

基于 GA-BP 神经网络电商平台公司数据资产价值研究 —— 以拼多多为例



金明丹*

重庆理工大学经济金融学院, 重庆 401320

摘要: 随着物联网技术的飞速进步, 数据已经变成了企业中最有价值的资产之一。因此, 对数据资产进行合理有效地评估具有重要意义。但是, 传统的资产数据评估手段确实有其固有的缺陷。为了解决这一问题, 需要建立一套适合于大数据分析和数据挖掘的数据资产评估模型。研究通过对文献的深入探讨, 提出了一个以拼多多电商平台为研究对象的基于逆向传播神经的算法。在此基础上, 列出了 7 个会影响数据资产价值的核心指标, 构建了一个 GA-BP 神经网络的评估模型。在对模型的结果进行深入分析后, 发现该模型在评定数据资产的价值上展现出了卓越的精确度和高效性, 为公司在数据资产的管理和决策过程中提供了坚实的科学支撑。

关键词: 数据资产; 拼多多; GA-BP 神经网络; 电商平台

DOI: [10.57237/j.wjeb.2024.02.003](https://doi.org/10.57237/j.wjeb.2024.02.003)

Research on the Value of Data Assets of Companies Based on GA-BP Neural Network E-commerce Platform — Take Pinduoduo as an Example

Mingdan Jin*

School of Economics and Finance, Chongqing University of Technology, Chongqing 401320, China

Abstract: With the rapid advancement of IoT technology, data has become one of the most valuable assets in a business. Therefore, it is important to have a reasonable and effective evaluation of data assets. However, traditional asset data valuation methods do have their inherent drawbacks. In order to solve this problem, it is necessary to establish a set of data asset evaluation models suitable for big data analysis and data mining. Through an in-depth discussion of the literature, this paper proposes an algorithm based on reverse propagation nerves based on Pinduoduo e-commerce platform. On this basis, 7 core indicators that will affect the value of data assets are listed, and an evaluation model of GA-BP neural network is constructed. After in-depth analysis of the results of the model, it is found that the model shows excellent accuracy and efficiency in evaluating the value of data assets, which provides solid scientific support for the company in the management and decision-making process of data assets.

Keywords: Data Assets; Pinduoduo; GA-BP Neural Network; E-commerce Platforms

基金项目: 重庆理工大学研究生创新项目资助 (基金号: gzlcx20233269).

*通信作者: 金明丹, a13803891724@163.com

收稿日期: 2024-10-14; 接受日期: 2024-11-19; 在线出版日期: 2025-01-09

<http://www.wjeconbus.com>

1 引言

进入 21 世纪, 数字经济的规模、覆盖面和影响力空前增强, 成为改变全球资源配置、经济结构和竞争格局的关键力量。依托中国海量数据资源, 加快数据要素市场化, 是把握数字时代转型发展主动权的关键。为此, 财政部要求自 2024 年 1 月 1 日起将数据资产纳入财务报表, 而数据资产估值是其中的前提条件。然而, 数据资产的独特属性对这一过程提出了挑战。与传统资产不同, 数据资产具有权属不明确、来源多样、管理复杂、结构动态等特点, 难以预先评估。现有的数据交易定价方法大多不规范, 缺乏统一的定价规则。因此, 本研究通过构建 GA-BP 神经网络模型, 对中国领先的电商平台拼多多的数据资产价值进行评估, 以解决数据资产估值的难题。研究结果将为拼多多和其他电子商务平台的数据资产价值驱动因素和优化策略提供宝贵见解。

数据资产的研究尚处于萌芽阶段, 如何界定数据资产并采用合适的方法来评估数据资产的价值是充分挖掘数据资产价值, 促成数据资产交易的关键问题。张迎 (2018) [1]强调了在交易平台上实现大数据产品市场化、标准化定价机制的挑战。首先, 大数据交易所通常根据交易量收取服务费, 这激励交易所高估并出售大数据资产以获取利益。其次, 大数据交易信息相对缺乏透明度, 可能导致大数据资产价格偏离正常市场价格。此外, 现有交易市场难以找到合适的可比物, 使得市场法不适合评估数据资产价值。李永红和张淑雯 (2018) [2]则认为数据资产为企业带来的收益往往与企业的产品交织在一起, 二者难以区分, 故收益法不适用于数据资产价值的评估。除了数据资产未来的收益难以确定, 数据资产未来的预测收益时限以及相应的折现率, 也是现阶段面临的难题, 因此学者普遍认为收益法并不是一种理想的评估方法。资产价值评估是否适用成本法, 目前仍存在争议。李泽红和檀晓云 (2018) [3]等明确指出, 成本法只适用于随着时间推移而贬值的资产, 而数据资产却会随着持续挖掘而增值, 所以成本法无法有效评估数据资产价值。因此, 我们认识到传统的数据资产评估方法存在明显的局限性, 市场法、收益法和成本法都未能充分考虑数据资产的独特价值和市场行为。

GA-BP 神经网络是一种结合遗传算法 (GA) 和反向传播 (BP) 神经网络的混合模型, 被广泛应用于数据资产价值评估。该方法通过以下步骤实现: 首先,

收集并整理与数据资产相关的各类数据, 进行清洗、预处理和特征提取, 构建能够准确反映数据资产价值的特征向量。其次, 构建 GA-BP 神经网络模型。其中, 遗传算法用于优化神经网络的初始权值和结构, 以提高模型的全局搜索能力, 避免陷入局部最优。BP 神经网络则负责学习数据特征与数据资产价值之间的复杂映射关系。最后, 利用准备好的训练数据集对 GA-BP 神经网络进行训练, 使模型能够准确地预测数据资产的价值。

GA-BP 神经网络结合了遗传算法 (GA) 的全局搜索能力和反向传播 (BP) 神经网络的非线性学习能力, 使其在处理复杂非线性问题时具有显著优势。通过 GA 的全局优化, 能够有效避免 BP 神经网络陷入局部最优, 从而找到更优的权重和网络结构。此外, GA-BP 神经网络还具备强大的自适应能力, 能够根据新的数据资产进行持续学习和更新, 以适应数据特征的动态变化。在实际应用中, 通过使用测试数据集对训练好的 GA-BP 模型进行评估, 可以定量地衡量模型的预测精度和鲁棒性。因此本文的贡献在于: 本文基于 GA-BP 神经网络, 提出了一种全新的数据资产评估方法。该方法通过对拼多多大数据的深度挖掘, 建立了数据资产与企业价值之间的复杂非线性关系模型。模型能够自适应地学习和调整, 从而实现对数据资产价值的动态评估。相比传统评估方法, 该方法具有更高的准确性和可靠性, 能够更全面地反映数据资产的多维特征, 如用户规模、活跃度等, 为数据资产的价值评估提供了更科学、更量化的依据。该方法的应用将为电子商务行业的数据资产管理提供有力支持, 并为投资者、合作伙伴和监管部门提供更客观的决策参考。

2 文献综述

2.1 传统数据资产评估方法

在众多评估数据资产价值的方法中, 市场法的应用要求我们参考市场上同类或相似数据资产的最新交易价值。通过这种对比分析, 我们能够更精确地估算出数据资产的真实价值。数据资产价值评估的有效性, 依赖于一个开放且活跃的交易市场。尽管贵州、江苏等地区已经逐步建立了数据信息资产交易的网络平台, 并且在网络上出现了如 HexBin、东湖等数据资产交

