



重慶大學  
CHONGQING UNIVERSITY

# 国家级（市级）大学生创新训练项目 申报书

项目名称：基于多尺度重编程的时序数据语义理解与预测  
方法研究

所在学院：本科生院

申 请 人：圣烨 陆韬宇 吕珊珊 田源棋

联系电话：15380212186

指导教师：高旻

本科生院 制

## 填写说明

1、本表填写内容必须与事实相符，逐项认真填写，表达准确严谨。空缺项要填“无”。

2、报告书填写内容应言简意赅，思路清晰，论证充分，字迹清楚，一律用计算机输入打印。

3、“项目开展支撑平台”指支撑本项目开展的国家级和省部级重点实验室（中心、平台等）、教学实验中心（实验室）、科研实验室等，表中填写有关实验室名称，可以多个。

4、需签字部分由相关人员以黑色钢笔或水笔签名。均用 A4 纸双面打印，于左侧装订成册。

5、格式要求：

（1）文中小标题为四号、仿宋、加黑；

（2）栏内正文为小四号、仿宋。

（3）栏内正文 1.5 倍行距。

项目名称	基于多尺度重编程的时序数据语义理解与预测方法研究		
项目起止时间	2024 年 11 月 — 2025 年 11 月		
项目所属学科	一级学科代码及名称	二级学科代码及名称	三级学科代码及名称
	520 计算机科学技术	52040 计算机软件	5204070 软件工程
项目开展支撑平台	本科生院工训中心 信息物理社会可信服务计算教育部重点实验室		
项目申请经费	10000 元（不超过 10000 元）		
项目来源 (可多选)	<input type="checkbox"/> 国家级和省部级重点实验室（中心、平台等）支持申报项目 <input checked="" type="checkbox"/> 教师科研项目转化 <input type="checkbox"/> 交叉学科创新项目 <input type="checkbox"/> 基于前期研究实践成果、继续深入研究实践的创新项目 <input checked="" type="checkbox"/> 教师发布课题 <input type="checkbox"/> 学生承担社会、企业委托项目等 <input type="checkbox"/> 学生自拟 <input type="checkbox"/> 其他_____		
负责人之前参与大学生科研训练（SRTP）项目情况	圣烨，负责人，CQU-SRTP-20230644，求解大规模线性方程组的分块Kaczmarz方法研究，2023，优秀		
项目成员之前参与大学生科研训练（SRTP）项目情况	陆韬宇，成员，CQU-SRTP-20230829，大学校园学生才艺技能社交平台的研发，2023，中等 吕珊珊，成员，CQU-SRTP-20223370，易腐品供应链库存-路径协同控制与优化，2022，良好 田源棋，成员，CQU-SRTP-20230767, 基于深度学习注意力检测的关键技术研究，2023，中等		

项目组成员（含项目申请学生）					
姓 名	学 号		学院	年级专业班	联系电话
圣烨	20233825		弘深学院	数学拔尖班	15380212186
吕珊珊	20221952		大数据与软件学院	软件工程	13814023667
陆韬宇	20236356		辛辛那提大学联合学院	电气工程及自动化	13350396440
田源棋	20231075		大数据与软件学院	软件工程	13792608261
主要指导教师					
姓 名	工号	职 称	学 院	研究方向	联系电话
高旻	31617	教授	大数据与软件学院	时序数据挖掘、个性化推荐系统	13608367500

一、项目内容概述

主要研 究内容 (限 200 字 内)	时序数据预测可以揭示数据随时间变化的趋势，为科学决策提供重要依据。当前广泛使用的预测模型难以利用数据之外的知识，往往泛化能力较差。近期，有研究初步探索了基于大语言模型的预测方法，借助 LLMs 的通用知识与推理能力进行预测。但是，这些工作难以在语义表示时捕捉短期波动和长期趋势呈现的复杂多尺度特征，为此，本项目研究基于多尺度重编程的时序数据语义理解方法，通过学习复杂多尺度语义表示，提升大语言模型的预测性能。
------------------------------	--

