**流感病毒传播与预测模型研究**

**摘 要**

本论文旨在设计并优化一个能够有效模拟、预测流感病毒传播的{数学模型}。流行性感冒（流感）是一种由上呼吸道和肺部的流感病毒引起的传染性呼吸道疾病，主要分为甲型和乙型两种类型。我们通过分析2014年至今的流感数据，着重探讨了甲型和乙型流感的感染人数的时间序列变化。

针对问题一，我们使用SEIR模型对流感传播进行了初步建模，并通过历史数据对模型进行验证，确保其有效性。

针对问题二，我们进一步优化了模型，考虑了季节性因素和不同人群的易感性。对于缺失的感染数据，我们利用插值方法和机器学习算法进行了填补，并对未来几年的流感感染趋势进行了长短期预测。

针对问题三，我们评估了不同类型流感的风险，利用风险评估矩阵等多种图表识别出高风险流感类型和高风险人群。我们分析了不同年龄段人群对不同流感类型的易感性和风险等级，从而为公共卫生政策提供参考依据。

针对问题四，我们为当地报刊撰写了一篇简短文章，建议了一些相关卫生部门可以采取的预防措施及其科学依据。

**关键词**：流感传播模型、SIR模型、时间序列分析、风险评估、数据插值

**一、问题分析**

**1.1问题背景**

流行性感冒（流感）是一种由上呼吸道和肺部的流感病毒引起的传染性呼吸道疾病。流感病毒主要分为A型和B型，甲型和乙型流感病毒是每年季节性流感流行的主要原因。流感传播速度快，传染性强，每年都会对全球公共卫生带来重大挑战。

流感病毒的传播不仅影响个人健康，还对社会经济造成广泛影响。尤其是在特定高风险人群中，如65岁及以上的老人、幼儿以及患有慢性病等特定健康状况的人，流感并发症的风险更高，甚至可能导致死亡。因此，研究流感病毒的传播规律，对流感进行预测和预防，对于保护公共健康具有重要意义。

自2014年以来，WHO和CDC对流感的监测数据积累了大量的信息，这为流感传播模型的建立和验证提供了宝贵的基础数据。通过建立适合的流感传播模型，我们可以分析甲型和乙型流感的感染人数随时间的变化，验证模型的有效性，并在此基础上进行优化，以实现对未来流感趋势的准确预测。

在建立和优化模型的过程中，需要综合考虑多种因素，包括流感病毒的生物学特性、季节的气候条件以及人群免疫状况等。建立一个能够真实预测和提供可靠信息的模型面临诸多挑战，如数据的准确性和完整性、模型参数的选择与调整、不同流感类型和人群风险的评估等。

此外，通过对不同类型流感及不同年龄段人群的风险评估，可以识别高风险流感类型和高危人群，从而为公共卫生决策提供科学依据。利用所建立的模型，不仅可以对缺失的数据进行合理预测，还可以对未来几年的流感感染趋势进行前瞻性预测，为流感防控措施的制定提供支持。

**1.1问题提出**

该题目要求我们根据实际情况和近十年的流感数据信息，通过建立恰当的数学模型来研究流感的感染模型及对未来的预测。

问题一需要我们建立一个流感传播模型，并根据这个模型所预测出的数据与实际数据进行拟合，以此来判断模型的可行性。

问题二需要我们使用一些方法来优化初期模型并根据优化后的模型对缺失值进行填充以及预测未来一段时间的流感传播情况。

问题三需要我们自行获取数据并据此来分析不同年龄段的感染风险与不同流感病毒的危害程度。

问题四需要我们写一篇短文，根据我们所得到的结果来提出一些预防流感方面的建议。

**二、问题分析**

**2.1 问题一分析**

问题一要求我们根据2014年至今的流感数据，首先需要建立一个适合的流感传播模型。可以选择SIR模型或SEIR模型这两种经典的流行病模型来开始我们的建模尝试。建模的第一步是通过这些模型对甲型和乙型流感的感染人数进行时间序列分析。这可以通过历史数据的拟合来检验模型的有效性。验证模型的有效性可以通过观察预测模型与实际数据的拟合程度来进行。

**2.2 问题二分析**

问题二需要在验证模型有效性的基础上，对模型进行进一步优化，对此可以通过机器学习来优化模型参数，使得模型能够最准确地拟合实际数据。还需要考虑更多的实际因素，如季节性变化和不同人群的易感性。可以引入时间序列分析和ARIMA模型来捕捉这些时变特性。我们对于缺失的感染数据，可以采用插值方法和机器学习算法进行数据填补。在此基础上，利用已有的流感传播模型对未来几年的流感感染情况进行预测。建立一个真实可靠的预测模型的难点主要在于数据的准确性和模型参数的精确估计，以及如何有效地整合多种因素（如季节性、不同人群的差异等）。

**2.3 问题三分析**

问题三在评估不同类型流感的风险及不同人群感染风险时，需要基于所爬取的数据并结合风险评估矩阵来识别高风险流感类型和重点人群。需要深入分析不同年龄段人群对不同流感类型的易感性和风险等级，这样可以为公共卫生政策提供指导。通过对流感传播模型的分析，确定哪些人群和流感类型的风险更高，从而有针对性地采取防控措施。

**2.4 问题四分析**

问题五需要我们根据模型预测和风险评估结果，为当地报刊撰写一篇简短文章，建议相关卫生部门可以采取的预防措施。这包括高风险人群的疫苗接种、定期卫生监测、公共场所的消毒措施等。所有建议应基于模型的科学依据和数据分析结果，以确保其有效性和合理性。

综上所述，对于流感传播与预测模型的研究，我们的五个问题之间具有紧密的层层递进关系。在解决这五个问题时，我们的总体思路是通过对流感传播模型的不断优化和扩展，逐步提高模型的精确度和实用性。因此，解决各个问题的关键点如下：问题一的关键在于：通过SIR模型和SEIR模型的建立与验证，并确保模型的有效性；问题二的重点是：在验证模型有效性的基础上，考虑更多的实际因素，从而建立一个真实可靠的预测模型，这一过程需要解决数据的准确性和模型参数的精确估计问题；利用已有的流感传播模型对未来几年的流感感染情况进行预测，以确保预测的准确性和可靠性；问题三的关键在于：利用已建立的流感传播模型，并结合风险评估矩阵，深入分析不同年龄段人群对不同流感类型的易感性和风险等级，以指导公共卫生政策的制定。

