒度残差学习神经网络本质上是将所有粒度时序预测结果的置信度权重都设为 。本研究利用**互信息**[6][32][37][42][51]（Mutual Information, MI）来衡量每种粒度数据的时序趋势一致性，如果某一粒度数据的当前状态与历史趋势之间有着相对较大的互信息，则表明该粒度数据的历史信息可能对时序预测结果更有帮助。互信息是衡量随机变量之间关系的基本量。与相关性（Correlation）不同，互信息能够捕捉变量之间的非线性统计依赖关系，因此可以度量变量间更真实的依赖关系 [3][29]。给定两个随机变量 和 ，互信息可以理解为了解 在多大程度上减少了 的不确定性，反之亦然。

将粒度为 的数据在第 个时间步上的特征表达记为 。为了充分提取隐藏在前 个时间步中的历史趋势特征，本文使用特征提取器中的自回归模型 （由于时序特征提取网络为双层 GRU 架构，因此可以直接选用 做为回归模型）来概括隐空间中的所有特征 并生成一个历史趋势潜表征 （ 本质上就是 ）。想要捕捉历史趋势和当前状态的一致性，就需要得到 与 之间互信息的最优估计：

式中： 表示历史趋势和当前状态的联合分布,即 。然而互信息的精确值难以被直接计算，因此本文仿照参考文献[39]估计互信息的最大下限做为替代：

式中：参数 是样本总数（超参），是对比损失函数，此处选用的是 ***InfoNCE* 损失**[39]：

式中： 表示从当前状态的边际分布中随机抽取的一个负样本，它来自与 属于

相同批次中其他样本（此处称之为负样本）的集合 。因为批量大小为 且其中只有 1 个正样本，因此集合 共有 个元素；另外 是针对粒度 的以 为参数的对数线性识别器：

式中：是一个是可学习的线性变换矩阵。对公式（2-13）完成优化后，可以证明最

优线性识别器的输出 与公式（2-11）中的互信息密度比 成

正比，并且在理论上与负样本数 的选择无关，接下来将证明这一点。

请注意，公式（2-13）中的 ***InfoNCE*** 损失本质就是在包含 1 个正样本和 个负样本的集合 中正样本上的分类交叉熵损失。因此，可以将优化后的

重写为概率形式 ，其中 表示样本 是正样本 ，其他的 表示负样本 。基于此，样本 是正样本的概率如下:

因此不难看出，公式（2-13）中的最优解 与互信息密度比 成正

比，且与负样本数 的选择无关。即：

又因为一对样本的点互信息（Pointwise Mutual Information, PMI）被定义为：

所以对应粒度 的最优判识别器 在最后一时间步 上所计算的当前状态 与历史趋势 的**识别分数** 就可以被近似当作该粒度数据的置信度权重：

然后使用 函数对各粒度的 PMI 值进行标准化，使值区间为 同时和为 1：

式中： 就是粒度为 数据的最终置信度权重。

### 2.4.2 加入置信度估计后的神经网络预测结果与优化过程

基于置信度权重，可以将公式（2-9）中的最终预测结果重写为置信度加权和：

组合针对时序预测准确性的 **MSE 损失函数**、**细粒度信息重构损失函数**、训练置信度估计模块的**对比损失函数**以及**网络参数正则项**，可以对公式（2-10）进行重写，得到针对 MgRL\_CE 神经网络的损失函数：

# 第3章 实验结果与分析

## 3.1 实验设置

### 3.1.1 数据集

本文针对三个从现实世界中收集的数据集展开了广泛而全面的实验：

**（1）电力数据**。UCI 电力数据集[[1]](#footnote-1)收集了2011年至2014年4年间共370位用户的每15分钟用电量（kW\*15mins，千瓦15分钟），其中部分用户在2011年之后才创建，这些用户缺失的用电量数据全部用0填充。为减弱数据缺失所造成的干扰，本研究从原始数据集中截取了2012年至2014年3年间的样本数据，同时剔除掉该区间内数据缺失量大于1天的用户样本，最终保留了其中320位用户的用电量数据。本文目标任务是**预测每位用户的次日用电量**，按照时间先后将数据集划分为训练集、验证集和测试集，分别涵盖24个月、6个月和6个月。输入网络的特征数据共有5种粒度：1天（最粗粒度）、12小时、4小时、1小时和15分钟（最细粒度）。