<p>项目研究 难点及创 新点</p>	<p><b>一、项目研究难点</b></p> <p>1. <b>时间窗口设计的适应性与精确性。</b>设计多尺度时序数据的重编程机制中，最大的挑战是如何确保时间窗口的划分能够充分捕捉到关键时序特征。合理的时间窗口设计需要能够精确地识别和反映短期的快速变动和长期的趋势变化。这涉及到如何根据具体任务的需求（如金融与气象等不同领域），选择窗口的长度和类型（如滑动窗口或固定窗口），以及如何设置窗口以适应时序数据中的非线性特性和复杂性。此外，自适应窗口策略的开发需要能够根据数据波动幅度或特定事件的发生频率动态调整窗口的大小，这在技术上需要高度的动态调整能力和精确的模型训练策略。</p> <p>2. <b>时序数据的数值与语义对齐。</b>如何将复杂的时序数值模式有效地转换为大型语言模型（LLM）能理解的语义形式是另一大难点。一个精确的数值到文本的映射机制不仅需要定义数值模式与具体语义描述之间的对应关系，还需要根据不同领域的特性来构建领域特定的符号化表达。例如，在金融领域中，某些市场波动可以映射为“牛市”或“熊市”，而相应波动在气象领域则可以映射为“降雨”、“晴天”等气象术语。此外，实现这一对齐还需要确保语义描述能够准确地反映出时序数据的上下文信息和变化趋势，这对于语义分析的深度和广度提出了更高的要求。</p> <p>3. <b>多尺度特征整合与预测模型优化。</b>构建一个适用于多领域时序数据预测的通用模型面临诸多难点，包括如何整合来自不同时间尺度的特征，以及如何设计模型以具备强大的泛化能力。特别是在数据量和复杂性持续增长的情况下，保持模型的适应性和扩展性是关键。这要求我们开发新的算法和架构，以在不牺牲预测性能的前提下，处理更广泛和更复杂的数据集。此外，还需要设计高效的数据处理流程和模型训练策略，以应对可能出现的大规模数据处理需求，确保模型能够在实际应用中达到高效和精准的预测效果。</p> <p><b>二、项目创新点</b></p> <p>1. <b>基于多尺度重编程的时序预测方法优化</b></p>
-----------------------------	--

	<p>结合多尺度重编程与时序数据的语义对齐，设计一套高效的时序预测方法。通过结合 LLM 的语义推理能力与时序特征，优化预测性能。例如，在长时间跨度的时序任务（如金融市场预测）中，实现高效、精准的预测。</p> <p><b>2. 多尺度时序数据的重编程机制设计</b></p> <p>开发一种多尺度重编程方法，将时序数据按不同的时间窗口（如短期、中期、长期）划分成多个片段，并为每个片段生成嵌入向量。</p> <p>这一机制将确保模型能够从不同时间层次捕捉特征，从而有效应对短期波动和长期趋势的预测任务。</p> <p><b>3. 时序数据语义对齐与嵌入表示</b></p> <p>在将时序数据切片后，设计一种语义对齐方法，将切片后的时序数据转化为 LLM 能理解的语义输入（即通过文本化描述或符号化表示进行转换）。这将允许 LLM 在处理时序数据时，能够结合其强大的语义理解能力来捕捉时序中的关键模式。</p>
--	--

二、项目组成员分工

姓名	主要承担项目研究内容
圣烨	负责设计和开发时序数据的多尺度重编程机制。实现时序数据的多层次切片，将其转化为适合 LLM 理解的语义片段。确保重编程后的数据片段在多尺度上保持一致性，并符合后续处理需求。
吕珊珊	设计捕捉短期波动和长期趋势的语义对齐机制。研究如何对齐不同尺度的语义片段，使 LLM 能够高效提取时序数据中的关键特征。调整和优化语义对齐算法，以保证对齐效果的准确性和鲁棒性。
陆韬宇	将重编程机制与语义对齐机制整合进时序数据预测系统。负责系统的整体架构设计和实现，包括数据流的处理和模型的集成。设定系统的部署环境，并确保在不同的应用领域中能够顺利适配和运行。
田源棋	负责在多个实际应用场景（如交通流量预测、水质预测等）中验证系统的有效性和鲁棒性。收集和分析预测系统在各场景下的性能数据，对结果进行评估。根据测试反馈，不断优化系统的预测精度和稳定性，并提供改进建议。

### 三、立项背景和依据（包括研究目的、意义、国内外研究现状分析及评价等）

#### 研究背景

在《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》中，对经济预测与市场监管以及风险管理与异常检测提出了明确要求，为基于多尺度重编程的时序数据语义理解与预测方法研究提供了坚实的政策背景。