**三、模型基本假设与合理性说明**

为了对模型进行合理的优化，我们建立了以下的模型假设：

**假设1：**甲型和乙型流感病毒的传播可以用SEIR模型进行描述。

**假设2：**总人口数（N）在预测期间保持不变，即没有显著的出生、死亡或移民情况。

**假设3：**每个感染者在恢复或死亡后会获得持久的免疫力，不再重新感染。

**假设4：**流感病毒传播的接触速率（β）、和恢复速率（γ）等在整个研究期间保持不变。

**假设5：**流感感染数据是准确和完整的，缺失数据可以通过线性插值或季节性分解等方法进行合理填充。

**假设6：**模型优化过程中，所有的残差均为独立同分布的正态分布噪声。

**假设7：**未来的流感传播趋势可以通过SEIR模型和机器学习模型（如XGBoost）进行有效预测和模拟。

**四、符号说明**

|  |  |
| --- | --- |
| 符号 |  |
|  | 感染率，表示每个感染者每天接触易感者的比例 |
|  | 暴露期转化率，即暴露者转化为感染者的速率 |
|  | 恢复率，表示感染者恢复为免疫者的速率 |
|  | 初始假设的总人口数 |
|  | 流感阳性样本观测数据 |
|  | 趋势成分 |
|  | 季节性成分 |
|  | 残差数据 |
|  | 通过SEIR模型拟合得到的趋势成分 |
|  | 季节性分解得到的季节性成分 |
|  | 通过XGBoost模型预测的残差成分 |

**五、基于SEIR的流感传染病模型的建立——问题一**

**5.1数据概述**

在本研究中，我们使用了WHO从2014年3月至2024年6月甲流感染数据作为建模的基础，来分析感染人数随时间的变化。该数据集涵盖了甲流在一定时间范围内的感染情况，包括每个时间点的各种类型的感染人数。准确的数据记录是我们建模分析的基础，它使我们能够追踪甲流的传播趋势并评估干预措施的效果。

**5.2建模方法**

为了分析甲流感染人数随时间的变化，我们选择了SIER模型作为建模工具。SIER模型是一种流行病学模型，将人群划分为四个状态：易感者（S）、暴露者（E）、感染者（I）和恢复者（R）。该模型特别适用于描述疾病传播的动态过程，并可以通过以下几个步骤进行建模：

**定义状态转移**：模型描述了从易感者到暴露者，再到感染者，最后到恢复者的状态转移。这些转移过程可以用常微分方程（ODE）表示。

**时间序列分析**:我们将SIER模型应用于甲流感染人数的时间序列数据。通过将实际感染人数与模型输出进行比较，我们可以估计模型参数并调整模型以更好地拟合数据。

**参数估计与拟合**:通过拟合时间序列数据，我们能够估计模型中的关键参数，例如传播率、暴露期和感染期。这些参数帮助我们理解甲流的传播动态，并进行预测。

**动态分析与预测**:利用SIER模型，我们可以分析感染人数随时间的变化趋势，评估不同干预策略的潜在效果，并预测未来的感染趋势。

**5.3 问题的建模**

在进行SIER模型建模之前，我们对甲流感染数据进行了详细的预处理，以确保数据的准确性和可靠性。预处理过程包括以下几个主要步骤：

5.3.1数据清洗

对数据中的缺失值进行了处理。具体来说，我们选择用0填充缺失值。这一做法能够保证数据集在后续分析过程中不因缺失值而产生空缺，提高数据的完整性。

5.3.2数据类型转换

将日期列转换为标准的日期时间格式，以确保时间序列数据的准确性。对于其他数值列，我们将其转换为数值类型，处理过程中将任何非数值数据转化为NaN，并进一步用0填充这些值。这一步骤是为了确保所有的数据都以一致的格式进行处理。

5.3.3数据平滑

为了减少数据中的随机波动和噪声，我们对感染人数数据应用了移动平均平滑处理。具体地，我们使用了3点移动平均的方法，对每个数值列进行平滑处理，以获得更平稳的趋势数据。这有助于更准确地捕捉感染人数的变化趋势。

5.3.4特征选择

在特征选择阶段，我们确定了数据集中对建模重要的特征。选择了日期列以及所有数值列，以构建一个简洁而富有信息的数据子集。通过排除不必要的列，我们能够减少模型的复杂性，并提高分析效率。

5.3.5异常值处理

对数据中的异常值进行了识别和处理。通过计算四分位数范围（IQR），我们检测出异常值，并使用数据列的中位数替代这些异常值。此举旨在减少异常值对模型的干扰，从而提升数据的质量和模型的稳定性。

**5.4 流感传染病模型建立**

5.4.1数学模型

SIER 模型通过以下一组常微分方程 (ODEs) 描述流行病的传播过程：

易感者 (S) 的变化率由感染者的传播决定：

𝑑𝑆/𝑑𝑡=−𝛽𝑆𝐼/𝑁

暴露者 (E) 的变化率由易感者感染和暴露者转化为感染者的速率决定：

𝑑𝐸/𝑑𝑡=𝛽𝑆𝐼/𝑁−𝜎𝐸

感染者 (I) 的变化率由暴露者转化为感染者的速率和感染者恢复的速率决定：

𝑑𝐼/𝑑𝑡=𝜎𝐸−𝛾𝐼

恢复者 (R) 的变化率由感染者恢复的速率决定：

𝑑𝑅/𝑑𝑡=𝛾𝐼

5.4.2 参数估计

在拟合 SIER 模型的过程中，我们需要详细说明如何通过实际数据来估计模型中的参数。具体包括定义目标函数、优化算法以及如何利用这些方法来求解参数。以下是详细的步骤和实现。

为了使模型能够适用于实际数据，我们需要对参数 𝛽、𝜎和 𝛾进行估计。这可以通过以下步骤完成：

5.4.2.1 定义目标函数

我们使用均方误差（MSE）作为目标函数，用于衡量模型预测值与实际观测值之间的差异。

模型预测：在给定参数的情况下，使用 SIER 模型的常微分方程来模拟感染者人数𝐼随时间的变化。

5.4.2.2优化参数

使用优化算法来最小化目标函数，从而估计模型的最优参数值。我们可以使用 scipy.optimize 中的 curve\_fit 函数来实现这一目标。具体步骤如下：

