**钱塘江流域水环境质量预测预警方案**

**一、基本原理**

许多现实世界的应用需要长序列时间序列的预测，如电力消耗计划。长序列时间序列预测(LSTF)要求模型具有较高的预测能力，即能够有效捕捉输出和输入之间精确的长期相关性耦合。最近的研究表明，潜在的transformer增加了预测容量。然而，有几个问题与防止直接应用程序崩溃的transformer有关，包括四个方面的复杂性、高内存使用量和对代码无关 架构的一致模拟。为了解决这些问题，我们使用了一个基于高效 transformer 的模型，名为Informer，具有显著的特点:

1. 一个自关注机制，它在时间复杂度和内存使用方面达到了，并且在序列的依赖对齐方面具有相当的性能。
2. 自我注意力提取通过将层叠层输入减半来突出主要注意力，并有效地处理长期序列。
3. 生成式解码器虽然概念简单，但它可以预测长时间序列的运行，而不是逐步进行，这极大地提高了长时间序列预测的推理速度。在我们的大规模数据集上进行了广泛的实验，验证了模型的有效性，并为问题提供了新的解决方案。

**二、方案计划**

**2.1 研究目的**

本研究拟通过研究水环境质量预警技术，建立一套全面、科学的水环境质量预警体系，通过及时有效地监测水质情况，预警潜在风险，引导决策部门采取有效的措施，实现保护钱塘江流域水资源，改善水环境的目标。

**2.2 研究意义**

钱塘江作为中国东南沿海的重要水系，其下游流域不仅是杭州经济发展的重要引擎，也是人口高度密集的生活区域。近年来，随着工业化和城市化进程的加速推进，该流域面临日益严峻的水污染问题。水质的恶化不仅对生态系统的平衡构成威胁，也对居住于此的公众健康带来了潜在风险。此外，国家信息化的快速发展对智慧水利提出了新的要求，“十四五”规划和 2035 年远景目标规划中明确指出，构建智慧水利体系、以流域为单位提升水情测报和智能调度能力是重点建设任务。这不仅体现了国家对水利信息化、智能化的高度重视，也为水环境质量的监测和管理指明了新的发展方向。然而，现有水质监测技术和管理手段已难以满足日益增长的监测需求和精度要求。传统监测方法受限于时效性、精确度以及智能化水平，难以实现对水环境的连续、全面监控，对水污染事件的预测和预警能力有限。因此，开展钱塘江流域水环境质量预测预警关键技术研究具有重要的现实意义。

**2.3研究目标**

本项目研究有以下几个方面的研究目标:

**2.3.1建立钱塘江下游流域水环境质量预警的指标体系**

针对钱塘江下游水环境监测网络的挑战，应用多元统计分析方法，通过聚类和因子分析探索流域监测断面水环境质量相似性和主要污染因子，辅助建立水环境质量评价体系，解决现有评价体系存在的问题，为钱塘江水质评价提供理论基础。

**2.3.2提出钱塘江水环境预测模型，实现水环境质量的中长期预测**

现有神经网络模型难以实现中长期时间序列预测，导致水环境质量预测精度下降。针对传统网络的不足，本研究拟基于Informer 网络模型进行改进建立钱塘江下游流域水质预测模型，提高中长期水环境质量预测的精度和准确性。

**2.3.3构建钱塘江预测预警的监测平台，为水质监管提供智能决策**

研发钱塘江下游流域预警平台达到实现:1)实时监测水环境质量:通过实时获取水文监测数据，并通过钱塘江水环境预测模型对水文数据进行预测，提升对水环境变化的实时感知能力;2)预警和快速响应:当水环境质量参数超出正常范围时系统会发出报警提示，以方便城市环保管理部门可以快速响应;3)数据分析和智能决策:通过预警监测平台提供对数据进行实时分析和可视化展示，同时在水利发生预警时，为环保管理部门提供智能决策。

**2.4基于 Informer 模型的水环境质量预警方法**

Informer 模型在水环境质量预警工作中，传统的预警方法很难捕捉到时间跨度很长的关键信息，会导致预警系统灵敏度低，影响突发污染时间的应急处理。针对现有技术的不足，本研究拟设计一种基于Informer模型的水环境质量预警方法，如图1所示。基于Informer模型的水环境质量预警方法具体包括如下：