易所，但这些交易所和平台的完善程度仍有待提高。目前，它们还不能提供全面和广泛的交易数据。正是这些局限，使得市场法在数据资产价值评估中的应用受到了一定的制约。刘枏（2021）[4]深入剖析了国内外的数据资产定价方法，发现市场应用受到了一定的约束。在收益法的应用中，我们可以区分为以下几种方法：权利金节省法、多期超额收益法。这些方法各有特点，共同构成了数据资产价值评估的重要手段。在无形资产的价值评估领域，权利金节省法被广泛采用。申海成与张腾（2019）[5]对此有深入研究。而郝雪镜[4]则提出了一种新颖的多期超额收益法，这种方法特别考虑了数据资产权属的模糊性和数据安全等风险因素，提出了一种区别于传统无形资产折现率的数据资产折现率模型。此外，戴宇欣（2017）[6]对品牌价值的评估进行了探索，建立了增量收益法、成本剩余法和市场比较法等多种方法，为不同类型的品牌提供了标准化的价值评估框架。在数据资产价值评估的领域中，成本法通过分析数据资产在其生命周期内各个阶段所产生的成本来评估其价值，本质上是对数据成本的系统梳理。这种方法认识到数据资产的成本与其市场价格之间的关联并不总是强烈的，且成本记录可能不尽完整。尽管如此，成本法在评估那些以成本分摊为目标的数据资产时，仍然具有一定的合理性。其优点在于计算过程的简洁性和易操作性。李永红与李金鹭（2017）[7]通过应用成本法对企业内部累积的数据资产进行评估，揭示了其成本主要包括初期准备和数据获取两大阶段，并指出当数据资产的规模过大或过小时，成本法可能无法准确反映其真实价值。

2.2 GA-BP 神经网络模型

遗传算法（GA）是一种受自然选择和进化理论启发的搜索算法，它通过模拟生物进化中的选择、交叉和变异过程，为各类搜索、优化和学习问题提供有效的解决方案。BP 神经网络，即反向传播神经网络，是一种多层前馈网络，通过最速下降法调整权重和阈值，以最小化误差平方和，从而学习输入至输出之间的映射关系。

Yang（2019）[8]利用改进的 BP 神经网络设计了物联网入侵检测系统，采用 LM-BP 算法优化了检测流程。Zhang 等（2019）[9]提出了 ATPSO-BP 模型，结合主动目标粒子群优化（ATPSO）算法和 BP 神经网络，增强了模型的权重和阈值寻优能力，提高了网络的泛

化能力和收敛速度。Jiang（2020）[10]研究了 BP 神经网络在建筑造价估算中的应用，建立了一个估计模型，并通过实验验证了其高精度和可靠性。为了克服 BP 神经网络可能陷入局部最优的问题，GA 被用来优化神经网络。Zhang 等人（2021）[11]将 GA-BP 神经网络应用于股价模式的分类和预测，证明了 GA-BP 模型相比传统 BP 网络具有更快的收敛速度和更高的预测准确性。墨蒙、赵龙章等（2018）[12]探讨了遗传算法和神经网络结合的方法，而马晓敏、王新等（2013）[13]则研究了遗传算法在优化神经网络结构、权重和学习规则方面的应用。刘春艳、凌建春等（2013）[14]指出 GA-BP 神经网络模型具有高稳定性和效率，能有效避免神经网络的局部最优问题。闵永军（2021）[15]通过构建 GA-BP 模型，对二手车价格进行了合理准确的评估。党雪宁[16]巧妙地构建了一个 GA-BP 神经网络模型，以精准评估数据资产的价值。该模型融合了遗传算法的全局搜索能力与反向传播神经网络的强大学习功能，全面考量并量化了数据资产价值的多重影响因素。在此基础上，作者搜集了诸如浏览量、访客数、IP 数等关键指标，并将它们纳入神经网络的输入层，为搜索引擎数据资产的价值预测提供了数据支撑。进一步地，通过遗传算法对神经网络的权重和阈值进行精细调优，党雪宁有效克服了 BP 神经网络在局部极值问题上的挑战，显著提升了模型预测的准确度。这一创新方法不仅展示了数据资产评估的新视角，也为未来的研究提供了宝贵的参考。

总体而言，GA-BP 神经网络模型结合了遗传算法的全局搜索能力和 BP 神经网络的强大学习能力，为电子商务企业数据资产价值评估提供了一套科学且合理的评估体系，为该领域的研究和实践提供了宝贵的参考。

2.3 文献综述

在数据资产价值评估领域，市场法、收益法和成本法是三种主要的评估方法。市场法依赖于活跃的交易市场来参考同类或相似数据资产的交易价值，但由于交易平台的不完善，这种方法的应用受到限制。收益法包括权利金节省法和多期超额收益法，后者特别考虑了数据资产权属的模糊性和数据安全风险。成本法则通过分析数据资产生命周期内的成本来评估其价值，尽管它的简洁性和易操作性是优点，但当数据资产规模过大或过小时，可能无法准确反映其真实价值。

通过梳理文献过程中发现 GA-BP 神经网络模型是一种结合了遗传算法 (GA) 和 BP 神经网络的方法。该遗传算法模拟自然选择和进化理论中的选择、交叉和变异过程, 而 BP 神经网络通过最速下降法调整权重和阈值, 以最小化误差平方和。这种模型已被应用于多个领域, 如物联网入侵检测、建筑造价估算和股价模式分类与预测, 显示出其高稳定性和效率。特别是在数据资产价值评估中, GA-BP 模型通过全面考量并量化数据资产价值的多重影响因素, 提高了预测的准确度。

GA-BP 神经网络模型为电子商务企业的数据资产价值评估提供了一种科学而合理的新体系。该模型融合了遗传算法的全局优化能力与 BP 神经网络的高效学习机制, 为数据资产评估领域贡献了重要的理论与实践指导。在处理复杂性高和不确定性强的评估问题时, GA-BP 神经网络展现了其无可比拟的优势, 标志着数据资产评估方法的一大进步。作为一种强有力的工具, GA-BP 神经网络能够精准处理复杂数据集, 学习经验, 并灵活适应环境变化, 使其成为评估数据资产动态价值的理想选择。随着对该模型研究的深入, 其在数据资产价值评估中的应用将愈发显著, 为传统评估方法带来革新。

3 GA-BP 神经网络模型与构建流程

3.1 GA-BP 神经网络模型

GA-BP 神经网络 (Genetic Algorithm-Back propagation Neural Network, 即遗传算法—反向传播神经网络), 是一种将遗传算法与反向传播神经网络相结合的机器学习模型, 它结合了遗传算法的全局搜索能力和反向传播神经网络的强大学习能力, 具有以下特点:

①非线性映射能力强: GA-BP 神经网络可以学习数据模型, 包括非线性参数, 这有利于评估数据资产价值, 因为数据资产价值通常受到多种因素的影响, 这些因素对于非线性系统而言也同样重要。

②鲁棒性强: GA-BP 神经网络具有较强的鲁棒性, 能够抵抗数据噪声和异常值的影响, 这使其非常适合用于评估真实世界数据资产, 因为真实世界数据往往存在噪声和异常值。