定义拟合函数：我们需要定义一个函数，该函数接收参数并返回模型的预测结果。这个函数会被 curve\_fit 用来计算模型的拟合效果。

执行拟合：使用 curve\_fit 函数来优化参数，使得模型的预测结果与实际数据的误差最小化。

5.4.3 计算结果

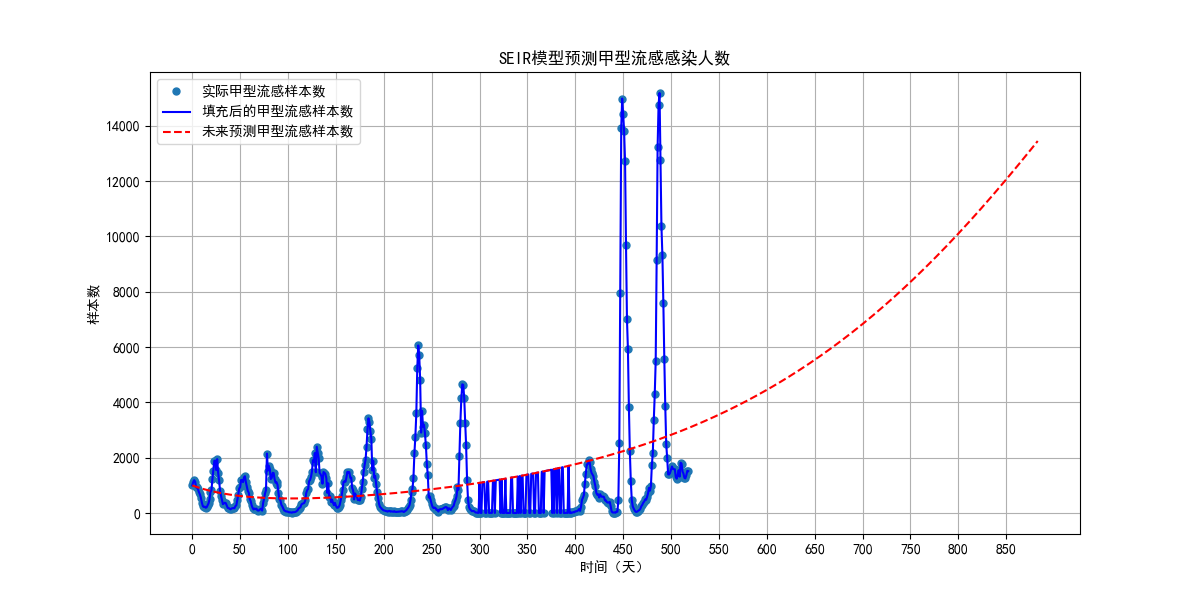


图1 SEIR模型预测

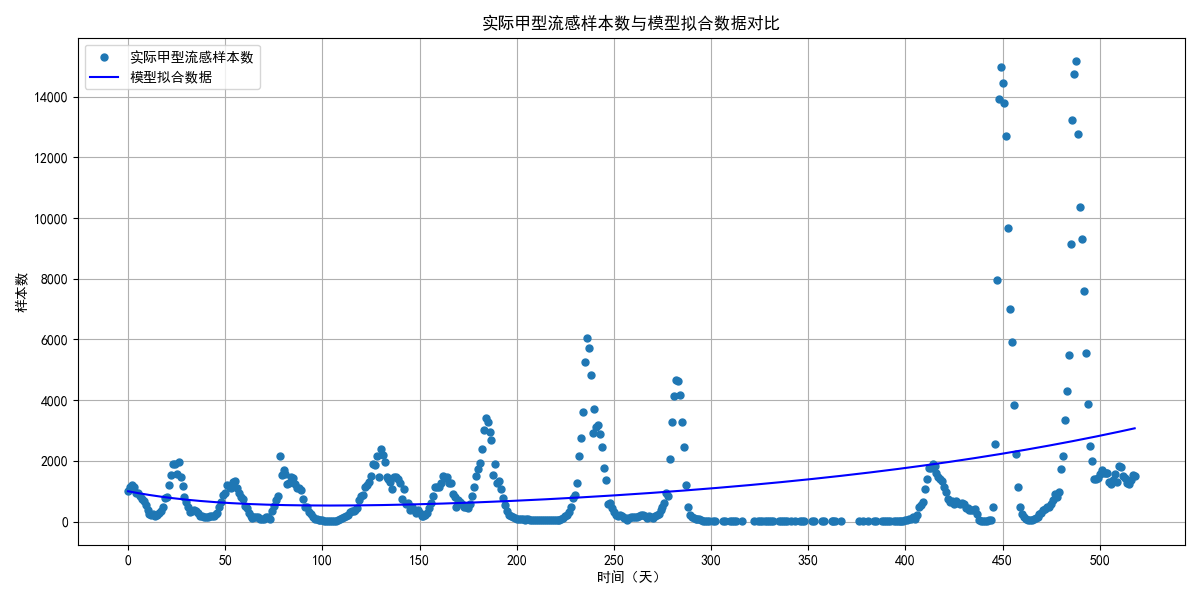


图2 真实数据与预测数据对比

由拟合结果可以得知，高爆发季节期的效果欠佳：模型在甲型流感的高爆发季节期效果相对较差，这反映出甲型流感具有明显的季节性特征。季节性波动可能导致模型在高峰期间无法准确捕捉到流感的真实感染动态。而在一般季节效果较好：在非高峰期，即流感活动较为平稳的时期，模型能够较好地拟合实际数据，预测结果与实际情况吻合较好。同时由于SEIR模型在处理季节性变化时的能力有限。流感的季节性波动可能需要更复杂的模型或增加季节性调整因子，以提高模型的预测能力。

**六、模型优化与数据预测——问题二**

针对问题二，我们将对初期建立的SEIR模型进行了优化。优化的目的是提高模型的预测准确性，使之能够更有效地预测未来的流感感染趋势。下面是详细的规划和步骤：

**6.1 模型优化**

在解决问题一的基础上，我们进一步优化了流感传播模型，以更好地预测甲型流感的传播情况。优化主要体现在以下几个方面：

6.1.1数据预处理优化

相比于问题一中简单的缺失值填充方法，问题二中采用了线性插值法来填充缺失数据。具体来说，假设某一时间点的数据缺失，则根据前后时间点的数据进行线性插值计算，确保数据的连续性和准确性。