**经济预测与市场监管：**在数字经济快速发展的背景下，市场行为的复杂性和快速变化对经济预测提出了更高要求。《纲要》中明确提出“加快数字化发展，建设数字中国”，强调了数字技术在提升经济运行效率和韧性中的重要作用。时间序列分析作为数字技术的重要组成部分，对于监控经济指标和市场趋势至关重要。通过精准的经济预测，可以为宏观经济政策的及时调整提供决策支持，确保市场稳定，促进国内大循环的畅通。

**风险管理与异常检测：**《纲要》中对“防范化解重大风险”提出了明确要求，特别是在金融市场和能源供应等领域。基于多尺度重编程的时间序列预测方法，能够对潜在风险进行早期识别和预警，为风险管理提供强有力的技术支持。这种方法有助于机构提前制定应对策略，减少金融市场波动和能源供应短缺等风险对经济和社会的负面影响，确保国家经济安全和社会稳定。

#### 研究目的

- 模型适应性：**研究如何将 LLM 的架构与时间序列数据的特性相结合，提高其在不同时间序列任务中的适应性。
- 预测精度提升：**通过优化模型参数和训练策略，提升 LLM 在时间序列预测中的准确性与鲁棒性。
- 多模态融合：**探索将 LLM 与其他模型（如 ARIMA、LSTM 等）进行融合，利用其语言理解能力来改进时间序列预测效果。
- 实用案例分析：**在金融、气象、交通等实际应用场景中验证研究成果，分析实验结果。

#### 研究意义

- 理论贡献：**通过对 LLM 在时间序列预测中的应用研究，丰富了时间序列分析的理论体系，推动了自然语言处理与时间序列分析的交叉研究。

2. 应用价值：提升时间序列预测的准确性和效率，有助于各行各业在决策制定中获得更可靠的预测支持，尤其是在金融市场、气象预报和供应链管理等领域。
3. 创新实践：为后续研究提供新的思路和方法，推动更智能化的预测模型发展，促进数据科学与人工智能领域的进一步创新。

## 国内外研究现状及评价

在时间序列预测领域，传统方法主要集中在统计模型和机器学习算法上，如 ARIMA、LSTM 等。然而，随着深度学习技术的飞速发展，时序预测的研究逐渐向更加复杂和多样化的模型演进。近年来，长短期记忆网络（LSTM）和卷积神经网络（CNN）在处理时序数据时取得了一定的成功，同时新的方法也在不断涌现。基于 Transformer 的模型，尤其是自注意力机制（Self-attention）在捕捉长期依赖关系方面表现出了强大的优势，如 Transformers for Time Series 等。除此之外，线性模型的结合，如线性变换模型（Linear Transformer）和自回归模型（AR-based Models），也在一定程度上简化了复杂度，同时保留了较强的预测能力，成为当前时序预测研究中的重要方向。这些新兴方法不仅提高了预测的准确性，也推动了时序预测领域从传统统计方法到深度学习模型转变。

国内外研究逐步认识到自然语言处理技术在时间序列预测中的潜力，特别是大型语言模型的出现为这一领域带来了新的视角。研究者们开始尝试将 LLM 应用于时间序列数据，以实现更高维度的数据理解和信息提取。

尽管已有部分研究探讨了 LLM 在时间序列预测中的应用，但 LLM 在时序预测上对于多尺度的划分和重编程仍处于探索阶段。因此，本课题的研究不仅能够填补这一领域的空白，还将推动相关技术的进一步发展和应用。



## 四、主要研究内容和目标（包括研究方案和技术路线）

### 研究方案：

#### 1. 多尺度时序数据的重编程机制设计：

- **内容：**开发一种多尺度重编程方法，将时序数据按不同的时间窗口（如短期、中期、长期）划分成多个片段，并为每个片段生成嵌入向量。这一机制将确保模型能够从不同时间层次捕捉特征，从而有效应对短期波动和长期趋势的预测任务。
- **关键问题：**如何设计适当的时间窗口，确保不同尺度的时序特征能够被重编程并嵌入到 LLM 的输入中。