③GA-BP 神经网络具有较强的自适应能力: GA-BP 神经网络具有较强的自适应能力, 能够随着新

数据的到来不断学习和改进, 这使得它能够用于评估我们的数据资产。

具体来说, GA-BP 神经网络可以用于评估数据资产价值, 主要体现在以下几个方面:

①数据资产特征提取: GA-BP 神经网络导入了众多的数据, 并采用这些特性来评估数据资产的价值。GA-BP 神经网络整合了众多的数据, 并采用这些特性来描述用户的停留时长、转换率等属性, 这些属性有助于揭示网站的实际价值。

②数据资产价值预测: GA-BP 神经网络可以基于数据资产特征预测数据资产价值。例如, GA-BP 神经网络可以基于网站流量数据获得预测的代码广告收入。

③数据资产价值评估: GA-BP 神经网络将数据资产价值与市场交易价值进行比较, 以评估数据资产价值是否合理。

总的来说, GA-BP 神经网络是一种用于评估数据资产的机器学习模型。它具有非线性映射能力强、鲁棒性强、自适应能力强等特点, 可以用于数据资产特征提取、数据资产价值预测、数据资产价值评估等方面。

3.2 GA-BP 神经网络评估构建

在数据驱动时代, 精准数据资产估值能力已成为组织做出明智决策和获取竞争优势的关键。然而, 传统估值方法难以应对数据资产的复杂性和动态性, 亟需更为有效的解决方案。GA-BP 神经网络, 融合遗传算法 (GA) 和反向传播神经网络 (BP) 的强大优势, 为数据资产估值领域带来了前所未有的机遇。下面是构建 GA-BP 神经网络的流程。

①数据采集和预处理: 奠定基础

构建有效的 GA-BP 神经网络进行数据资产评估的旅程始于对相关数据的细致收集。这包括定量和定性信息, 全面描述正在评估的数据资产。定量数据可能包括数据量、使用情况统计数据 and 数据质量指标等指标。另一方面, 定性数据可能包括数据的战略重要性、其创收潜力以及其对监管要求的遵守情况等因素。

收集数据后, 需要对其进行预处理, 以确保其适合训练神经网络。这一关键步骤包括清理数据以消除错误和不一致之处、将数据标准化为通用比例以及将分类变量编码为数值表示。通过精心预处理数据, 我们为神经网络有效学习和做出准确预测奠定了坚实的基础。

②特征工程：提取重要指标

特征工程在将原始数据转换为神经网络可以有效掌握的有意义的特征方面起着关键作用。此过程通常需要领域专业知识来识别与数据资产评估最相关且信息量最大的特征。常见的特征工程技术包括特征选择、特征转换和特征创建。

特征选择涉及从原始数据集中明智地选择相关特征子集，以降低维度并提高模型性能。特征变换涉及将特征转换为更合适的表示形式，例如缩放、规范化和离散化。另一方面，特征创建涉及从现有特征中生成新特征，例如计算特征的比率、差异或组合。通过精心设计特征，我们使神经网络能够提取数据的精髓并做出明智的估值决策。

③网络架构设计：制定神经网络结构

GA-BP 神经网络的架构决定了其学习复杂模式和做出准确预测的能力。该网络通常由三层组成：输入层、隐藏层和输出层。

输入层接收从数据中提取的预处理和工程特征。输入神经元的数量与特征的数量相对应。隐藏层是网络的主力，负责学习输入特征和目标输出之间的复杂关系。隐藏层的数量和每层神经元的数量会影响网络的容量和复杂性。最后，输出层产生预测的数据资产值。输出神经元的数量与目标值的数量相对应（例如，一个神经元用于预测单个值）。

通过精心设计网络架构，我们制定了一个模型框架，使神经网络能够有效地学习数据资产估值的复杂细节。

④基于 GA 的超参数优化：优化至优

超参数是控制神经网络学习过程的配置设置。超参数的示例包括学习率、激活函数和权重正则化系数。选择最佳超参数对于实现良好性能至关重要。

GA（遗传算法）可用于优化 BP 神经网络的超参数。GA 模仿自然选择的过程来进化出一批潜在解决方案，并根据适应度函数迭代地提高其性能。通过利用基于 GA 的超参数优化，我们可以将神经网络微调至最佳性能，确保数据资产估值准确可靠。

⑤网络训练与评估：改进与验证

一旦确定了网络架构和超参数，GA-BP 神经网络就会进入训练阶段。训练过程包括迭代调整网络连接的权重和偏差，以最小化预测数据资产值与实际数据资产值之间的误差。

在训练期间，我们会在单独的验证数据集上对网络进行评估，以评估其泛化能力并防止过度拟合。当

网络对训练数据的记忆过好时，就会发生过度拟合，导致在未见过的数据上表现不佳。通过仔细评估网络的性能，我们可以提高其泛化能力，并对新数据资产做出准确的预测。

GA-BP 神经网络，为数据资产估值领域提供了强大的技术支撑，具有广阔的应用前景。随着研究的不断深入，GA-BP 神经网络将发挥更大的作用，助力企业释放数据资产的巨大潜力。

4 拼多多数据资产案例分析

4.1 拼多多发展概况

2015 年 9 月，拼多多由上海寻梦信息技术有限公司孵化而成立，创始人黄峥为其注入了生机。黄峥曾供职于谷歌和亚马逊中国，拥有丰富的互联网经验和敏锐的市场洞察力。他立足于中国电商市场发展现状，洞察到沉着的市场用户对价格敏感障碍，并由此创立了拼多多。拼多多以“团购”为核心，通过用户分享、多人协作的方式，实现低价位购买。这种创新的模式迅速获得大量价格型消费者的青睐，特别是下沉的市场用户。而拼多多的经营理念为“本分”，意味着企业坚守本职，全心全意为用户创造更大的价值。在 2018 年 7 月，拼多多在纳斯达克成功上市，首发价格为 19 美元，市值一举冲至 240 亿美元，迅速成为资本市场的宠儿。

在用户规模、商家生态和货品供给这三个关键领域，拼多多展现出了无可匹敌的竞争力。用户群体的高度忠诚和活跃，商家的丰富多样和稳定性，以及供给侧的丰富性和超高性价比，共同巩固了拼多多在中国电商市场的领先地位。相较于其他电商平台，拼多多的用户更加看重价格和性价比，虽然品牌和质量也很重要，但价格更是购买的关键因素。拼多多的飞速发展，部分归功于用户在社交网络上的拼团和分享行为，这种社交裂变效应为平台带来了持续的发展动力。

4.2 拼多多数据资产价值评估因素

拼多多的数据资产价值受众多因素影响，其中包括用户群体的规模、市场的需求、技术实力以及用户体验。为了深入评估这些因素如何塑造拼多多的数据价值，我们进行了广泛的数据收集工作。研究发现，用户行为对这些因素起着至关重要的作用。因此，本

研究将聚焦于用户视角，以预测搜索引擎的价值，并以流量相关指标作为评估基准。我们选取了七个关键指标：浏览量、访客数、月总交易额、活跃用户数、IP 地址数量、复购率和客单价，来探讨它们如何影响拼多多的数据资产价值。