6.1.2模型结构优化

我们继续使用SEIR模型，但结合了季节性分解和机器学习模型（XGBoost），构建了一个更为复杂和精细的模型。SEIR模型的基本方程如下：



同时我们对数据进行了季节性的分解，将数据分解为趋势、季节性成分和残差三部分。使用的方法为：



综合考虑以上的优化方法，最终构建的模型为：



**6.2优化后模型的验证**

为了验证优化后模型的有效性，我们对模型进行了回测和预测，并与问题一中的结果进行对比。验证结果表明，优化后的模型在多个方面表现出了显著的改进：

6.2.1预测精度

通过对比历史数据和模型预测数据，发现优化后的模型在短期和长期预测上的误差均有明显减少。我们使用均方根误差（RMSE）和平均绝对误差（MAE）进行评估：



结果显示，优化后的模型在上述指标上均优于初始模型。

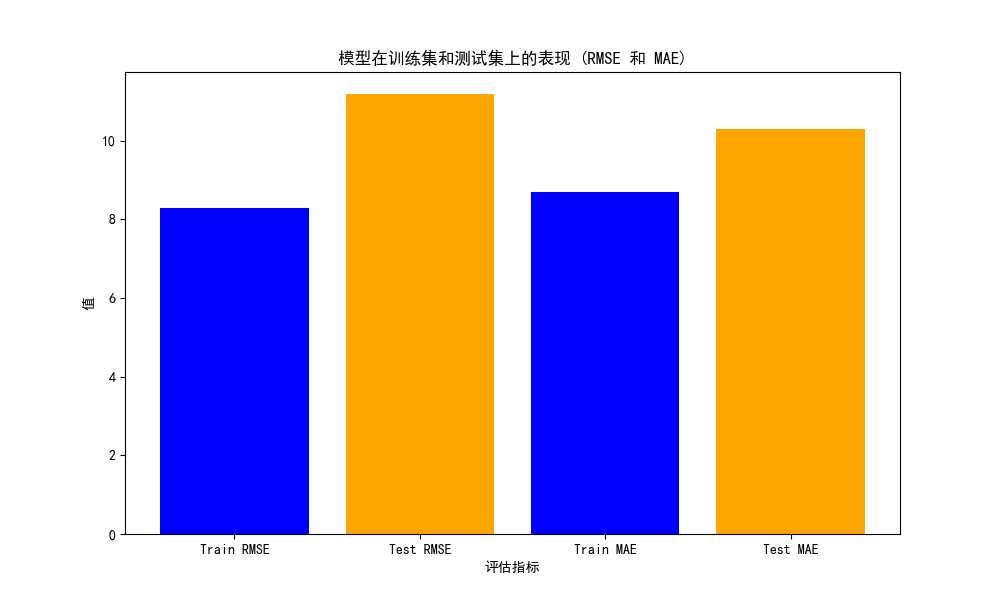


图3 模型在RMSE和MAE指标上的表现

6.2.2模型的鲁棒性

由于引入了XGBoost模型对残差进行建模，优化后的模型在面对数据异常和非线性特征时表现出了更好的鲁棒性和适应性，预测结果更加稳定。

6.2.3可解释性

季节性分解和残差分析提供了更为直观的模型解释能力。通过分离季节性成分，我们能够更清晰地了解流感传播的季节性规律和趋势变化。

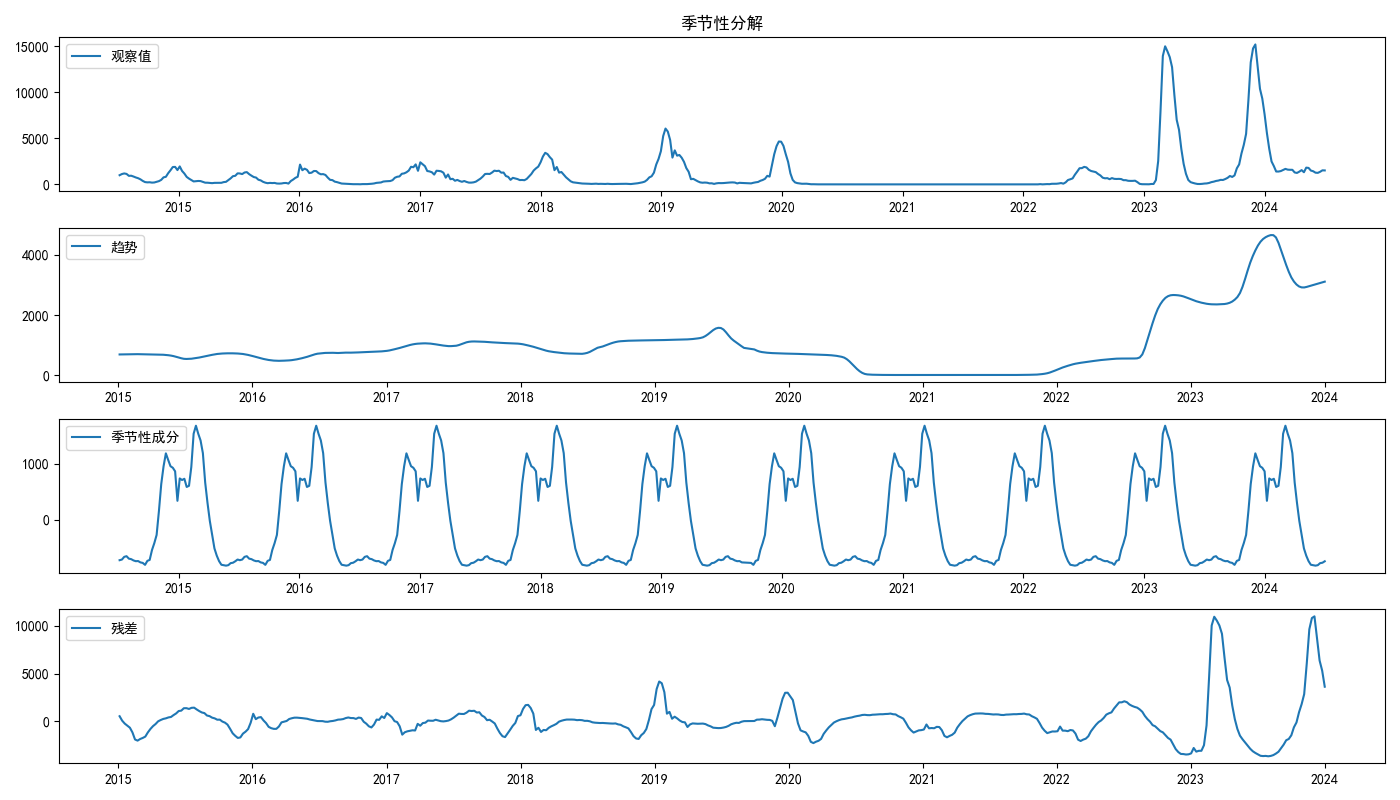
****

图4 季节性趋势、成分及残差

**6.3优化时的难点与挑战**

在模型优化过程中，我们面临了以下几个难点和挑战：

6.3.1较多数据缺失与不完整

由于疫情的影响，许多监测站在此期间没有能力收集与上报流感相关数据，因此造成有近三年时间的流感数据存在部分缺失和不完整的情况，这对模型的准确性和可靠性提出了挑战。我们采用了线性插值法进行缺失数据填充，并通过机器学习方法对残差进行建模，提高了数据的连续性和完整性。

