E-Commerce Letters 电子商务评论, 2026, 15(11), 487-494
Published Online November 2026 in Hans. https://www.hanspub.org/journal/ecl
https://doi.org/10.12677/ecl.2026.15111264

Hans汉斯

基于深度学习的电商用户评论情感分类研究

张维薇[1]，曹稀蓉[1]，程　实[2*]

[1]南通大学杏林学院，江苏　南通
[2]南通大学人工智能与计算机学院，江苏　南通

收稿日期：2026年5月18日；录用日期：2026年6月8日；发布日期：2026年11月18日

摘　要

电商平台中的用户评论不仅反映消费者对商品质量、物流服务和售后体验的真实感受，也是平台开展口碑监测、商品优化和风险预警的重要依据。由于中文电商评论普遍存在篇幅短、口语化表达多、评价维度分散以及噪声字符较多等特点，人工逐条分析往往效率低、成本高，难以满足实际业务需求。为提升电商评论情感识别的自动化水平，本文以Online Shopping 10 Cats数据集为基础，构建了融合BERT、BiGRU、Attention和FGM对抗训练的情感分类模型，并将其用于电商评论正负极性识别。实验结果表明，该模型在F1-score上达到0.9525，较对比模型具有一定优势。进一步分析发现，Attention机制有助于突出评论中的关键评价片段，FGM对抗训练能够增强模型对噪声表达的适应能力。研究结果说明，该方法可为电商平台评论监测、差评识别、商品口碑分析和用户反馈整理提供技术支持。

关键词

电商评论，情感分类，BERT，BiGRU，注意力机制

Research on Sentiment Classification of E-Commerce User Reviews Based on Deep Learning

Weiwei Zhang[1], Xirong Cao[1], Shi Cheng[2*]

[1]Xinglin College, Nantong University, Nantong Jiangsu
[2]School of Artificial Intelligence and Computer Science, Nantong University, Nantong Jiangsu

Received: May 18, 2026; accepted: June 8, 2026; published: November 18, 2026

Abstract

User reviews on e-commerce platforms not only reflect consumers' genuine feelings about product

*通讯作者。

文章引用: 张维薇, 曹稀蓉, 程实. 基于深度学习的电商用户评论情感分类研究[J]. 电子商务评论, 2026, 15(11): 487-494. DOI: 10.12677/ecl.2026.15111264

quality, logistics services and after-sales experience, but also serve as an important basis for platforms to conduct reputation monitoring, product optimisation and risk warning. Due to characteristics such as generally short length, colloquial expression, scattered evaluation dimensions and numerous noisy characters in Chinese e-commerce reviews, manual analysis of each review is often inefficient and costly, making it difficult to meet actual business needs. To improve the automation level of sentiment recognition in e-commerce reviews, this paper constructs a sentiment classification model based on the Online Shopping 10 Cats dataset, integrating BERT, BiGRU, Attention and FGM adversarial training, and applies it to recognising positive and negative sentiment in e-commerce reviews. Experimental results show that the model achieves an F1-score of 0.9525, demonstrating certain advantages over comparison models. Further analysis indicates that the Attention mechanism helps highlight key evaluation fragments within reviews, and FGM adversarial training can enhance the model's adaptability to noisy expressions. The research results suggest that this method can provide technical support for review monitoring, negative review detection, product reputation analysis and user feedback management on e-commerce platforms.

Keywords

E-Commerce Reviews, Sentiment Classification, BERT, BiGRU, Attention Mechanism

Copyright © 2026 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

随着电子商务的发展，用户评论已经成为影响消费决策的重要信息来源。情感分析旨在从文本中识别观点、情绪倾向及评价对象，已经成为自然语言处理中的重要研究方向[1]。对于平台和商家而言，评论不仅是消费者表达满意度和投诉意见的渠道，也包含着商品质量、物流速度、包装体验、客服响应等多方面的业务信息。如何从海量评论中快速识别情绪倾向，并进一步提炼出用户关注的核心问题，已经成为电商运营中的一项基础工作。