**（2）期货数据**。使用开源的量化投资平台 Qlib[[2]](#footnote-2) 收集了沪深300股指期货（IF）的高频交易数据，包括买卖双方5个档位的限价订单簿（Limit Order Book，LOB），交易信号的频率为0.5秒。数据集的时间范围为2022年1月4日至2022年12月30日，涵盖了2022年共242个交易日中的所有数据，每个交易日都有28800条交易记录。同样地，按照时间的先后顺序划分期货数据集：训练集、验证集和测试集分别覆盖了8个月、2个月和2个月。本研究直接对原始的 LOB数据进行建模，也即仅使用买卖双方 1至5档的价格和成交量共20个基础特征，而不再手工构建其他因子。将**预测股指期货的分钟频收益率**作为目标，其形式为 ，其中 表示股指期货在第 分钟买方1档价格和卖方1档价格的平均值。输入的特征粒度共有5种，分别是1分钟（最粗粒度）、30秒、10秒、1秒和0.5秒（最细粒度）。使用 Z-Score 法对所有特征数据进行标准化预处理。

**（3）股指数据**。基于财经数据工具包 AKShare[[3]](#footnote-3) 从公开的数据源中采集了沪深300股票指数（000300.SH）的分钟频交易数据。采集区间为2016年初至2023年底，共含有8年间的1945个交易日，每个交易日均有240条交易记录。训练集、验证集和测试集按照时间先后分别跨越6年（2016年至2021年）、1年（2022年）和1年（2023年）。此处提取了6个常用的市场因子作为特征输入，包括最高价、开盘价、最低价、收盘价、成交量和成交额，所有特征在输入模型前均通过 Z-Score 法进行了标准化处理。根据参考文献[57][61]，本文也**选择股票的日收益率作为预测目标**，即 其中 代表沪深300股指在第 天的均价。该数据集的特征粒度同样也有5种，分别是1天（最粗粒度）、1小时、15分钟、5分钟、1分钟（最细粒度）。

### 3.1.2 比较方法

所选的与本文模型进行比较的基准方法大致可分为如下四组：

**（1）经典的基于单一粒度数据的时间序列预测模型**，包括 **GRU**[8]、**LSTM**[21]、**Transformer**[52]、**DeepAR**[46] 和 **Informer**[60]。DeepAR 通过构建一种深度自回归方法能够生成准确的概率预测；Informer 则是一种基于 Transformer 的更加高效的模型。这些模型的输入都只有最粗粒度的数据。

**（2）目前效果极佳的基于单一粒度数据的股票趋势预测模型**。受傅立叶变换的启发，SFM[59] 旨在将单粒度数据分解为多频率分量，从而发掘不同频率的交易模式并最终捕获价格的变动趋势。ALSTM[43] 在基础 LSTM 上引入了一个时间注意力聚合层以提高时序信息发掘能力。ADV**-**ALSTM[12] 是 ALSTM 采用对抗学习法进行训练的一种变体，据称是目前股票日频趋势预测的 SOTA方法。

**（3）使用不同粒度数据的模型变体**。细粒度 GRU（Fine-Grained GRU）只使用了最细粒度的数据。多粒度 GRU（Multi-Grained GRU）的输入由最粗和最细两种粒度数据对齐后的特征直接拼接（Concatenation）而成。集合框架（Ensemble）表示分别使用五种粒度数据的五个单独训练模型的集合预测结果。

**（4）针对本文所提方法 MgRL\_CE 的两种消融模型**。MgRL表示本文 2.2 节中介绍的多粒度残差学习神经网络，其不含有 MgRL\_CE 中的置信度估计机制。MgRL**\_**Attention 则代表用经典的软注意力机制[2]（Soft Attention Mechanism）取代 MgRL\_CE 中置信度估计机制的模型，具体而言就是将公式（2-4）中的 作为查询，将 作为键-值对进行注意力加权计算后得到新的隐变量，再通过 网络得到对应粒度的时序预测结果 。