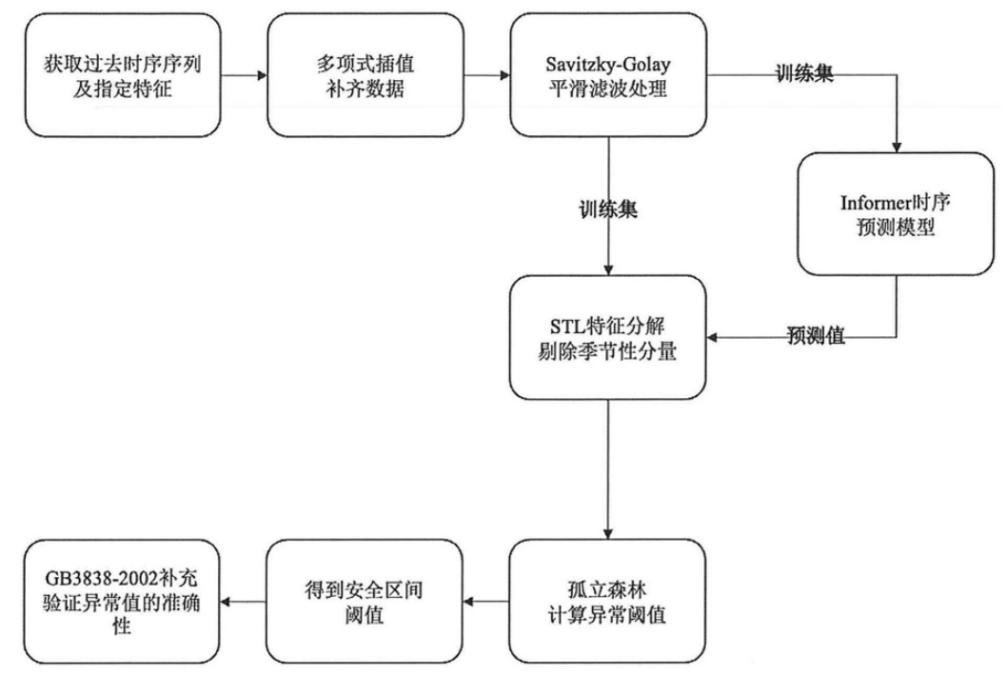


图1 Informer 水环境质量预测建模过程图

1）获取过去一段时间水环境质量指标的时间序列数据并选择一维水环境质量特征，组成时间戳和指定水环境质量特征构成的二维时间序列数据，对二维数据中的缺失值采用多项式插值的方式补齐；

2）在步骤 1 的基础上，对水环境质量特征数据进行 Savitzky-Golay（SG）滤波平滑降噪；

3）在步骤 2 的基础上，将 SG 处理之后的数据输入 Informer 模型中，输出水环境质量指标的多步预测值，获得未来的水环境质量指标预测值；

4）在步骤 2 的基础上，将 SG 处理之后的数据进行 STL（Seasonal and Trenddecomposition using Loess，STL）特征分解，将其分解为趋势分量、季节分量和残余分量三部分，并剔除季节分量；结合时间戳作为孤立森林模型的输入，isoforest(孤立森林)来对分解后的多维特征数据进行建模，将 isoforest 非异常点的中心偏离程度作为安全区间的阈值设定标准；

5）在步骤 3 的基础上，得到的未来水环境质量指标值，将其进行 STL 特征分解并剔除季节性分量，与 isoforest 的安全阈值进行预警；

6）在步骤 5 的基础上，得到的异常值再结合地表水环境质量标准 gb3838-2002计算水环境质量等级，根据水环境质量等级解释其异常的合理性。

**三、实验结果**

经过我们团队对钱塘部分流域的数据集进行模型分析得出如图2，图3所示，使用多变量（包括水温、PH值、溶解氧含量）来预测单变量的水体质量。其中两幅图在大模型的预测下拟合效果还是比较符合预期的。

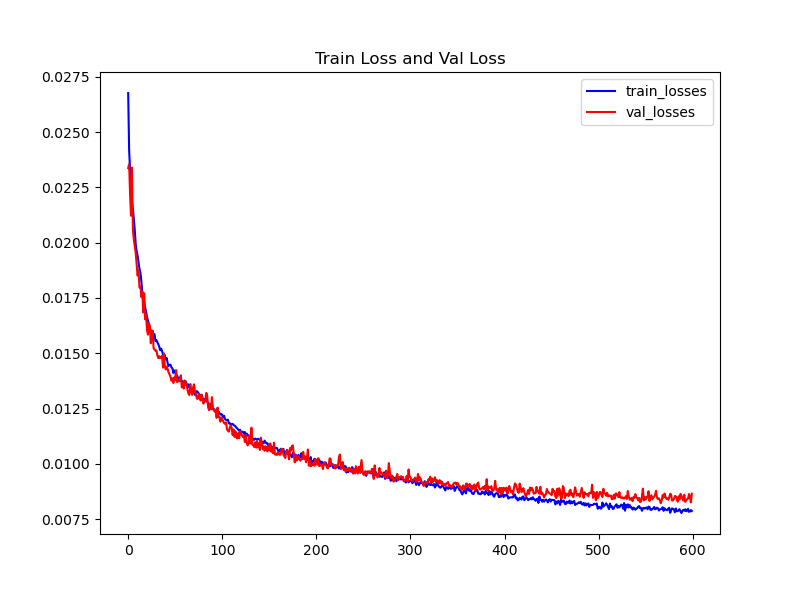


图2 预测与训练的loss值

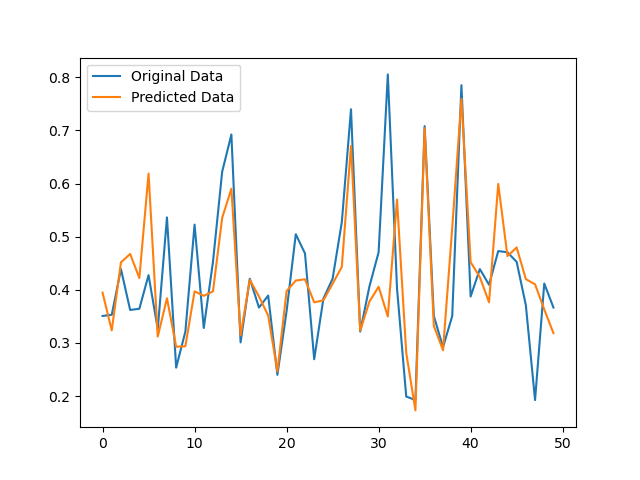


图3 预测与实际的数据值