#### 2. 时序数据语义对齐与嵌入表示：

- **内容：**在将时序数据切片后，设计一种语义对齐方法，将切片后的时序数据转化为 LLM 能理解的语义输入（即通过文本化描述或符号化表示进行转换）。这将允许 LLM 在处理时序数据时，能够结合其强大的语义理解能力来捕捉时序中的关键模式。
- **关键问题：**如何将时序数据的数值模式与文本模式对齐，使得 LLM 能够基于对齐后的数据生成准确的预测结果。

#### 3. 基于多尺度重编程的时序预测方法优化：

- **内容：**结合多尺度重编程与时序数据的语义对齐，设计一套高效的时序预测方法。通过结合 LLM 的语义推理能力与时序特征，优化预测性能，尤其在长时间跨度的时序任务（如金融市场预测）中，实现高效、精准的预测。
- **关键问题：**如何充分利用多尺度特征和语义对齐的信息，构建一个适用于多领域时序数据预测的通用模型。

### 技术路线：

#### 1. 多尺度时序数据的重编程机制设计：

##### 1) 时序数据的多尺度切片：

- **时间窗口划分：**首先根据任务需求，设计多种时间窗口，将时序数据按不同的时间跨度切分为多个片段。短期窗口可以捕捉快速波动（例如 1

分钟、5 分钟、1 小时），长期窗口则用于捕捉趋势或周期性变化（例如每天、每周、每月）。这个切片过程可以通过滑动窗口（sliding window）或固定窗口（fixed window）方法实现，确保不同时间尺度上的特征都能够被处理。

- **自适应窗口策略：**针对某些场景中的非线性时序数据，可以设计自适应窗口策略，即根据时序数据的波动幅度或事件发生频率自动调整窗口长度，确保模型在波动剧烈的时期能获得更精细的刻画，而在相对平稳的时期则使用较大时间尺度来节约计算资源。

## 2) 数据嵌入与多尺度融合：

- **片段嵌入：**对每个时间片段使用嵌入技术生成嵌入向量，具体可以通过线性变换、卷积神经网络（CNN）或图神经网络（GCN）等嵌入方式。嵌入的目标是将每个片段的时序特征映射到高维空间，保留其时间上的依赖性和模式信息。
- **多尺度特征融合：**通过拼接或加权平均的方式，将不同时间尺度上的嵌入向量融合到一起，形成统一的输入表示。对于加权融合，可以使用注意力机制（Attention），根据任务需求动态分配不同时间尺度的重要性，从而让模型聚焦于关键的时间片段。

## 3) 片段上下文重编程：

- **全局上下文建模：**在多尺度片段嵌入的基础上，利用 LLM 的自注意力机制，将不同时间尺度的片段进行上下文重编程，捕捉不同片段之间的时间依赖性。例如，利用 Transformer 的多头注意力（multi-head attention），实现对短期和长期信息的同时建模，使得模型能够在不同时间片段之间建立关联，并充分利用全局时间上下文进行时序预测。

## 2. 时序数据语义对齐与嵌入表示：

### 1) 时序数据到语义嵌入的转换：

- **数值到语义映射：**将时序数据的数值信息映射到语义空间。通过预定义的词汇表或符号化规则，将特定的数值模式（如异常点、突发变化、周期性变化）转化为语义描述。例如，温度数据中的剧烈变化可以映射为“升温”或“降温”等词汇。

- **领域特定符号表构建**: 根据应用领域（如金融、气象、交通等），设计一个包含领域特定术语的符号表。例如，在金融领域中，某些市场波动可以映射为“牛市”或“熊市”，在气象领域则可以映射为“降雨”、“晴天”等气象术语。此符号表将为时序数据生成语义描述提供依据。

## 2) 语义对齐与重编程:

- **文本与时序数据对齐**: 设计一种对齐机制，将时序片段的嵌入与其对应的语义描述对齐，使得 LLM 能够将时序数据视为类似于文本数据来处理。具体来说，使用语义相似度算法（如余弦相似度）将数值嵌入和文本语义嵌入进行匹配，确保数值变化能够合理地映射为自然语言中的语义表达。
- **上下文语义描述生成**: 通过上下文分析，生成更复杂的语义描述。例如，对于一段气温数据，可以生成“过去三天的气温逐渐上升，预计未来一天将持续升高”的描述，这些描述可以直接作为 LLM 的输入。上下文描述可以通过规则设计或预训练模型生成。