4.3 拼多多样本数据来源及处理

本文以浏览量、月均访客人数、月均交易额、活跃用户数、IP 数、复购率、客单价这七个指标作为拼多多数据资产价值评估的核心影响因素。为了消除量纲差异，提高模型训练效率，我们将这七个指标进行了归一化处理，并将处理后的数据作为神经网络的输入层。同时，我们搜集了拼多多过去三年每月末的股票单价，并结合股票总市值，计算得出股票的内在价值作为预测目标。将此预测值作为神经网络输出层的训练样本，建立了如下价值预测模型：

$$V=C\times (S_{max}+S_{min})/2$$

- 其中：
- V：拼多多市场价值；
- C：拼多多预测当日股数；
- S_{max}：百度日最高股价；
- S_{min}：百度日最低股价；

本文从东方财富网站搜集拼多多 2021 年—2023 年每月的相关数据进行后续实验。

5 神经网络结构搭建

5.1 传递函数

在神经网络设计中，传递函数是定义网络行为的关键因素，主要有三种类型：tansig（双曲正切 S 型传递函数）、logsig（对数 S 型传递函数）和 purelin（线性传递函数）。在本研究中，我们选择了 tansig 和 purelin 的组合作为神经网络的传递函数，以期达到最佳的网络性能。

5.2 神经网络层数

在神经网络的构建过程中，精心选择合适数量的隐藏层是至关重要的，因为它直接影响着模型的性能。隐藏层的数量如果过多，不仅会增加训练时间，还可能引起模型的过拟合现象；而隐藏层数量不足，则可能导致模型欠拟合。因此，在追求模型精度的同时，我们必须避免过度增加隐藏层的数量。在本研究中，我们首先利用遗传算法（GA）来自动优化网络的层数。随后，借鉴了党雪宁错误!未找到引用源。的研究，我们选择了一个相同的网络结构,进行了 50 次迭代训练。我们比较了网络的运行时间和结果精度，并采用了双层隐藏层的网络结构。

5.3 各层神经元节点数

在本研究中，我们选取了以下指标作为神经网络的输入因素：浏览量、访客数、月总交易额、活跃用户数、IP 地址数量、复购率和客单价，共计 7 个输入层神经元节点。作为网络的输出，我们定义了拼多多数据资产市值这一被解释变量，其在网络中由单一节点表示。在广泛研究相关文献并进行多轮测试之后，我们确定了网络结构：第一隐藏层包含 13 个神经元，第二隐藏层则包含 1 个神经元。此配置的目的是为了优化网络结构，从而提升模型的预测精度。

5.4 学习率

我们设定神经网络的初始学习率为 0.001，并采用梯度下降法在训练过程中逐步优化。通过不断调整，直到模型收敛并达到局部最优解，我们最终确定了适宜的学习率。这一过程确保了学习率能够有效地促进网络的学习与进步。为了提高数据分析的准确性，我们对原始数据进行了标准化处理。具体来说，我们将浏览量和总交易额缩减了 100 倍，并对市值应用了对数转换。经过这样的处理，数据之间的差异性得到了有效缩小，从而使得分析结果更为精确。以下表 1 是处理后的数据结果展示。

表 1 处理后的样本数据表

日期	浏览量	访客数	总交易额	活跃用户数	IP 数	复购率	客单价	市值
2021-1	1.835	5.5	1.457	7.31	4.9	0.77	4.15	3.46709138
2021-2	1.892	5.6	1.523	7.42	5	0.78	4.17	3.497859481
2021-3	2.024	5.7	1.601	7.53	5.1	0.79	4.19	3.409792842
2021-4	2.053	5.8	1.68	7.64	5.2	0.80	4.21	3.418548205
2021-5	2.102	5.9	1.76	7.75	5.3	0.81	4.23	3.377335397
2021-6	2.151	6	1.841	7.86	5.4	0.82	4.25	3.394276527

日期	浏览量	访客数	总交易额	活跃用户数	IP 数	复购率	客单价	市值
2021-7	2.18	6.1	1.922	7.97	5.5	0.83	4.27	3.292444373
2021-8	2.209	6.2	2.004	8.08	5.6	0.84	4.29	3.223270039
2021-9	2.238	6.3	2.086	8.19	5.7	0.85	4.31	3.273244409
2021-10	2.267	6.4	2.169	8.3	5.8	0.86	4.33	3.250851642
2021-11	2.296	6.5	2.252	8.41	5.9	0.87	4.35	3.18364844
2021-12	2.209	6.6	2.336	8.52	6	0.88	4.37	3.04383758
2022-1	2.322	6.7	1.667	7.88	7.2	0.77	4.17	3.052135809
2022-2	2.379	6.8	1.703	7.92	7.3	0.78	4.15	3.028924888
2022-3	2.436	6.9	1.812	8.08	7.16	0.79	4.19	3.002312962
2022-4	2.465	7	1.928	8.24	7.4	0.80	4.21	3.012153622
2022-5	2.514	7.1	2.046	8.38	7.5	0.81	4.23	3.050499625
2022-6	2.563	7.2	2.167	8.46	7.38	0.82	4.25	3.046346862
2022-7	2.592	7.3	2.289	8.52	7.6	0.83	4.27	3.041226832
2022-8	2.621	7.4	2.412	8.6	7.7	0.84	4.29	3.051700353
2022-9	2.65	7.5	2.536	8.68	7.42	0.85	4.31	3.103423975
2022-10	2.679	7.6	2.661	8.76	7.9	0.86	4.33	3.015007482
2022-11	2.708	7.7	2.787	8.84	7.6	0.87	4.35	3.120350146
2022-12	2.725	7.8	2.914	8.92	6.8	0.88	4.37	3.22142728
2023-1	3.4	6.8	2.914	8.68	6.9	0.69	5.08	3.269150881
2023-2	3.305	6.6	3.028	8.76	6.4	0.70	5.12	3.260312331
2023-3	3.51	7.31	3.143	8.84	7.62	0.71	5.16	3.213693033
2023-4	3.38	7.32	2.719	7.88	5.8	0.72	5.2	3.129128942
2023-5	3.32	7.38	2.822	7.96	5.7	0.72	5.24	3.108118323
2023-6	3.48	7.42	2.925	8.05	6.3	0.72	5.28	3.155767081
2023-7	3.35	6.8	3.028	8.14	6.5	0.72	5.32	3.190062632
2023-8	3.3	6.7	3.131	8.23	6.6	0.72	5.36	3.236000359
2023-9	3.42	7.31	2.947	7.31	7.26	0.72	5.4	3.279728526
2023-10	3.37	7.42	2.96	7.35	7.26	0.73	5.44	3.312050351
2023-11	3.34	7.5	3.007	7.38	7.3	0.73	5.48	3.381905301
2023-12	3.69	7.58	3.112	7.42	7.38	0.73	5.52	3.449148117

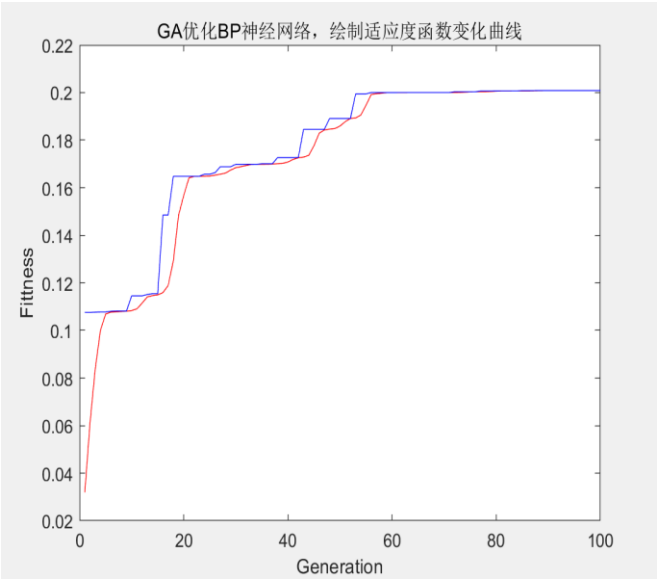


图 1 适应度曲线

当数据输入到 GA-BP 模型后，模型自动启动，寻求最优参数。训练过程中，模型在第七轮达到顶峰性能，此时的验证集均方误差（MSE）降至 0.018971。图 1 展示了迭代过程：x 轴标注迭代次数，y 轴显示适