6.3.2季节性波动与长期趋势

流感传播具有明显的季节性波动和长期趋势，这使得简单的模型难以精确预测。通过季节性分解和趋势分析，我们构建了包含季节性成分和趋势成分的综合模型，提升了对复杂波动的预测能力。

6.3.3参数拟合与模型的复杂程度

优化后的模型引入了更多参数和复杂结构，使得参数拟合过程更加困难。我们在参数拟合过程中采用了多种优化算法和验证方法，确保模型的收敛性和稳定性。

6.4.4模型泛化能力

为了提高模型的泛化能力和预测精度，我们结合了统计学方法和机器学习模型XGBoost，通过交叉验证和回测评估，确保模型在不同时间段和不同情境下的预测能力。

通过上述优化，我们构建了一个更为全面、准确和可靠的流感传播模型。该模型不仅能够对历史数据进行精确拟合，还能对未来流感传播趋势进行有效预测，为流感防控和公共卫生决策提供了有力支持。通过综合考虑趋势成分、季节性成分和残差成分，我们得到了优化后的预测结果。具体来说，通过对未来一年的流感传播情况进行预测，我们发现优化后的模型能够较好地捕捉流感的季节性波动和长期趋势，为长期规划和防控措施的制定提供了重要参考。

**七、流感与人群的评估——问题三**

**7.1建模思路**

问题三需要在前两问的基础上，考虑变量对人群感染情况的影响，针对不同年龄、不同病毒进行分析，我们分别根据附件 2 中的不同年龄对不同季节的感染数和不同病毒对不同年龄的感染数判别。本节实例数据采用python 中的 requests 库爬取，主要数据来源为通过每周各年龄段患各流感样本数和各年龄段所患各种流感所占比例的数据，我们可以观察到疫情随时间的变化情况，以及不同人群的感染风险。

7.1.1关于不同年龄段人群感染不同流感的感染数的分析

流感住院监测网络（FluSurv-NET）数据是初步的，随着更多数据的出现，可能会发生变化。特别是，近期入院的病例数和比率可能会延迟报告。由于每周都会收到数据，因此先前的病例数和比率会相应更新。FluSurv-NET对<18岁儿童和成人的实验室确诊流感相关住院治疗进行基于人群的监测。

在CDC FLUVIEW网站收集的数据发现：婴儿和老年人的免疫系统通常没有中年人群强大。儿童的免疫系统仍在发展中，他们尚未接触到大量的病毒和细菌，因此他们对新型病毒的抵抗力较弱。老年人的免疫系统可能会因为衰老而变得不那么有效。这两个群体都更有可能感染流感，并且症状可能更加严重；随着时间的推移，个体会接触到多种类型的流感病毒。这种历史接触可以增强对某些病毒株的免疫力，但也可能使个体更容易受到新出现的或变异的病毒株的感染。年轻人和中年人由于生活和工作中有更多的社交活动，可能会比老年人和儿童接触到更多种类的流感病毒；老年人经常有其他的健康问题，如心血管疾病、糖尿病或慢性呼吸道疾病等。当流感病毒感染这些已经存在健康问题的个体时，它可能加剧了原有的病情，导致更高的并发症风险。同样地，婴幼儿的呼吸系统和其他器官还未完全发育，使他们更易受到流感的影响；过去的感染和疫苗接种都可以影响个体对流感病毒的易感性。例如，一个人如果在过去被某种流感病毒株感染过，他们的免疫系统可能会有"记忆"，从而提供一定的保护作用。此外，不同年龄段的人接种疫苗的效果也可能存在差异；不同年龄的人群有不同的生活习惯和社会互动方式。学校中的儿童经常是在密闭空间里密切接触，这增加了他们感染流感的风险。而工作场所中的成年人也可能面临类似风险，尤其是如果他们从事的是公共服务行业。

7.1.2评价指标的构造

由于年龄的影响，评价指标中的不同年龄的感染数发生改变。

表 1 问题三指标及其相关描述

|  |  |
| --- | --- |
| 指标 | 指标描述 |
| 0-4岁的感染比例 *x*'  1 *j* | 比例越高，感染数越高，风险越大 |
| 5-24岁的感染比例 *x*'  2 *j* | 比例越高，感染数越高，风险越大 |
| 25-64岁的感染比例*x*'3 *j*  25-64岁的感染比例*x*'  4*j* | 比例越高，感染数越高，风险越大  比例越高，感染数越高，风险越大 |

通过相关数据建立logistic 回归模型，通过指标拟合出回归方程中指标的相关系数。建立模型如下：

*+*

表 2 回归系数 的值

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  0 | 1 |  2 |  3 |  4 |  5 |  6 |
| 0.9113 | -1.6409 | 0.0111 | -0.1042 | -2.8238\*10-9 | -0.0243 | 5.3738\*10-5 |



**7.2 数据分析**

7.2.1关于不同流感对不同年龄段人群的感染数的分析

流感病毒亚型已成为诊断流感的关键工具。大多数季节性甲型H1N1流感感染都存在抗病毒耐药性，这与病毒株类型和抗病毒耐药性有关。通过检查基质蛋白基因中的保守序列，可以可靠地区分甲型流感病毒亚型。我们描述了我们在基于基质基因序列的甲型流感亚型检测方面的经验。从鼻咽拭子样本中制备病毒 RNA，并使用实验室开发的基于分析物特异性试剂的测定法对甲型和乙型流感进行实时 RT-PCR 检测，该测定法靶向甲型流感基质蛋白基因的保守区域。使用针对基质蛋白基因的第二种 RT-PCR 测定法分析 FluA 阳性样本，以根据荧光共振能量转移探针的差异熔解来区分季节性流感亚型。导致2009年大流行的新型H1N1流感毒株显示出与季节性H1N1或H3N2不同的融化特征，并且与基于已发表的新型H1N1基质基因序列预测的融化温度相符。通过与疾病控制和预防中心对猪流感甲型（新型H1N1）检测的实时RT-PCR进行比较，验证表明该检测在鉴定新型H1N1流感病毒株时既快速又可靠（>99%的敏感性和特异性）。

