

基于GEP的高速公路通行费预测方法研究

焦虹桥

甘肃省兰州高速公路处

摘要: 高速公路通行费预测对于交通规划和管理具有关键性作用, 本研究采用基因表达式编程 (GEP) 算法提高了通行费预测的准确度和效率。研究探讨了GEP算法的理论基础和特征, 随后详细说明了预测模型的构建过程, 模型开发涉及数据处理、训练以及后续的验证和参数优化, 通过实际案例分析模型展现其在高速公路通行费预测方面的有效性和实用性, 研究结果表明, GEP算法在处理复杂的交通数据预测问题时表现出卓越的性能, 为高速公路通行费收入预测提供了一种创新方法。

关键词: 高速公路通行费; 预测模型; 基因表达式编程 (GEP)

【DOI】10.12252/j.issn.2096-627X.2022.06.205

引言

高速公路通行费收入的准确预测对于交通系统的有效管理和规划至关重要, 然而现有的预测方法在应对数据的复杂性和非线性方面存在局限, 为此本研究引入了基因表达式编程 (GEP) 算法, 该算法在处理复杂和非线性数据方面显示出显著优势, 研究着重于探讨GEP算法在高速公路通行费预测中的应用, 包括模型的构建、数据处理、训练及验证等关键步骤, 研究通过实例分析展示了模型的应用效果和前景, 旨在为高速公路通行费收入预测领域提供一种有效的解决方案。

一、高速公路通行费预测的理论与实践背景

高速公路通行费预测在交通规划与管理中占据了重要的地位, 它涉及对未来收入流的估算, 为决策制定提供基础支撑。预测的准确性直接影响到资金分配、基础设施建设和运营维护计划的制定, 在传统预测方法中常见的线性回归和时间序列分析往往受限于其假设条件和对历史数据的依赖, 难以准确捕捉到交通流量和通行费收入的非线性特征, 此外通行费收入受多种因素影响, 如政策变动、经济发展水平、节假日效应及气候条件等, 这些因素的复杂交互作用使得预测任务变得更为困难。近年来随着大数据技术的发展和计算能力的提高, 机器学习方法被引入到交通数据分析中, 其中基因表达式编程 (GEP) 作为一种有效的机器学习技术, 因其在处理复杂数据和挖掘深层次数据模式方面的优势, 逐渐成为高速公路通行费预测的新兴工具。GEP能够自动演化出解决特定问题的最佳或近似最佳数学模型, 无须事先对数据分布或关系的假设, 使其特别适合处理高速公路通行费数据的非线性和不确定性特点。

二、GEP算法在通行费预测中的应用

1. GEP算法原理与特点

基因表达式编程 (GEP) 算法是一种高效的演化算法, 融合了遗传算法 (GA) 的全局搜索特性和遗传编程 (GP) 的表达式树结构优势。GEP的核心在于其独特的基因编码方式, 其中每个染色体由多个基因组成, 每个基因包含两部分: 头部和尾部。头部由函数和终结符组成, 而尾部仅包含终结符, 这种结构使得GEP染色体在进化过程中保持稳定, 同时保留高度的多样性, 例如给定一个头部长度为 hh , 最大的参数个数 nn 为2, 那么尾部的长度 tt 将由公式 $t = h \times (n - 1) + 1$ 计算得出, 这使得GEP算法在解决复杂问题时具备高效的编码和解码机制, 在进化过程中GEP通过选择、变异、插串和重组等遗传操作来生成新的个体。变异操作通过改变染色体上的随机位点, 启用新的解决方案探索, 从而扩展解空间的范围; 而插串操作则允许不同染色体间的基因片段交换, 增加了种群的多样性; 重组操作类似于传统遗传算法中的交叉, 通过交换父代染色体的部分基因来产生新后代, 在通行费预测的应用中, GEP可以根据历史数据自动演化出复杂的数学模型, 这些模型能够准确捕捉通行费收入与影响因素之间的非线性关系, 例如当考虑季节性因素、经济指标和交通流量等多个变量时, GEP通过其强大的功能符和终结符集合构建出能够描述这些变量间复杂动态关系的模型, 从而为高速公路运营提供精确的收入预测^[1]。

2. 构建通行费预测模型

构建高速公路通行费预测模型的关键在于精确捕捉通行费收入与各种影响因素之间的关系, 在使用GEP算法时首先需要历史通行费数据进行深入分析, 以识别其潜在的模式和趋势, 这一分析包括季节性变化的