### 3.1.3 实验细节

为了公平比较，需要要严格控制超参数、实验平台和评价指标等实验细节：

**（1）超参数**：DeepAR 模型也选择2层 GRU 作为时序特征提取的骨架模块。其他类深度循环神经网络方法（GRU、LSTM、SFM、ALSTM、ADV-ALSTM）都采用2层结构，同时所有隐藏层大小均设置为64。对于所有模型，本文均使用小批量 Adam 算法（mini-batch Adam）进行优化直至收敛，同时通过网格搜索法在验证集上调整学习率（搜索范围为 ）。公式（2-10）中的 设为1，公式（2-20）中的 和 也均设置为1。L2 正则化系数 则在 这一范围中进行调整。Transformer 和 Informer 模型的编码器同样也是2层架构，编码器头的数量从 中进行选择。另外本研究还参考所有对比方法原文中列出的值，调整了相应模型的的超参数，每种方法均报告了最佳性能。

**（2）实验平台**：所有实验均在一张 Nvidia RTX 3090Ti 24GB GPU 上进行。

**（3）评价指标**：根据之前的一些研究[60][50][56]，本文选取了四种常用的评估指标，包括决定系数（R2）皮尔逊相关系数（CORR）、均方根误差（RMSE）和平均绝对误差（MAE）。对于 R2 和 CORR，数值越大表示预测精度越高，模型效果越好；对于 RMSE 和 MSE，数值越小则代表模型预测效果越好。

**表3.1. MgRL\_CE 以及其他各类基准方法在电力数据、期货数据和股指数据上的预测结果**

（最优结果用**粗体**表示，次优结果用下划线表示）

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **所用方法** | **电力数据** | | | |  | **期货数据** | | | |  | **股指数据** | | | |
| **R2 ↑** | **CORR ↑** | **RMSE ↓** | **MAE ↓** |  | **R2 ↑** | **CORR ↑** | **RMSE ↓** | **MAE ↓** |  | **R2 ↑** | **CORR ↑** | **RMSE ↓** | **MAE ↓** |
| GRU |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| LSTM |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Transformer |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| DeepAR |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Informer |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| SFM |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ALSTM |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ADV-LSTM |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Fine-Grained GRU |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Multi-Grained GRU |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Ensemble |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| MgRL |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| MgRL**\_**Attention |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| MgRL\_CE |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

## 3.2 实验结果

### 3.2.1 与各类基准方法的比较结果

为了验证本文所提 MgRL\_CE 模型针对时序预测任务的有效性，将其与各类当前最优的时序预测模型进行对比，既包括经典的基于单一粒度数据的时间序列预测模型，又包括目前效果极佳的基于单一粒度数据的股票趋势预测模型，还包括使用不同粒度数据的模型变体。与各类基准方法的比较结果如表1.所示。总体而言，**MgRL\_CE 模型在三个数据集上的四种评估指标基本都取得了最优结果**。具体来说，基于表1.可以得到如下发现：首先，多粒度 GRU（Multi-Grained GRU）与仅使用最粗粒度（GRU）或最细粒度（Fine-Grained GRU）的单粒度模型相比，取得了更为精准的预测结果，这证明了使用多粒度数据特征进而引入更丰富的信息的确能够提升模型性能。其次，与简单使用拼接融合多粒度信息的 Multi-Grained GRU 相比，MgRL\_CE 模型能够减少多粒度间的信息冗余，同时动态估计各粒度数据的置信度，进而更充分地提取和融合特征，得到了比基准方法更好的结果。最后，尽管多粒度残差学习神经网络 MgRL 和带有软注意力机制的 MgRL**\_**Attention 模型都呈现出了比 Multi-Grained GRU 更准确的预测结果，但是它们都仍显著差于引入置信度估计模块的 MgRL\_CE。这说明不同粒度数据对预测结果的有效性的确是动态变化的，使用固定权重静态地表示有效性具有局限性，而自监督的置信度估计模块能够较好地衡量这种动态变化。