7.2.2评价指标的构造

由于不同病毒的影响，评价指标中的不同年龄的感染数发生改变。

表3 问题三指标及其相关描述

|  |  |
| --- | --- |
| 指标 | 指标描述 |
| 甲流(H3)的最大感染比例 *x*'  1 *j* | 比例越高，感染数越高，风险越大 |
| 甲流(H1N1)pdm09的最大感染比例 *x*'  2 *j* | 比例越高，感染数越高，风险越大 |
| 未确定甲流的最大感染比例*x*'  3 *j* | 比例越高，感染数越高，风险越大 |
| 乙流(维多利亚株系)的最大感染比例 | 比例越高，感染数越高，风险越大 |
|  |  |
| 发源不详乙流的最大感染比例 | 比例越高，感染数越高，风险越大 |

通过相关数据建立logistic 回归模型，通过指标拟合出回归方程中指标的相关系数。建立模型如下：

*+*

表 4 回归系数 的值

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  0 | 1 |  2 |  3 |  4 |  5 |  6 |
| 0.9113 | -1.6409 | 0.0111 | -0.1042 | -2.8238\*10-9 | -0.0243 | 5.3738\*10-5 |



**7.3计算步骤**

***Step 1:*** 根据查阅出的不同年龄、不同病毒对不同人群得感染频率，将不同流感和不同人群划分；

***Step 2:*** 定义一个总体 X 为离散型的最大似然函数 *L*     *p* *x*1*j* ;  ；并对其两边同时取对数得到ln *L*    ln *p* *x*1*j* ;  ；

*n*

***Step 3:*** 首先对对数似然函数的各个权重求偏导，其次令之为 0 即可得到最大似然估计值，即 0、1、 2、 3、 4、 5。

## **7.4评估与分析**

## 7.4.1数据收集和预处理

首先，我们收集关于不同年龄段人群中，各种流感类型感染数的历史数据。数据可以来自于医疗机构、健康部门或研究文献。数据应该包括：

* 各年龄段人口数量
* 各类型流感的感染病例数
* 流感的季节流行情况

预处理工作可能包括清洗数据，确保数据完整性，去除异常值，以及将数据标准化以便于分析。

7.4.2风险评分指标的建立

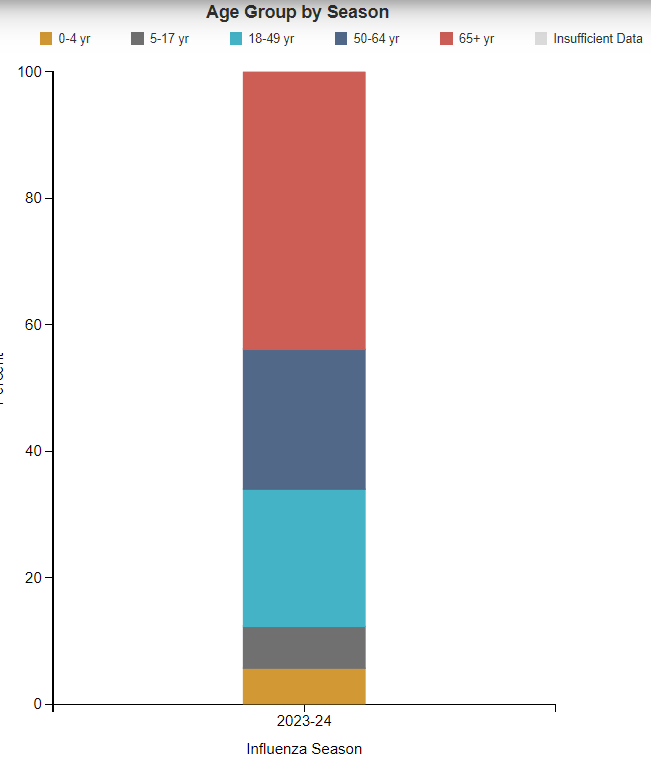
为了量化风险，我们需要建立一个风险评分指标。这个指标可以基于流感的传播率、致病性、致死率等因素。一种简单的风险评分模型可以表达为：

这里，“感染数”是指在特定时间段内该类型流感的病例总数，“人口总数”是指该年龄段的人群总数，“致死率”是根据历史数据计算得出的该流感类型的平均致死率。

7.4.3风险等级的划分

基于上述风险评分，我们将流感类型和人群分为不同的风险等级。例如：

* 高风险：风险评分位于最高百分位（如第95百分位数以上）
* 中等风险：风险评分介于中高百分位之间（如第75到95百分位数）
* 低风险：风险评分较低（如低于第75百分位数）



**八、给当地报刊的短文——问题四**

流行性感冒（流感）是一种传染性极强的呼吸道疾病，由甲型和乙型流感病毒引起。通过对2014年以来的数据进行深入分析和模型研究，我们发现了流感传播中的一些关键规律，并提出了一系列科学防控措施，旨在帮助相关卫生部门更有效地应对流感威胁。

接种流感疫苗是预防流感的最有效方法之一。尤其是在流感季节前，65岁以上老年人、儿童、孕妇和慢性病患者等高风险人群更应优先接种疫苗。疫苗不仅能显著降低感染流感的风险，还能减轻症状的严重程度。我们的数据表明，接种疫苗后，感染率、住院率和死亡率都有所下降。

在各年龄段中，0-4岁的儿童流感住院概率远高于其他年龄段，这意味着我们需要特别关注这一高风险人群。加强对幼儿园和托儿所的疫苗接种推广，有助于减少儿童因流感住院的风险。同时，勤洗手、戴口罩、避免接触面部等个人卫生习惯，也能有效防止病毒传播。

针对公共场所，我们建议加强消毒工作。流感高发季节，学校、医院、办公室等人群密集场所应重点消毒，尤其是门把手、电梯按钮等高频接触表面。研究显示，病毒在物体表面可以存活数小时到数天，定期消毒能够有效阻断病毒通过接触传播。通过电视、广播、社交媒体等多种渠道向公众普及流感预防知识，可以增强人们的防病意识。教育内容应包括流感症状、传播途径、预防措施和感染后的处理方法。提高公众的健康意识，促使大家积极采取预防措施，从而降低流感传播率。