应度值。结果显示，随着迭代次数增加，适应度值稳步提升，这反映了 BP 神经网络对特定数据集的精准适应。图 2 揭示了遗传算法（GA）对 BP 神经网络性能的显著提升，证明了 GA 在寻找更优解决方案方面的有效性。优化后，模型的最佳适应度值和平均适应度值均达到新高。图 3 则呈现了均方误差（MSE）的变化趋势，尽管第 13 轮后测试集的 MSE 略有上升，但整体趋势稳定，显示出模型的良好泛化能力。综合分析这些图表 GA 算法成功优化了 BP 神经网络，极大地增强了其性能。

图 4 数据揭示了 GA-BP 神经网络在数据资产评估方面的卓越性能，训练集、验证集和测试集的拟合度分别达到了 0.99781、0.99345 和 0.9732，整体拟合度为 0.99169。这些高度精确的数字强力证明了该神经网络的优越性。经过深入训练，我们得到了如下表所示的预测值与真实值之间的误差，以及优化前后误差的对比，凸显了优化过程带来的显著提升。这些成果不仅验证了 GA-BP 神经网络的高精度，也证明了其在评估数据资产时的高效性和可靠性。在图 5 中，比较了 BP 和 GA-BP 模型的预测值与真实值。可以看出从图 5

中可以观察到，GA-BP 模型的预测值与真实值之间的偏差较小，相对于 BP 模型来说更准确。从图 6 中可以看出，GA-BP 模型的预测误差明显低于 BP 模型，这