**表3.2. MgRL 模型与两种变体的时间序列预测表现**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **所用方法** | **电力数据** | | | |  | **期货数据** | | | | |  | | **股指数据** | | | | |
| **R2 ↑** | **CORR ↑** | **RMSE ↓** | **MAE ↓** |  | **R2 ↑** | **CORR ↑** | **RMSE ↓** | **MAE ↓** |  | | **R2 ↑** | | **CORR ↑** | **RMSE ↓** | **MAE ↓** |
| MgRL\_Add |  |  |  |  |  |  |  |  |  |  | |  | |  |  |  |
| MgRL**\_**Cat |  |  |  |  |  |  |  |  |  |  | |  | |  |  |  |
| MgRL |  |  |  |  |  |  |  |  |  |  | |  | |  |  |  |

### 3.2.2 多粒度残差学习的有效性

为进一步研究多粒度残差学习神经网络的有效性，说明跨粒度残差堆积较比简单的求和或拼接在特征融合方面更具优势，在此又构建了两种 MgRL 的模型变体，分别是将跨粒度残差堆积替换为求和运算的 MgRL\_Add 和替换为拼接运算的 MgRL\_Cat。两种模型的具体架构如图3.1所示，将其与图1.3中的结果做对比，能够明显发现各种方法之间的主要区别。具体而言，变体 MgRL\_Add 是将经过粒度对齐模块后的的特征进行加和，得到 并将其输入至对应粒度的特征提取模块；变体 MgRL\_Cat 是将经过粒度对齐模块后的的特征进行拼接，得到 并将其输入至对应的特征提取模块。表2.呈现了这三种模型的预测表现，多粒度残差学习网络的精度显著优于 MgRL\_Add 和 MgRL\_Cat。这说明：所提出的跨粒度残差堆积的确有助于各特征提取模块学习对应粒度的特殊信息，从而提高了模型的时序预测精度。

### 3.2.3 特征粒度选择对预测精度的影响

本文还研究了输入特征的粒度选择对模型预测精度的影响，相关结果在表3.中。粒度选择中的 表示由粗至细的五种特征粒度。表中从上至下，首先呈现了只使用单粒度数据的模型预测结果，然后展示了将粒度增加至2种、3种以及最全面的5



**图3.1 MgRL 的两种模型变体：使用求和运算或拼接运算替代跨粒度残差堆积**

种后的预测精度。此处并未对所有的粒度组合进行验证，而是以最粗粒度 和最细粒度 作为基准，选择最有代表性的组合方式做呈现，尽可能在保证实验简洁度的同时，展示出粒度种类增加带来的影响。比如在验证3种粒度数据的组合时，在最粗和最细粒度特征的基础上又加入了差别较大的中间粒度 。所有的组合均在 MgRL 模型和 MgRL\_CE 模型（表中未使用括号标注的粒度选择）上进行了实验。从表中不难看出，随着粒度种类的增多，MgRL\_CE 模型的预测精度始终都在不断改善。而未引入置信度估计模块的 MgRL 模型在粒度种类增加的初期（从1种粒度到3种粒度）的确有性能增量，但从3种粒度增加至5种粒度时，预测精度反而明显下降。同时在各类粒度组合之下，MgRL\_CE 均显著优于 MgRL。这再次说明，各粒度特征对预测目标的有效性时动态变化的，随着粒度种类的增加，这种动态性变得越来越复杂，如果不能充分估计各粒度的置信度，会导致信息增量无法被有效识别。总的来说，跨粒度残差堆积能够较为全面地提取和融合各粒度的数据特征，置信度估计模块可以动态识别各粒度的有效性变化，从而更加充分地利用各粒度信息提高模型的预测精度。