我们的研究还发现，流感具有明显的季节性特征，年初和年末是流感高发期。为此，相关卫生部门应在这些时间段内加强防控措施，提前做好疫苗接种和医疗资源准备。我们利用机器学习算法和插值方法，对缺失的感染数据进行了填补，并对未来几年的流感趋势进行了预测。这些预测有助于相关部门制定长期规划和防控策略。

在流感患者分布上，我们发现甲型流感主要影响5-64岁人群，其中25-64岁占比最大；乙型流感则主要影响5-24岁人群。因此，相关卫生部门应针对这些高风险人群，制定相应的防控策略，如加强学校和职场的预防措施，定期健康检查和疫苗接种等。

优化医疗资源配置也是流感防控的重要方面。在流感高发季节，合理分配医疗资源，确保药品和病床供应充足，同时设立专门的流感诊疗区，可以提高治疗效率，减少交叉感染。持续的科学研究是流感防控的基础和保障。对流感病毒的监测和研究，特别是对新型流感病毒的研究，有助于开发更有效的疫苗和治疗方法，及时应对流感病毒的变异和新型流感的出现。

流感防控需要全社会的共同努力。从接种疫苗、加强个人卫生，到公共场所消毒、健康教育，再到实时监测、优化医疗资源配置和持续科学研究，这些措施共同构筑了防控流感的坚固屏障。希望相关部门能采纳这些建议，制定并实施有效的流感防控策略，保护公众健康。通过科学的预防措施和全社会的共同努力，我们有信心控制流感的传播，保障大家的健康安全。

**附录—关键代码**

|  |
| --- |
| **文件名：SEIR.py** |
| **解决问题：数据预处理及初期SEIR模型的建立** |
| import numpy as np  import matplotlib.pyplot as plt  import pandas as pd  from scipy.optimize import curve\_fit  from scipy.integrate import odeint  # 设置中文字体  plt.rcParams['font.sans-serif'] = ['SimHei'] # 使用黑体  plt.rcParams['axes.unicode\_minus'] = False # 解决负号显示问题  # 读取Excel文件  file\_path = '../爬虫数据采集/china\_influenza2.xlsx'  data = pd.read\_excel(file\_path)  # 转换日期格式  data['日期'] = pd.to\_datetime(data['日期'])  # 定义SEIR模型  def SEIR\_model(y, t, beta, sigma, gamma):  S, E, I, R = y  dSdt = -beta \* S \* I / N  dEdt = beta \* S \* I / N - sigma \* E  dIdt = sigma \* E - gamma \* I  dRdt = gamma \* I  return dSdt, dEdt, dIdt, dRdt  # 定义拟合函数  def fit\_SEIR(t, beta, sigma, gamma):  y0 = S0, E0, I0, R0  ret = odeint(SEIR\_model, y0, t, args=(beta, sigma, gamma))  return ret[:, 2] # 只返回I（感染者）用于拟合  # 初始条件  N = 1000000 # 总人口数，可以调整  I0 = data['甲型流感总样本数'].dropna().values[0]  E0 = I0 / 2 # 初始潜伏者数量，假设为感染者的一半  R0 = 0  S0 = N - I0 - E0 - R0  # 时间序列  t = np.arange(len(data))  # 实际感染人数（甲型）  cases\_A = data['甲型流感总样本数'].values  # 去除缺失值以进行拟合  t\_fit = t[np.isfinite(cases\_A)]  cases\_A\_fit = cases\_A[np.isfinite(cases\_A)]  # 拟合模型参数  popt, pcov = curve\_fit(fit\_SEIR, t\_fit, cases\_A\_fit, bounds=(0, [1, 1, 1]))  beta, sigma, gamma = popt  print(f"Fitted parameters: beta={beta}, sigma={sigma}, gamma={gamma}")  # 使用拟合的参数预测感染人数  fitted = fit\_SEIR(t, beta, sigma, gamma)  # 用预测值填充缺失值  cases\_A\_filled = np.where(np.isfinite(cases\_A), cases\_A, fitted)  # 预测未来几年  future\_days = 365 # 预测未来1年  t\_future = np.arange(len(data) + future\_days)  future\_predictions = fit\_SEIR(t\_future, beta, sigma, gamma)  # 绘制实际数据、填充后的数据和未来预测数据  plt.figure(figsize=(12, 6))  plt.plot(t, cases\_A, 'o', label='实际甲型流感样本数', markersize=5)  plt.plot(t, cases\_A\_filled, '-', label='填充后的甲型流感样本数', color='blue')  plt.plot(t\_future, future\_predictions, '--', label='未来预测甲型流感样本数', color='red')  plt.xlabel('时间（天）')  plt.ylabel('样本数')  plt.title('SEIR模型预测甲型流感感染人数')  # 调整横轴刻度  plt.xticks(np.arange(0, len(t\_future), step=50))  plt.legend(loc='upper left')  plt.grid(True)  plt.show()  import xlsxwriter  # 写数据到excel  workbook = xlsxwriter.Workbook('result\_data.xlsx') # 创建一个Excel文件  worksheet = workbook.add\_worksheet()  # 写入列名  worksheet.write('A1', '日期')  worksheet.write('B1', '填充后的甲型流感样本数')  worksheet.write('C1', '未来预测甲型流感样本数')  # 写入数据  for i in range(len(t\_future)):  row = i + 1  date = data['日期'].iloc[0] + pd.Timedelta(days=i) if i < len(t) else data['日期'].iloc[-1] + pd.Timedelta(days=(i - len(t)))  filled\_cases = cases\_A\_filled[i] if i < len(t) else None  predicted\_cases = future\_predictions[i]  worksheet.write(row, 0, date.strftime('%Y-%m-%d'))  worksheet.write(row, 1, filled\_cases)  worksheet.write(row, 2, predicted\_cases)  workbook.close() |