表明 GA-BP 模型更适合这个特定的问题，具体情况如下图 6 所示。

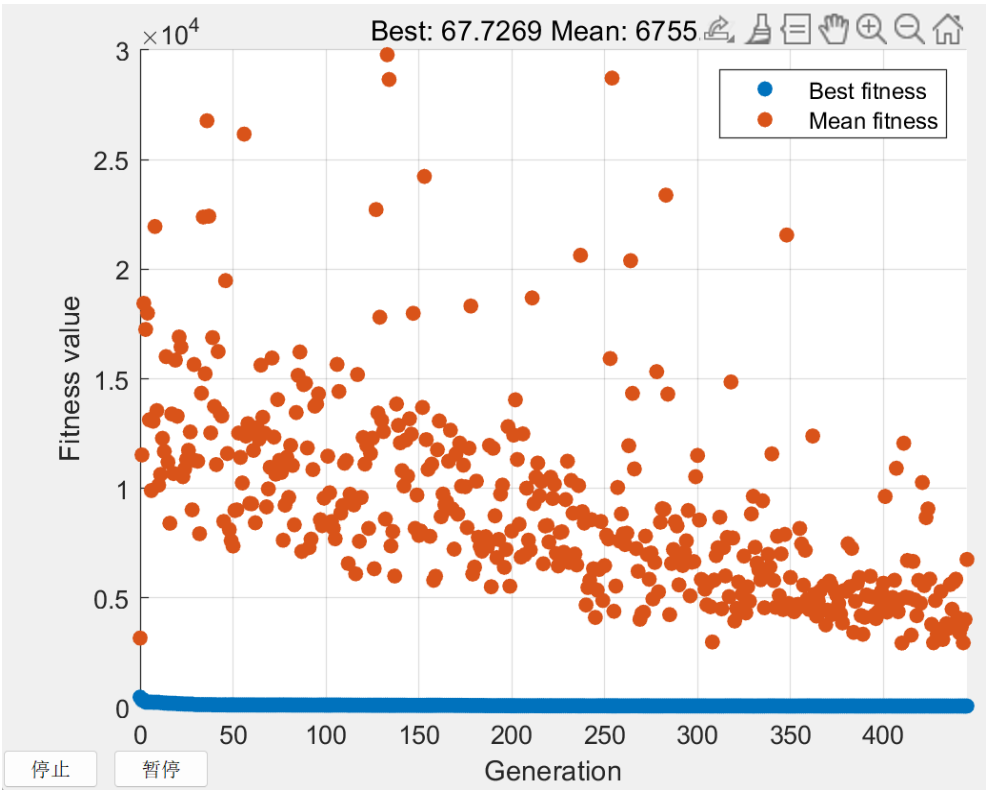


图 2 进化过程中的个体适应值变化图

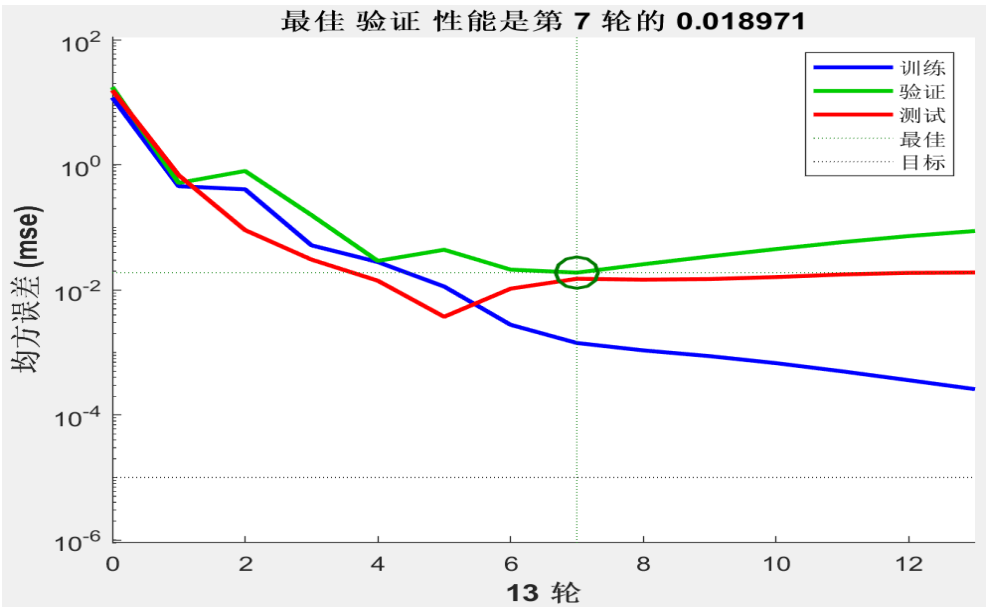


图 3 均方误差（MSE）曲线图

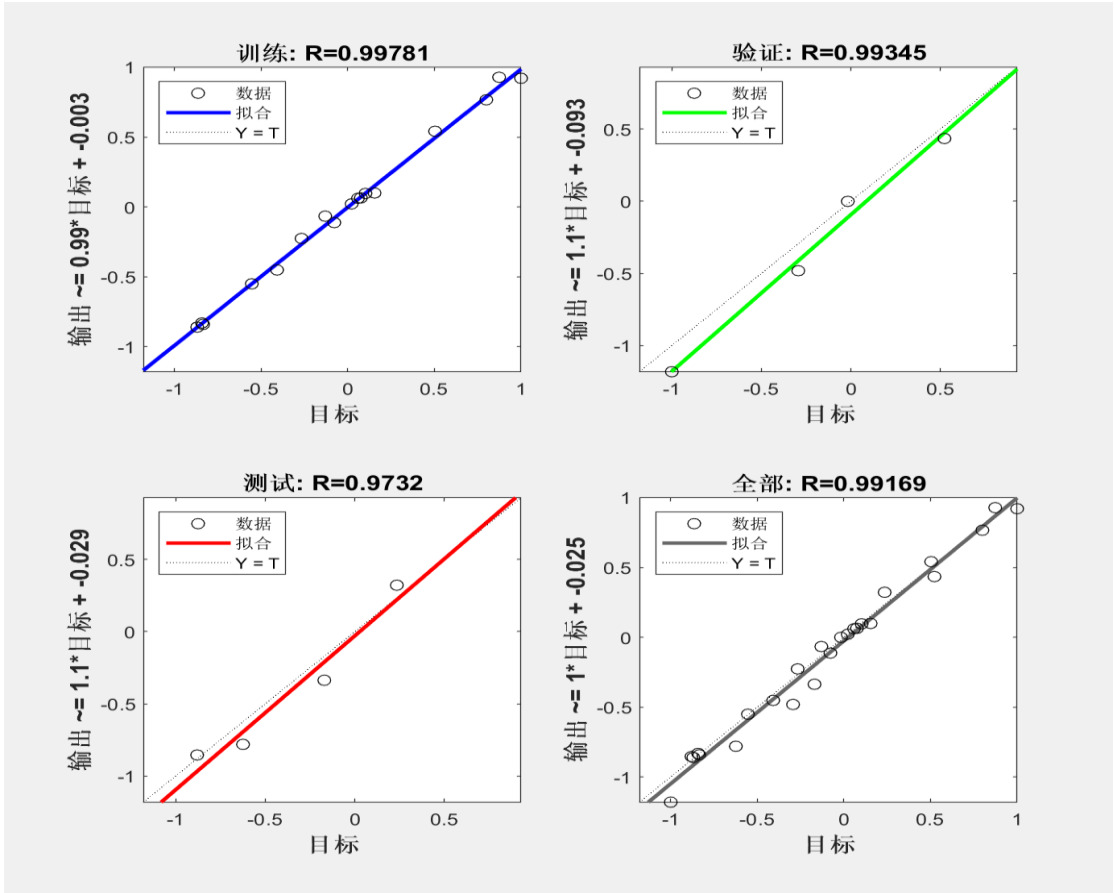


图 4 模型拟合图

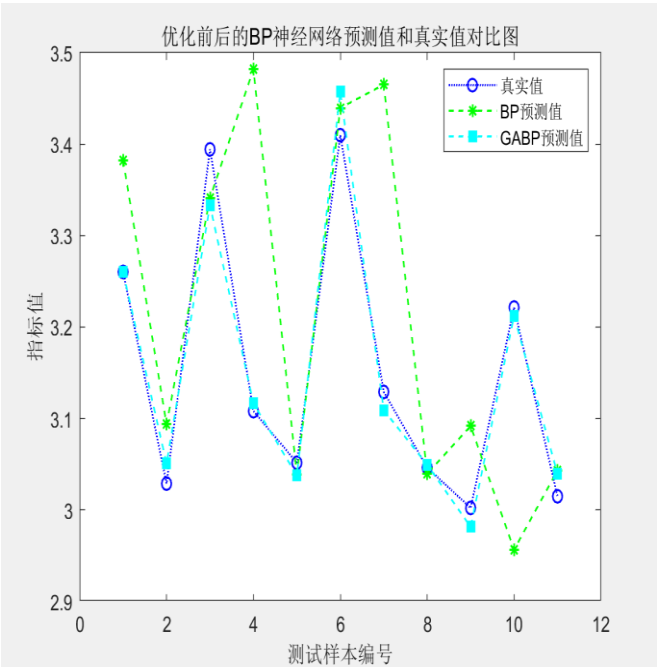


图 5 BP 与 GA-BP 预测值与真实值对比图

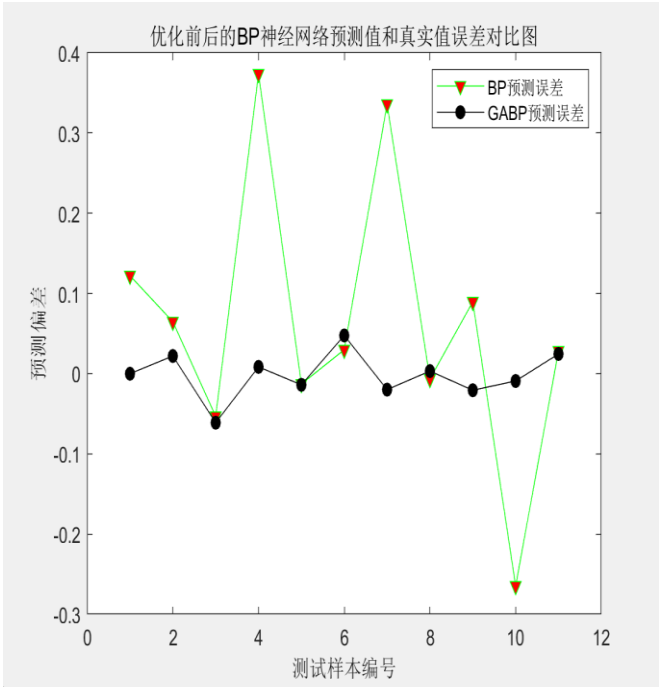


图 6 BP 与 GA-BP 预测误差图

6 建立 GA-BP 神经网络体系

6.1 建模流程

1、声明全局变量

(1)训练集与测试集的确定

我们从样本数据中让模型随机挑选了前 29 组作为训练集的输入数据，构成了一个 7 行 29 列的矩阵(p)。同时，为了预测百度搜索引擎的市值，我们设定了一个 1 行 29 列的矩阵(t)作为训练集的输出数据。

(2)输入层、隐藏层、输出层个数确定

输入层神经元个数为 R=7，隐藏层神经元个数为 S1=13，输出层神经元个数 S2=1。

(3)编码长度

$S=R \times S1+S1 \times S2+S1+S2=7 \times 13+13 \times 1+13+1=118$

2、数据归一化

在 MATLABr2023b 软件对数据进行了归一化处理。具体命令如下：`[p, minp, maxp, t, mint, maxt] = premnmx(p, t)`