**表3.3. 特征粒度的选择对预测精度的影响**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **粒度选择** | **电力数据** | | | |  | **期货数据** | | | |  | **股指数据** | | | |
| **R2 ↑** | **CORR ↑** | **RMSE ↓** | **MAE ↓** |  | **R2 ↑** | **CORR ↑** | **RMSE ↓** | **MAE ↓** |  | **R2 ↑** | **CORR ↑** | **RMSE ↓** | **MAE ↓** |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (MgRL) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (MgRL) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (MgRL) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (MgRL) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (MgRL) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (MgRL) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# 第4章 总结与展望

总的来说，本研究揭示了探索多粒度特征模式对于提高时序预测精度的重要性，为解决当前多粒度时序预测所面临的两大困难，提出了一种多粒度学习框架。具体而言，为了克服多粒度数据之间的语义重叠，设计并实现了一种新颖的多粒度残差学习方法，以避免模型被冗余的粗粒度趋势信息所支配。同时，为了进一步加强多粒度残差学习神经网络的有效性，本研究还设计了自监督的置信度估计模块。在电力数据、股指数据和期货数据三种实际数据上所进行的实验证明了：本文所提出了基于多粒度残差学习和置信度估计的神经网络模型，能够有效地提高时间序列的预测能力。

在未来的学习和研究中，以下两个方向还值得深入探索：首先是多粒度数据生成方面，目前不同粒度的数据全部由固定算法生成，未来可以研究如何使多粒度数据的生成过程可学习，而不是直接人工指定。其次是特征建模方面，可以将强化学习、信息论等思想融入到建模过程中，让多粒度输入数据提供更多信息，进一步提高最终预测效果。