## 3) 对齐后的输入重编码:

- **统一表示结构**: 将时序数据片段和语义描述融合成统一的表示结构输入 LLM。这可以通过将语义描述和嵌入拼接在一起，或者将语义描述视为一种提示（prompt）输入，指导 LLM 对时序数据进行更高层次的理解和预测。
- **跨模态对齐**: 引入跨模态嵌入技术（如跨模态 Transformer），确保时序数据的数值特征和文本描述能够共同作用于模型中，实现对时序数据的语义理解与预测。

## 3. 基于多尺度重编程的时序预测方法优化:

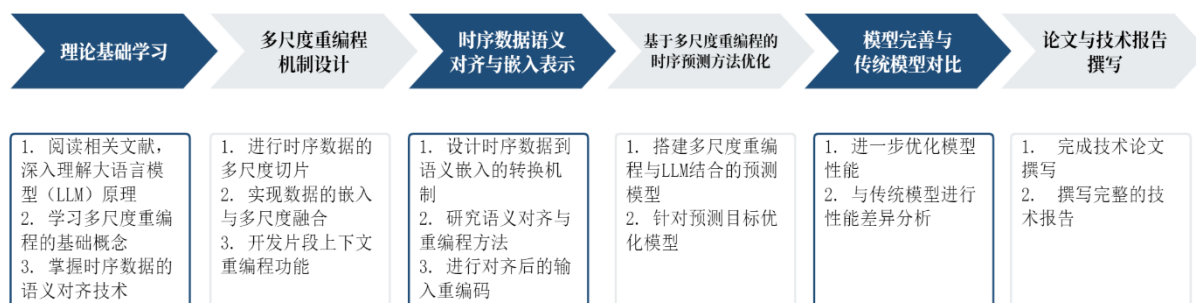
### 1) 多尺度重编程与 LLM 结合的预测模型:

- **LLM 的重编程接口设计**: 通过预训练大语言模型（如 GPT-2、BERT），接入多尺度重编程后的时序数据，确保 LLM 可以处理这些非语言数据。设计重编程接口时，需要考虑如何在不大规模修改 LLM 结构的前提下，使其能够有效解析时序数据的特征。具体可以通过输入转化（input transformation）模块，将时序数据预处理为 LLM 可以识别的格式。

- **引入 Prompt 工程：**使用 Prompt 工程对 LLM 进行任务提示，帮助 LLM 理解时序数据中的特定模式。例如，可以使用“分析过去 7 天的趋势，预测接下来一天的变化”这样的提示，引导 LLM 在预测任务中聚焦于特定时间段的关键特征。

## 2) 预测目标的优化：

- **损失函数优化：**设计适合多尺度时序预测的损失函数，平衡短期与长期预测误差。可以使用加权均方误差 (Weighted MSE) 或分段损失函数 (Piecewise Loss)，确保模型在不同时间尺度上的表现都有优化。



注：如为重庆大学大学生科研训练计划 (SRTP) 结项项目，需写明在原项目的基础上进一步深入研究的内容。

## 五、研究基础（对项目的参与动机、已有知识储备、相关研究和训练基础）

### 1. 项目基础

多尺度重编程技术提供了一种新的思路，通过将时序数据的不同时间尺度划分并重编程为 LLM 能够理解的语义信息，能够让模型捕捉更复杂的时序模式，提升对时序数据的语义理解能力。这种方法不仅可以提高模型在时序数据上的预测能力，还能扩展 LLM 在非语言数据上的应用范围，实现更加准确和高效的时序预测。因此，研究基于多尺度重编程的时序数据语义理解与预测方法，具有重要的理论和实际意义。

小组成员均已阅读了论文 TimeLLM，了解了 TimeLLM 的基本框架，学习 TimeLLM 中的重编程技术，进一步了解了我们研究所要使用的多尺度重编程技术。阅读了一些 TimeLLM 的后续论文，探索 LLMs 在时序预测领域应用的可能性。

### 2. 项目组成员

圣烨，陆韬宇，吕珊珊和田源棋四位成员分别来自于弘深学院数学拔尖班，辛辛那提学院电气工程及其自动化专业，大数据与软件学院软件工程专业和大数据与软件学院软件工程专业。均具有高等代数、数学分析、离散数学、和编程基础；专业知识扎实，在程序设计开发、数据库设计、计算机网络、人工智能算法方面有知识储备；且对时间序列这一研究课题有过了解，且系统地接受过大学生科研训练，并且均成功结题，获得了不错的成绩。