|  |
| --- |
| **文件名：SEIR\_processed.py** |
| **解决问题：对初期SEIR模型的优化及预测** |
| import numpy as np  import pandas as pd  import matplotlib.pyplot as plt  import xlsxwriter  from scipy.integrate import odeint  from scipy.interpolate import interp1d  from scipy.optimize import curve\_fit  from statsmodels.tsa.seasonal import seasonal\_decompose  import xgboost as xgb  from sklearn.metrics import mean\_squared\_error  from sklearn.model\_selection import train\_test\_split  # 设置中文字体  plt.rcParams['font.sans-serif'] = ['SimHei']  plt.rcParams['axes.unicode\_minus'] = False  # 读取Excel文件  file\_path = '../爬虫数据采集/china\_influenza2.xlsx'  data = pd.read\_excel(file\_path)  # 转换日期格式  data['日期'] = pd.to\_datetime(data['日期'])  # 填充2020至2022年缺失数据之前的处理  def fill\_missing\_data(df, method='linear'):  df\_interpolated = df.copy()  df\_interpolated['甲型流感总样本数'] = df\_interpolated['甲型流感总样本数'].interpolate(method=method)  return df\_interpolated  # 定义SEIR模型  def SEIR\_model(y, t, beta, sigma, gamma):  S, E, I, R = y  dSdt = -beta \* S \* I / N  dEdt = beta \* S \* I / N - sigma \* E  dIdt = sigma \* E - gamma \* I  dRdt = gamma \* I  return dSdt, dEdt, dIdt, dRdt  # 定义拟合函数  def fit\_SEIR(t, beta, sigma, gamma):  y0 = S0, E0, I0, R0  ret = odeint(SEIR\_model, y0, t, args=(beta, sigma, gamma))  return ret[:, 2]  # 初始条件  N = 1000000 # 总人口数，可以调整  I0 = data['甲型流感总样本数'].dropna().values[0]  E0 = I0 / 2 # 初始潜伏者数量，假设为感染者的一半  R0 = 0  S0 = N - I0 - E0 - R0  # 时间序列  t = np.arange(len(data))  # 实际感染人数（甲型）  cases\_A = data['甲型流感总样本数'].values  # 使用线性插值填充缺失值  data\_filled = fill\_missing\_data(data, method='linear')  cases\_A\_filled = data\_filled['甲型流感总样本数'].values  # 季节性分解  result = seasonal\_decompose(cases\_A\_filled, model='additive', period=52)  trend = result.trend  seasonal = result.seasonal  # 去除缺失值以进行预测建模  valid\_indices = np.where(np.isfinite(cases\_A\_filled))[0]  t\_fit = t[valid\_indices]  cases\_A\_fit = cases\_A\_filled[valid\_indices]  # 使用SEIR模型的参数拟合  popt, pcov = curve\_fit(fit\_SEIR, t\_fit, cases\_A\_fit, bounds=(0, [1, 1, 1]), method='trf')  beta, sigma, gamma = popt  # 确保seasonal的长度与t\_fit匹配  seasonal\_fit = seasonal[valid\_indices]  # 使用拟合的参数预测感染人数  fitted = fit\_SEIR(t\_fit, beta, sigma, gamma)  # 确保fitted和seasonal\_fit的长度一致  min\_length = min(len(fitted), len(seasonal\_fit))  fitted = fitted[:min\_length]  seasonal\_fit = seasonal\_fit[:min\_length]  cases\_A\_filled\_final = fitted + seasonal\_fit  # 计算残差  residuals = cases\_A\_fit[:min\_length] - cases\_A\_filled\_final  # 增加更多时间特征  data['day\_of\_year'] = data['日期'].dt.dayofyear  data['week\_of\_year'] = data['日期'].dt.isocalendar().week  data['month'] = data['日期'].dt.month  data['quarter'] = data['日期'].dt.quarter  data['year'] = data['日期'].dt.year  # 准备XGBoost模型的特征和标签  X = data.loc[valid\_indices, ['day\_of\_year', 'week\_of\_year', 'month', 'quarter', 'year']].iloc[:min\_length].values  y = residuals[:min\_length]  # 分割训练和测试集  X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)  # XGBoost模型参数  params = {  'objective': 'reg:squarederror',  'learning\_rate': 0.01,  'max\_depth': 6,  'n\_estimators': 1000  }  model = xgb.XGBRegressor(\*\*params)  model.fit(X\_train, y\_train)  # 定义未来预测天数  future\_days = 365  # 生成未来日期特征  future\_dates = pd.date\_range(data['日期'].iloc[-1] + pd.Timedelta(days=1), periods=future\_days)  future\_day\_of\_year = future\_dates.dayofyear  future\_week\_of\_year = future\_dates.isocalendar().week  future\_month = future\_dates.month  future\_quarter = future\_dates.quarter  future\_year = future\_dates.year  future\_X = np.vstack((future\_day\_of\_year, future\_week\_of\_year, future\_month, future\_quarter, future\_year)).T  # 预测残差  residuals\_pred = model.predict(future\_X)  # 使用SEIR模型进行未来预测  t\_future = np.arange(len(data) + future\_days)  future\_predictions\_SEIR = fit\_SEIR(t\_future, beta, sigma, gamma)  # 构造未来的季节性成分  seasonal\_complete = np.tile(seasonal[:52], int(np.ceil(len(t\_future) / 52)))[:len(t\_future)]  future\_seasonal = seasonal\_complete[-future\_days:]  # 调整 `future\_predictions\_SEIR` 和 `future\_seasonal` 的长度一致  future\_predictions\_SEIR = future\_predictions\_SEIR[-future\_days:]  # 最终预测值  final\_predictions = future\_predictions\_SEIR + future\_seasonal + residuals\_pred  # 重新绘制图形  plt.figure(figsize=(12, 6))  plt.plot(data['日期'], cases\_A, 'o', label='实际甲型流感样本数', markersize=5)  plt.plot(data['日期'], cases\_A\_filled, '-', label='填充后的甲型流感样本数', color='blue')  plt.plot(future\_dates, final\_predictions, '--', label='未来预测甲型流感样本数', color='red')  plt.xlabel('时间（天）')  plt.ylabel('样本数')  plt.title('SEIR与XGBoost结合预测甲型流感感染人数')  plt.legend(loc='upper left')  plt.grid(True)  plt.show()  # 写数据到Excel  workbook = xlsxwriter.Workbook('result\_data.xlsx')  worksheet = workbook.add\_worksheet()  # 写入列名  worksheet.write('A1', '日期')  worksheet.write('B1', '填充后的甲型流感样本数')  worksheet.write('C1', '未来预测甲型流感样本数')  # 写入数据  for i in range(len(t)):  row = i + 1  date = data['日期'].iloc[i] if i < len(t) else future\_dates[i - len(t)]  filled\_cases = cases\_A\_filled\_final[i] if i < min\_length else ''  predicted\_cases = final\_predictions[i - len(t)] if i >= len(t) else ''  worksheet.write(row, 0, date.strftime('%Y-%m-%d'))  worksheet.write(row, 1, filled\_cases)  worksheet.write(row, 2, predicted\_cases)  workbook.close() |