其中，minp 和 maxp 分别代表训练集输入数据的最小值和最大值；mint 和 maxt 则分别代表训练集输出数据的最小值和最大值。

3、建立模型

本文使用“tansig”，“purelin”函数创建 GA-BP 神经网络，模型如下：

`net=newff(minmax(p1), [S1, 1], {'tansig', 'purelin'}, 'trainlm')`

4、然后设置训练参数：

(1)设置训练迭代过程

`net.trainParam.show=100;`

(2)最大收敛次数

`net.trainParam.epochs=2000;`

(3)收敛误差

`net.trainParam.goal=1.0e-13;`

(4)学习速率

`net.trainParam.lr=0.001`

5、确定种群个数并初始化种群

设定种群个数，在 MATLABR2023b 中输入 `pop_num=50`，初始化过程如下：

`Prec=[1e-6, 1]%epslin 为 1e-6，实数编码`

`normGeomSelect=0.09%`选择函数的参数

`arithXover=2%`交叉函数的参数

`nonUnifMutation=[2gen3]%`变异函数的参数

`initPpp=initializega(pop_num, bounds, 'gabpEval', [], [1e-6])`

6、优化算法

`[x, endPop, bPop, trace]=ga(bounds, 'gabpEval', [], initPpp, [1e-6], 'maxGenTerm', gen, 'normGeomSelect', [0.09], 'arithXover', [2], 'nonUnifMutation', [2gen3])`

7、获取最优参数

`[W1, B1, W2, B2, val]=gadecod(x)`

8、参数赋值以及训练网络

在 matlab 中输入以下代码，对模型中的参数进行赋值，进而对模型进行训练：

`net.IW{1, 1}=W1;`

`net.LW{2, 1}=W2;`

`net.b{1}=B1;`

`net.b{2}=B2;`

`net=train(net, p1, t1)`

$$W1 = \begin{bmatrix} -0.2154 & -0.8466 & -0.7595 & 0.0567 & 0.0637 & -0.2114 & 0.4890 \\ 0.7515 & -0.4264 & -0.3471 & -0.0507 & -0.4531 & -0.3381 & 0.6014 \\ 0.5814 & -0.3521 & 0.5728 & 0.5559 & 0.3642 & 0.5299 & -0.3146 \\ 0.1099 & 0.2522 & 0.5867 & -0.7393 & 0.1435 & 0.1746 & 0.1765 \\ -0.5431 & 0.0535 & 0.0301 & 0.4927 & -0.0770 & -0.0252 & 0.7234 \\ 0.0729 & 0.2002 & -0.9030 & 0.3091 & -0.8195 & 0.1332 & 0.0594 \\ 0.1801 & -0.1194 & -0.5920 & -0.1202 & 0.1268 & 0.1092 & 0.1343 \\ 0.7277 & -0.3927 & 0.9793 & -0.8045 & 0.1365 & -0.3511 & -0.2033 \\ 0.0290 & -0.5744 & 0.4449 & -0.2268 & -0.2779 & -0.3773 & -0.1657 \\ 0.6485 & 0.6624 & 0.0845 & 0.2273 & 0.5166 & 0.4867 & 0.4391 \\ -0.5937 & -0.1060 & -0.2677 & 0.0692 & 0.5600 & -0.2338 & 0.5535 \\ 0.5823 & 0.2102 & 0.2990 & -0.7129 & 0.5659 & 0.9442 & 0.1656 \\ 0.0049 & -0.6070 & -0.3841 & 0.4810 & 0.2678 & -0.9256 & 0.0306 \end{bmatrix}$$

`W2=[0.2309 0.1393 0.4905 0.6708 0.5586 0.0844 0.1336 0.1119 0.0048 0.0201 0.3502 -0.3368 -0.4953]`

`B1=[0.4140 0.6973 0.8937 -0.1384 0.1804 0.0772 0.6200 0.0715 -0.2585 0.1587 0.4172 -0.0022 -0.3813]`

`B2=0.0214`

6.2 结果分析

经过上述过程，在 MATLABR2023b 软件中，将输出的数据进行反归一化，将反归一化的数据看作预测

值，同时测试数据作为真实值，比较两者之间的差异。
c=postmnmx(s_ga, mint2, maxt2)
得到的情况如下表 2 所示：

表 2 预测值与真实值比较

日期	真实值	预测值	误差	误差百分比
2021-3	3.224766074	3.011696862	0.213069213	0.066072765
2021-4	3.224929108	3.244664157	-0.01973505	-0.00611953
2021-5	3.194544866	3.21521291	-0.020668044	-0.006469793
2021-6	3.201923967	3.139985369	0.061938598	0.019344181
2021-7	3.059994848	3.0744715	-0.014476653	-0.00473094
2021-8	3.098138197	3.014805244	0.083332952	0.026897752
2021-9	3.055515938	3.228737842	-0.173221903	-0.05669154
2021-10	3.047048898	3.035036724	0.012012174	0.003942232
2021-11	2.92087434	2.893266156	0.027608184	0.009452027
2021-12	2.867549978	3.017295126	-0.149745147	-0.052220588
2022-1	2.850946279	2.991909345	-0.140963066	-0.049444308
2022-2	2.82610073	2.702373556	0.123727174	0.043780171
2022-3	2.826107212	2.977191673	-0.151084461	-0.053460272
2022-4	2.82610073	3.082752201	-0.256651471	-0.090814693
2022-5	2.803880278	2.931690367	-0.127810089	-0.04558329
2022-6	2.892867849	2.980586242	-0.087718393	-0.030322295
2022-7	2.79216748	3.011019499	-0.21885202	-0.078380692
2022-8	2.954970548	2.847062315	0.107908233	0.036517533
2022-9	2.898313723	2.873510264	0.024803459	0.008557893
2022-10	2.840896143	2.769045895	0.071850248	0.025291402
2022-11	3.015904378	3.095330379	-0.079426001	-0.026335716
2022-12	3.013304999	3.20386807	-0.190563071	-0.063240552
2023-1	3.093018725	2.848327389	0.244691336	0.079110848
2023-2	3.045029437	2.935213291	0.109816146	0.036064067
2023-3	2.982121919	3.012074615	-0.029952696	-0.010044088
2023-4	2.95685488	3.064311203	-0.107456323	-0.036341426
2023-5	2.938434659	3.095282672	-0.156848012	-0.053378084
2023-6	2.963121713	3.122831641	-0.159709928	-0.053899213
2023-7	3.076764198	2.891948184	0.184816014	0.060068306
2023-8	3.118892725	2.993251934	0.125640791	0.040283781
2023-9	3.11492775	3.211599542	-0.096671791	-0.031035003
2023-10	3.129512659	3.175859675	-0.046347017	-0.014809659
2023-11	3.292005615	3.005157993	0.286847622	0.087134609
2023-12	3.288664983	3.275785741	0.012879242	0.003916253

通过对表格中的预测值和实际值进行比较分析，可以看出 GA-BP 神经网络模型的误差率大多数情况下控制在 10%以内。这一成果有效证明了该模型不仅拥有出色的拟合能力，而且还具备稳定的预测性能。因此，该模型完全适用于对拼多多数据资产价值进行准确预测的任务。

表 3 数据处理表

预测值（对数）	预测值
0.029493944	1691.4568
0.066072765	1027.2990
-0.00611953	1756.5647
-0.006469793	1641.3943
0.019344181	1380.3378

预测值（对数）	预测值
-0.00473094	1187.0568
0.026897752	1034.6781
-0.05669154	1693.3153
0.003942232	1084.0186
0.009452027	782.1070
-0.052220588	1040.6271
-0.049444308	981.5430
0.043780171	503.9339
-0.053460272	948.8371
-0.090814693	1209.9076
-0.04558329	854.4573
-0.030322295	956.2826
-0.078380692	1025.6980
0.036517533	703.1732
0.008557893	747.3263
0.025291402	587.5514
-0.026335716	1245.4617