识别、节假日效应的评估以及经济发展水平对通行费收入的影响，例如通过分析历史数据可以发现通行费收入 R 与节假日 H 经济指标 E 和交通流量 T 之间存在关系 $R = f(H, E, T)$ ，其中 f 是一个未知的复杂函数，接下来确定GEP算法中的函数集和终结符集，这些集合包括构建模型所需的运算符和变量。函数集包含如加（+）、减（-）、乘（×）、除（÷）等基本运算符以及更复杂的函数，如指数 e^x 对数 $\log(x)$ 等；终结符集则包括表示影响因素的变量 H, E, T ，根据这些选择生成初始种群，该种群由随机生成的多个符合GEP编码结构的染色体构成，通过适应度函数评价这些染色体的性能，如使用均方误差（MSE） $\frac{1}{n} \sum_{i=1}^n (R_i - \hat{R}_i)^2$ 作为适应度函数，其中 R_i 和 \hat{R}_i 分别是实际和预测的通行费收入，随后通过遗传操作优化种群，直至满足预定的终止条件，最终演化出的模型能够基于输入的影响因素准确预测通行费收入，其性能可以通过独立测试集进行验证，这种方法不仅提升了预测的准确性，也为高速公路管理和政策制定提供了科学依据。

3. 数据处理与模型训练

在高速公路通行费预测模型的构建过程中，数据处理和模型训练是至关重要的步骤。数据处理的首要任务是从原始的通行费数据中提取有用信息，并将其转换为适合GEP算法处理的格式。假设原始数据包括每日的通行车辆数、收费类型、气候条件等多个变量，首先需要进行数据清洗，去除异常值和缺失值。例如，可以使用插值方法填补缺失数据，或者应用离群值检测技术剔除异常数据点。接下来，进行数据标准化处理，例如，将车辆数和收入数据转换为标准分数，方法是减去均值后除以标准差，从而消除不同量纲带来的影响。然后，对数据进行特征工程，如通过时间序列分析提取季节性因素，或者构建经济指数作为预测模型的输入变量。例如，可以构建一个综合指数 I ，综合考虑GDP增长率、汽车销量和节假日数量等因素，公式为 $I = \omega_1 \times \text{GDP增长率} + \omega_2 \times \text{汽车销量} + \omega_3 \times \text{节假日数量}$ ，其中 $\omega_1, \omega_2, \omega_3$ 为权重系数。数据处理完成后接着进行模型训练，首先初始化GEP算法的参数，如种群大小设为50，最大迭代次数设为5000，随后开始迭代过程，每次迭代中根据适应度函数选择优秀个体，然后通过遗传操作生成新的个体，例如设定变异概率为0.01，交叉概率为0.7，保持

算法的探索和利用平衡，训练过程中可以使用交叉验证方法评估模型性能，如使用五折交叉验证，将数据集分为五份，轮流使用其中四份进行训练，剩下一份用于验证^[2]。

4. 模型验证与参数调优

在高速公路通行费预测的遗传表达式编程（GEP）模型验证与参数调优中，首先通过比较模型预测结果与实际通行费收入数据的差异来进行模型验证，主要使用均方根误差（RMSE）和平均绝对误差（MAE）这样的评估指标来反映模型预测的准确度，然后在参数调优方面的关键在于找到最适合数据集的算法配置，包括种群大小、遗传算子概率、基因头尾长度等，通常通过网格搜索或随机搜索等方法来实现，例如设置种群大小的搜索范围为30到100，遗传算子概率的搜索范围为0.01到0.1，在这些范围内进行多次尝试并评估模型性能，从而找到最优的参数组合，如可能发现种群大小为60，变异概率为0.05，交叉概率为0.8时模型达到最佳性能，同时还需考虑防止过拟合的问题，确保模型不仅在训练集上表现良好，也能在未见数据上保持稳定的预测性能，一种方法是采用正则化技术，如在适应度函数中添加正则化项，减少模型复杂度，增强其泛化能力，通过这些方法可以有效提升GEP预测模型在高速公路通行费预测中的精确性和可靠性，为相关决策提供强有力的数据支持。

三、模型应用与案例分析

1. 实例数据分析

在实例数据分析阶段通过收集和分析多年高速公路的通行数据，对高速公路的运营效率和经济性进行了深入探讨。数据收集涵盖不同时间段的车流量、车型分布、天气状况、节假日排布和区域经济发展等多种因素，确保了数据的全面性和代表性，通过时间序列分析识别通行费收入的季节性变化和长期趋势，揭示其车流量与经济活动、气候条件之间的关联。对于特殊事件如节假日减免政策的实施，分析了其对通行费收入的短期和长期影响，数据分析不仅包括描述性统计，也运用了高级统计技术如聚类分析和主成分分析，从而在多维度上理解数据的特性，这一阶段的分析为构建高效的预测模型奠定了坚实的基础，确保了后续预测模型的准确性和可靠性^[3]。表1呈现了实例数据的关键统计和预测模型的性能评估结果。