# 引用文献

1. Ankargren S, Unosson M, Yang Y. A flexible mixed-frequency vector autoregression with a steady-state prior. Journal of Time Series Econometrics. 2020 Aug 18;12(2):20180034.
2. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arxiv preprint arxiv:1409.0473. 2014 Sep 1.
3. Belghazi MI, Baratin A, Rajeshwar S, Ozair S, Bengio Y, Courville A, Hjelm D. Mutual information neural estimation. InInternational conference on machine learning 2018 Jul 3 (pp. 531-540). PMLR.
4. Binkowski M, Marti G, Donnat P. Autoregressive convolutional neural networks for asynchronous time series. InInternational Conference on Machine Learning 2018 Jul 3 (pp. 580-589). PMLR.
5. Box G. Box and Jenkins: time series analysis, forecasting and control. InA Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century 2013 Oct (pp. 161-215). London: Palgrave Macmillan UK.
6. Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. InBiocomputing 2000 1999 (pp. 418-429).
7. Cheng M, Yuan F, Liu Q, Xin X, Chen E. Learning transferable user representations with sequential behaviors via contrastive pre-training. In2021 IEEE International Conference on Data Mining (ICDM) 2021 Dec 7 (pp. 51-60). IEEE.
8. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arxiv preprint arxiv:1406.1078. 2014 Jun 3.
9. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arxiv preprint arxiv:1412.3555. 2014 Dec 11.
10. Cui Z, Chen W, Chen Y. Multi-scale convolutional neural networks for time series classification. arxiv preprint arxiv:1603.06995. 2016 Mar 22.
11. Deng S, Zhang N, Zhang W, Chen J, Pan JZ, Chen H. Knowledge-driven stock trend prediction and explanation via temporal convolutional network. InCompanion proceedings of the 2019 world wide web conference 2019 May 13 (pp. 678-685).
12. Feng F, Chen H, He X, Ding J, Sun M, Chua TS. Enhancing Stock Movement Prediction with Adversarial Training. In IJCAI 2019 Aug 10 (Vol. 19, pp. 5843-5849).
13. Foroni C, Marcellino MG. A survey of econometric methods for mixed-frequency data. Available at SSRN 2268912. 2013 Feb 6.
14. [Ghysels E. Macroeconomics and the reality of mixed frequency data. Journal of Econometrics. 2016 Aug 1;193(2):294-314.
15. [Ghysels E, Sinko A, Valkanov R. MIDAS regressions: Further results and new directions. Econometric reviews. 2007 Feb 5;26(1):53-90.
16. Giannone D, Reichlin L, Small D. Nowcasting: The real-time informational content of macroeconomic data. Journal of monetary economics. 2008 May 1;55(4):665-76.
17. Harvey AC. Forecasting, structural time series models and the Kalman filter (1990).
18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).
19. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14 2016 (pp. 630-645). Springer International Publishing.
20. Hillmer SC, Tiao GC. An ARIMA-model-based approach to seasonal adjustment. Journal of the American Statistical Association. 1982 Mar 1;77(377):63-70.
21. Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997 Nov 15;9(8):1735-80.
22. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics. 1970 Feb 1;12(1):55-67.
23. Hou M, Xu C, Liu Y, Liu W, Bian J, Wu L, Li Z, Chen E, Liu TY. Stock trend prediction with multi-granularity data: A contrastive learning approach with adaptive fusion. InProceedings of the 30th ACM International Conference on Information & Knowledge Management 2021 Oct 26 (pp. 700-709).
24. Huang S, Wang D, Wu X, Tang A. Dsanet: Dual self-attention network for multivariate time series forecasting. InProceedings of the 28th ACM international conference on information and knowledge management 2019 Nov 3 (pp. 2129-2132).
25. Huber F, Koop G, Onorante L, Pfarrhofer M, Schreiner J. Nowcasting in a pandemic using non-parametric mixed frequency VARs. Journal of Econometrics. 2023 Jan 1;232(1):52-69.
26. Hunter JS. The exponentially weighted moving average. Journal of quality technology. 1986 Oct 1;18(4):203-10.
27. Huynh-Thu VA, Geurts P. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. Scientific reports. 2018 Feb 21;8(1):3384.
28. Kingma DP, Ba J. Adam: A method for stochastic optimization. arxiv preprint arxiv:1412.6980. 2014 Dec 22.
29. Kinney JB, Atwal GS. Equitability, mutual information, and the maximal information coefficient. Proceedings of the National Academy of Sciences. 2014 Mar 4;111(9):3354-9.
30. Kitaev N, Kaiser Ł, Levskaya A. Reformer: The efficient transformer. arxiv preprint arxiv:2001.04451. 2020 Jan 13.
31. Koochali A, Schichtel P, Dengel A, Ahmed S. Probabilistic forecasting of sensory data with generative adversarial networks–forgan. IEEE Access. 2019 May 17;7:63868-80.
32. Kwak N, Choi CH. Input feature selection by mutual information based on Parzen window. IEEE transactions on pattern analysis and machine intelligence. 2002 Dec;24(12):1667-71.
33. Lai G, Chang WC, Yang Y, Liu H. Modeling long-and short-term temporal patterns with deep neural networks. InThe 41st international ACM SIGIR conference on research & development in information retrieval 2018 Jun 27 (pp. 95-104).