预测值（对数）	预测值
-0.063240552	1599.0722
0.079110848	705.2245
0.036064067	861.4167
-0.010044088	1028.1929
-0.036341426	1159.6080
-0.053378084	1245.3249
-0.053899213	1326.8800
0.060068306	779.7371
0.040283781	984.5821
-0.031035003	1627.7944
-0.014809659	1499.2004
0.087134609	1011.9475
0.003916253	1887.0601

在 MATLABR2023b 软件中，根据预设的参数完成了模型的搭建与训练。通过遗传算法优化的神经网络所预测的结果与实际数据的对比效果，如图 7 中展示。

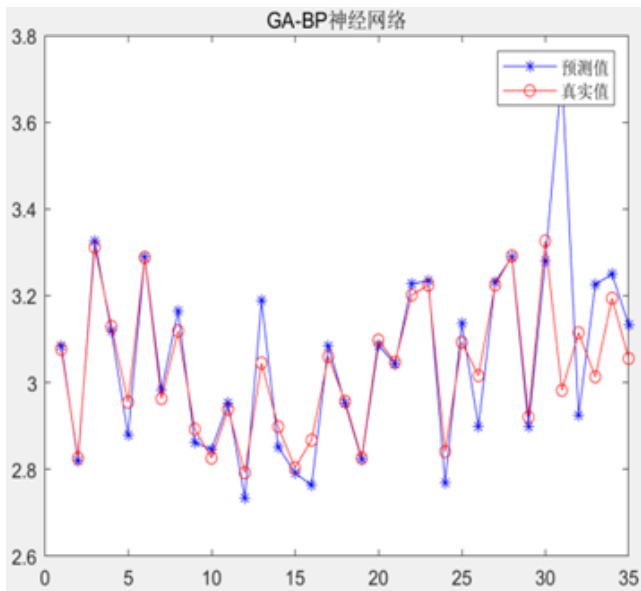


图 7 GA-BP 神经网络拟合图

图 7 展现了两种模型在预测结果上的整体趋势。本研究采用 GA-BP 神经网络模型对拼多多数据资产价值进行了评估。鉴于拼多多企业在市场份额和财务状况上的稳定性，我们将这 35 组数据的平均值视为搜索引擎数据资产预测价值的可靠估计。据此，百度搜索引擎数据资产的估值被确定为 1137.12 亿元。通过对比预测值与真实值的平均差异，我们发现两者之间的差距很小，这表明模型在预测时与实际数据之间有良好的拟合度。因此，我们得出的结论是基于模型的预测结果是真实和可信的，为相关领域提供了有价值的参考。

经过一系列详尽的分析，我们利用 35 组拼多多数据通过 GA-BP 神经网络模型得出了预测值，并与拼多多的实际数据进行了对比。结果显示，预测误差非常

小。鉴于拼多多企业目前的稳定发展态势，认为可以采用这 35 组数据的平均值，作为拼多多数据资产价值预测的可靠依据。具体计算过程如下：

$$V = (1691.4568 + 1027.2990 + 1756.5647 + 1641.3943 + 1380.3378 + 1187.0568 + 1034.6781 + 1693.3153 + 1084.0186 + 782.1070 + 1040.6271 + 981.5430 + 503.9339 + 948.8371 + 1209.9076 + 854.4573 + 956.2826 + 1025.6980 + 703.1732 + 747.3263 + 587.5514 + 1245.4617 + 1599.0722 + 705.2245 + 861.4167 + 1028.1929 + 1159.6080 + 1245.3249 + 1326.8800 + 779.7371 + 984.5821 + 1627.7944 + 1499.2004 + 1011.9475 + 1887.0601) / 35 = 1137.12 \text{ (亿元)}$$

7 研究结论

本文采用了一种结合遗传算法（GA）优化反向传播（BP）神经网络的混合模型——GA-BP 神经网络，用于对拼多多数据资产进行价值评估。该模型在处理复杂非线性问题时表现出卓越的拟合能力和泛化性能。通过模型评估，得出拼多多数据资产的价值约为 1137.12 亿元，与市场普遍估值范围相一致，充分验证了 GA-BP 模型在数据资产评估领域的有效性。

深入分析发现，GA-BP 模型能够将数据资产价值量化为可衡量的指标。研究结果表明，数据资产价值与平台的活跃用户数、交易额等核心指标呈现显著正相关关系。此外，消费者行为，如浏览频率、复购率等，对数据资产价值也产生了重要影响。然而，模型在样本选择、网络结构等方面仍有优化空间。未来，可以通过引入更多维度的数据、调整模型参数，以及探索更先进的深度学习算法来进一步提升模型的预测精度，为企业提供更准确、全面的数据资产评估。

参考文献

- [1] 张迎. 大数据时代下的情报学思考 [J]. 商情, 2018, 000(010): 114.
- [2] 李永红, 张淑雯. 数据资产价值评估模型构建 [J]. 财会月刊, 2018, (09): 30-35.
- [3] 李泽红, 檀晓云. 大数据资产会计确认、计量与报告 [J]. 财会通讯, 2018, (10): 58-59+129.
- [4] 刘枏, 郝雪镜, 陈俞宏. 大数据定价方法的国内外研究综述及对比分析 [J]. 大数据, 2021, 7(06): 89-102.
- [5] 申海成, 张腾. 知识产权评估的驱动因素、存在问题及对策 [J]. 会计之友, 2019, (02): 126-130.

- [6] 戴宇欣, 袁梦. 品牌价值评估方法标准化的探讨 [J]. 标准科学, 2017, (11): 102-105.
- [7] 李永红, 李金骛. 互联网企业数据资产价值评估方法研究 [J]. 经济研究导刊, 2017, (14): 104-107.
- [8] Yang A, Zhuansun Y, Liu C, et al. Design of Intrusion Detection System for Internet of Things Based on Improved BP Neural Network [J]. IEEE Access, 2019, 7: 106043-106052.
- [9] Zhang PFF. An ATPSO-BP neural network modeling and its application in mechanical property prediction [J]. Computational Materials Science, 2019, 163.
- [10] Jiang Q. Estimation of construction project building cost by back-propagation neural network [J]. Journal of Engineering, Design and Technology, 2020, 18(3): 601-609.
- [11] Zhang D, Lou S. The application research of neural network and BP algorithm in stock price pattern classification and prediction – Science Direct [J]. Future Generation Computer Systems, 2021, 115: 872-879.
- [12] 墨蒙, 赵龙章, 龚媛雯, 等. 基于遗传算法优化的BP神经网络研究应用 [J]. 现代电子技术, 2018, 41(09): 41-44.
- [13] 马晓敏, 王新. 基于遗传算法的BP神经网络改进 [J]. 云南大学学报 (自然科学版), 2013, 35(S2): 34-38.
- [14] 刘春艳, 凌建春, 寇林元, 等. GA-BP神经网络与BP神经网络性能比较 [J]. 中国卫生统计, 2013, 30(02): 173-176+181.
- [15] 李放, 闵永军. 遗传算法优化的BP神经网络在二手车价格评估的应用 [J]. 软件, 2021, 42(08): 91-94.
- [16] 党雪宁, 李明. 基于 GA-BP 神经网络的搜索引擎数据资产价值研究 [J]. 商展经济, 2023, (02): 126-129.