表1 实例数据统计

数据类别	描述	2019年	2020年	2021年
日均车流量	每天通过特定收费站的平均车辆数	12, 500	13, 800	14, 200
节假日车流量	特定节假日的平均车流量	18, 000	19, 500	20, 000
车型分布	不同车型（小型、大型）的车流量比例	小型：70%，大型：30%	小型：72%，大型：28%	小型：75%，大型：25%
收入增长率	年度通行费收入增长率	5.2%	6.0%	6.5%
特殊事件影响	节假日减免政策对收入的影响	-3.0%	-2.5%	-2.0%
经济活动指数	区域经济活动对车流量的影响	1.05	1.08	1.10

2. 预测结果与评估

基于之前分析的实例数据，通过训练得到的模型对未来一段时间内的通行费收入进行了预测，预测结果显示模型能够有效捕捉到通行费收入的季节性波动和长期趋势，为评估模型的性能采用了多种指标如均方根误差（RMSE）、平均绝对误差（MAE）和确定系数（ R^2 ）。模型预测的准确度在不同时间尺度上进行了验证，包括短期（一个月内）、中期（一季度）和长期（一年以上）的预测，通过与传统的时间序列预测模型和其他先进的机器学习模型进行比较，GEP模型在多数情况下展现出更优的预测性能，特别是在应对复杂的非线性模式和数据中的异常值时表现出色，最后针对模型在某些特定场景下的局限性进行了分析，并提出了相应的改进策略，以进一步提升模型的预测准确度和泛化能力^[4]。表2展示了不同模型在高速公路通行费预测任务上的性能对比。

表2 模型预测性能评估

评估指标	GEP模型	传统时间序列模型	其他机器学习模型
RMSE	0.75	1.20	0.85
MAE	0.58	0.95	0.65
R^2	0.92	0.85	0.89
短期预测准确度	95%	88%	90%
长期预测准确度	90%	80%	85%

3. 模型的实际应用和前景

在基于基因表达式编程（GEP）的高速公路通行费预测模型的优化阶段，模型的调整与改进是至关重要的。优化的核心在于增强模型对复杂数据的解释力和预测准确性，同时确保其对新数据具有良好的泛化能力，在实践中这通常涉及多个方面：（1）进行参数调整，如调整遗传算法的变异率、交叉率以及种群大小，以找到最佳的参数组合，这些参数对模型的训练过程和预测性能有着显著影响。（2）考虑模型的复杂性和运算效率的平衡，避免过度复杂的模型导致过拟合，例如可以通过剪枝技术减少模型中不必要的表达式，保持模型的

简洁性。此外，引入新的数据特征或调整现有特征的处理方式也是优化的重要环节，例如考虑到高速公路通行费收入受天气、节假日、经济活动等多种因素影响，将这些因素纳入模型可以提高预测的准确度。（3）持续监测模型在实际应用中的表现，根据实际应用中遇到的新挑战和问题对模型进行迭代更新和细微调整，通过这些优化策略可以确保GEP模型在高速公路通行费预测中保持高效且准确，为高速公路管理和决策提供可靠的数据支持^[5]。

结论

本研究采用基因表达式编程（GEP）算法成功构建了高速公路通行费预测模型，证明了该模型在预测精度和泛化能力上的优越性，通过实例分析验证了模型的实用性，展示其在实际应用中的潜力。未来，随着技术的进步这种方法有望在更广泛的领域内应用，为解决复杂预测问题提供有效工具，本研究不仅为高速公路通行费预测提供了新的视角，也为相关领域的研究提供了借鉴。

参考文献

- [1] 蒋强, 唐敏. 基于收费政策的高速公路通行费预测方法研究[J]. 公路工程, 2019, 44(05): 263-268.
- [2] 钱文海, 葛斌, 刘玲娜等. 基于GEP的微分方程模型在高速公路通行费预测中的应用[J]. 武汉理工大学学报, 2018, 40(08): 55-62+69.
- [3] 刘宁, 黄樟灿, 谈庆. 基于GEP的高速公路通行费预测方法研究[J]. 计算机应用研究, 2019, 36(07): 1998-2002.
- [4] 蔡晓, 李燕鸿, 李培骏. 高速公路通行费收入预测方法研究[J]. 公路与汽运, 2018, (02): 173-177+182.
- [5] 孙娜, 李珏, 陈强. 高速公路通行费收入短期预测方法研究[J]. 交通企业管理, 2016, 31(09): 32-34.