34. Li Z, Zhao H, Liu Q, Huang Z, Mei T, Chen E. Learning from history and present: Next-item recommendation via discriminatively exploiting user behaviors. InProceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining 2018 Jul 19 (pp. 1734-1743).
35. Lin J, Michailidis G. A multi-task encoder-dual-decoder framework for mixed frequency data prediction. International Journal of Forecasting. 2023 Sep 10.
36. Liu G, Mao Y, Sun Q, Huang H, Gao W, Li X, Shen J, Li R, Wang X. Multi-scale Two-way Deep Neural Network for Stock Trend Prediction. InIJCAI 2020 Jul (pp. 4555-4561).
37. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. IEEE transactions on Medical Imaging. 1997 Apr;16(2):187-98.
38. Mudelsee M. Climate time series analysis. Atmospheric and. 2010;397.
39. Oord AV, Li Y, Vinyals O. Representation learning with contrastive predictive coding. arxiv preprint arxiv:1807.03748. 2018 Jul 10.
40. Oreshkin BN, Carpov D, Chapados N, Bengio Y. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. arxiv preprint arxiv:1905.10437. 2019 May 24.
41. Ostle, Clare, et al. The rise in ocean plastics evidenced from a 60-year time series. Nature communications, 2019, 10.1: 1622.
42. Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on pattern analysis and machine intelligence. 2005 Jun 20;27(8):1226-38.
43. Qin Y, Song D, Chen H, Cheng W, Jiang G, Cottrell G. A dual-stage attention-based recurrent neural network for time series prediction. arxiv preprint arxiv:1704.02971. 2017 Apr 7.
44. Reis MS. Multiscale and multi-granularity process analytics: A review. Processes. 2019 Jan 24;7(2):61.
45. Roberts S, Osborne M, Ebden M, Reece S, Gibson N, Aigrain S. Gaussian processes for time-series modelling. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2013 Feb 13;371(1984):20110550.
46. Salinas D, Flunkert V, Gasthaus J, Januschowski T. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International journal of forecasting. 2020 Jul 1;36(3):1181-91.
47. Shih SY, Sun FK, Lee HY. Temporal pattern attention for multivariate time series forecasting. Machine Learning. 2019 Sep 15;108:1421-41.
48. Singh AK, Ibraheem SK, Muazzam M, Chaturvedi DK. An overview of electricity demand forecasting techniques. Network and complex systems. 2013;3(3):38-48.
49. Song H, Rajan D, Thiagarajan J, Spanias A. Attend and diagnose: Clinical time series analysis using attention models. In Proceedings of the AAAI conference on artificial intelligence 2018 Apr 29 (Vol. 32, No 1.)
50. Tang X, Yao H, Sun Y, Aggarwal C, Mitra P, Wang S. Joint modeling of local and global temporal dynamics for multivariate time series forecasting with missing values. In Proceedings of the AAAI Conference on Artificial Intelligence 2020 Apr 3 (Vol. 34, No. 04, pp. 5956-5963).
51. [Tishby N, Pereira FC, Bialek W. The information bottleneck method. arxiv preprint physics/0004057. 2000 Apr 24.
52. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Advances in neural information processing systems. 2017;30.
53. Vapnik V, Golowich S, Smola A. Support vector method for function approximation, regression estimation and signal processing. Advances in neural information processing systems. 1996;9.
54. Wu L, Li Z, Zhao H, Liu Q, Chen E. Estimating fund-raising performance for start-up projects from a market graph perspective. Pattern Recognition. 2022 Jan 1;121:108204.
55. Wu L, Li Z, Zhao H, Pan Z, Liu Q, Chen E. Estimating early fundraising performance of innovations via graph-based market environment model. InProceedings of the AAAI Conference on Artificial Intelligence 2020 Apr 3 (Vol. 34, No. 04, pp. 6396-6403).
56. [Wu Z, Pan S, Long G, Jiang J, Chang X, Zhang C. Connecting the dots: Multivariate time series forecasting with graph neural networks. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining 2020 Aug 23 (pp. 753-763).
57. Xu W, Liu W, Xu C, Bian J, Yin J, Liu TY. Rest: Relational event-driven stock trend forecasting. InProceedings of the web conference 2021 2021 Apr 19 (pp. 1-10).
58. [Yang, W., Tian, Z., & Hao, Y. (2022). A novel ensemble model based on artificial intelligence and mixed-frequency techniques for wind speed forecasting. Energy Conversion and Management, 252, 115086.
59. [Zhang L, Aggarwal C, Qi GJ. Stock price prediction via discovering multi-frequency trading patterns. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining 2017 Aug 13 (pp. 2141-2149).
60. Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI conference on artificial intelligence 2021 May 18 (Vol. 35, No. 12, pp. 11106-11115).
61. Ding Y, Liu W, Bian J, Zhang D, Liu TY. Investor-imitator: A framework for trading knowledge extraction. InProceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 2018 Jul 19 (pp. 1310-1319).

1. https://archive.ics.uci.edu/dataset/321/electricityloaddiagrams20112014 [↑](#footnote-ref-1)
2. https://github.com/microsoft/qlib [↑](#footnote-ref-2)
3. https://github.com/akfamily/akshare [↑](#footnote-ref-3)