四位组员的学习成绩优秀，综合绩点均名列前茅，在保证学业成绩的情况下有充足的时间和精力投入项目研究，有求知、探索和创新意识。

圣烨，弘深学院数学拔尖班，GPA 3.826，专业排名：1

陆韬宇，辛辛那提学院电气工程及其自动化专业，GPA 3.6，专业排名：13

吕珊珊，大数据与软件学院软件工程专业，GPA 3.72，专业排名：13

田源棋，大数据与软件学院软件工程专业，GPA 2.86，专业排名：56

## 六、研究计划和进度（就文献查询、社会调查、方案设计、实验研究、数据处理、研制开发、撰写论文或研究报告、结题和答辩、成果推广、论文发表、专利申请等工作逐项计划时间，时间节点精确到月份）

2024.11-2024.12

阅读相关文献，学习 LLM 大语言模型以及多尺度重编程与时序数据的语义对齐技术。

2024.12-2025.02

多尺度时序数据的重编程机制设计，包括进行时序数据的多尺度切片、数据嵌入与多尺度融合、片段上下文重编程。

2025.02-2025.04

时序数据语义对齐与嵌入表示，包括时序数据到语义嵌入的转换、语义对齐与重编程及对齐后的输入重编码。

2025.06-2025.08

基于多尺度重编程的时序预测方法优化，搭建多尺度重编程与 LLM 结合的预测模型并且进行预测目标的优化。

2025.09-2025.10

完善模型并且进行进一步的优化，对比与传统模型的差异。

2025.10-2025.11

撰写论文以及完整的技术报告。

## 七、项目研究支撑条件（项目所依托的实验室、课题组等各类单位能提供的直接支持项目开展的软硬件设施和其他校内外资源）

实验室：本科生院工训中心

硬件设施：2 块 4090，2 块 3090GPU，还可线上租用 GPU。

项目成员：

圣烨，陆韬宇，吕珊珊和田源棋四位成员分别来自于弘深学院数学拔尖班，辛辛那提学院电气工程及其自动化专业，大数据与软件学院软件工程专业和大数据与软件学院软件工程专业。均具有高等代数、数学分析、离散数学、和编程基础；专业知识扎实，在程序设计开发、数据库设计、计算机网络、人工智能算法方面有知识储

备；且对时间序列这一研究课题有过了解，且系统地接受过大学生科研训练，并且均成功结题，获得了不错的成绩。

四位组员的学习成绩优秀，综合绩点均名列前茅，在保证学业成绩的情况下有充足的时间和精力投入项目研究，有求知、探索和创新意识。

圣烨，弘深学院数学拔尖班，GPA 3.826，专业排名：1

陆韬宇，辛辛那提学院电气工程及其自动化专业，GPA 3.6，专业排名：13

吕珊珊，大数据与软件学院软件工程专业，GPA 3.72，专业排名：13

田源棋，大数据与软件学院软件工程专业，GPA 2.86，专业排名：56

## 八、预期提供的成果及形式

1. 开发一种有效的多尺度重编程机制，将时序数据切片为适合 LLM 理解的多层次语义片段。

设计一种能够捕捉短期波动与长期趋势的语义对齐机制，使得 LLM 在处理复杂时序数据时能够有效提取其中的关键特征。

实现基于该重编程机制的时序数据预测系统，并在多个应用领域（如交通流量预测、水质预测等）中验证其有效性和鲁棒性。

2. 通过项目的实施获取一项知识产权（包括但不限于）：

（1） 在国内外期刊/学术会议发表 CCF-A 或 CCF-B 论文 1-2 篇

（2） 申请发明专利 0-1 项

3. 提交完整的技术报告。

## 九、项目经费概算

资料文印费：3600 元；用于购买项目实施所需的图书资料；文印、出版等费用；

交通费：2000 元；用于学生在市内或校区之间的交通费；

材料费：4400 元；用于租用项目实施所需的云 gpu、cpu；

总计 10000 元

## 十、评审、审批意见

指导教师意见:

指导教师签字:

年 月 日

学院推荐意见:

主管院长签字:

(公 章)

年 月 日

学校意见:

(公 章)

年 月 日