在实际场景中，人工阅读和整理评论存在明显局限。一方面，电商平台评论数量大、更新快，人工处理难以覆盖全部内容；另一方面，评论语言往往较为口语化，常伴随省略、重复、错别字、表情符号和转折表达，给情感判断带来一定困难。例如，同一条评论中可能同时提到“质量不错”和“物流太慢”，其整体情感并非简单的词汇累加即可判断。对于平台而言，如果无法及时识别负面评论，可能会影响售后响应和商品治理；对于商家而言，如果不能准确归纳评论中的高频问题，也难以有效改进商品和服务。

因此，面向电子商务场景构建自动化评论情感分类方法具有较强的现实意义。近年来，深度学习和预训练语言模型在文本分类任务中表现出较好的效果，为电商评论分析提供了新的技术手段。早期情感分类多依赖词袋表示和传统机器学习方法[2]，但这类方法对上下文语义和隐含情绪关系的刻画能力有限。相比之下，深度学习模型能够更好地学习文本表示，从而提升对复杂表达的识别能力[3] [4]。不过，电商评论具有短文本、信息密集、局部情绪突出等特点，直接使用通用文本分类模型时，仍可能存在对关键片段关注不足、对噪声表达不够稳健等问题。

基于此，本文围绕电商评论分析需求，构建了一种融合 BERT、BiGRU、Attention 和 FGM 的情感分类模型，并利用 Online Shopping 10 Cats 数据集进行验证。因此，本文尝试从电商评论分析的实际需求出发，构建一套可用于评论监测、差评识别和口碑分析的分类方法。

2. 电商评论特征分析与业务需求

电商评论与一般新闻文本或社交媒体长文本不同，其语言特点和业务用途都具有较强针对性。首先，评论篇幅普遍较短，用户通常围绕少数几个维度给出评价，如“质量”“物流”“包装”“价格”“服务”等。其次，评论语言较为随意，常出现口语、省略、重复、错别字及非规范表达，这使得传统规则方法较难稳定识别其情感倾向。再次，电商评论中的情绪往往不是单一维度的，例如一条评论可能同时对商品和物流做出评价，而平台和商家更关注的是评论背后对应的业务问题。

从电商应用角度看，评论情感识别至少可以服务于以下几类需求：

(1) 口碑监测：通过批量识别正负评论，快速掌握商品整体评价趋势；

(2) 差评预警：及时发现负面评论集中出现的商品或服务问题；

(3) 商品优化：从高频评价内容中识别用户关注点，为商品改进提供依据；

(4) 客服辅助：将负面评论自动分类后分发给相应处理环节，提高响应效率。

因此，电商评论情感分类不仅是一个文本分类问题，更是一个面向业务管理的信息提取问题。本文后续的方法设计与实验分析，也都围绕这一目标展开。

3. 数据来源与文本特征分析

本文使用的语料来自 Online Shopping 10 Cats 数据集。该数据集包含多个商品类别下的中文用户评论，样本规模较大，能够较好反映电商评论中的实际表达情况。数据集基本信息如表 1 所示。

Table 1. Basic information on the Online Shopping 10 Cats dataset
表 1. Online Shopping 10 Cats 数据集基本情况

指标	内容
数据集名称	Online Shopping 10 Cats
数据类型	中文电商用户评论文本
商品类别数	10
标签类型	二分类标签(正向/负向，对应 0/1)
总样本数	62,713
正向样本数	31,728
负向样本数	30,985

从数据分布来看，该数据集的正负样本较为均衡，这有助于避免模型过度偏向某一类别。为了更直观地展示语料特征，本文对评论文本进行了初步统计分析。结果表明，评论中的高频词主要集中在“质量”“物流”“包装”“服务”“价格”等与消费体验直接相关的词汇上，这说明电商评论的核心信息往往围绕实际购物过程展开，而不是抽象的情绪表达。评论高频词分布图如图 1 所示。

4. 面向电商评论的情感分类方法

4.1. 模型设计思路

本文模型以 BERT 为基础[5] [6]，结合 BiGRU [7]、Attention 机制[8]和 FGM 对抗训练[9] [10]，构建面向电商评论的情感分类框架。该框架的设计目标并不是追求复杂结构本身，而是尽可能契合电商短评的语言特征：一方面利用 BERT 获取上下文语义表示，另一方面借助 BiGRU 和 Attention 强化对关键评

价片段的识别，同时通过 FGM 提升模型面对噪声表达时的稳定性。

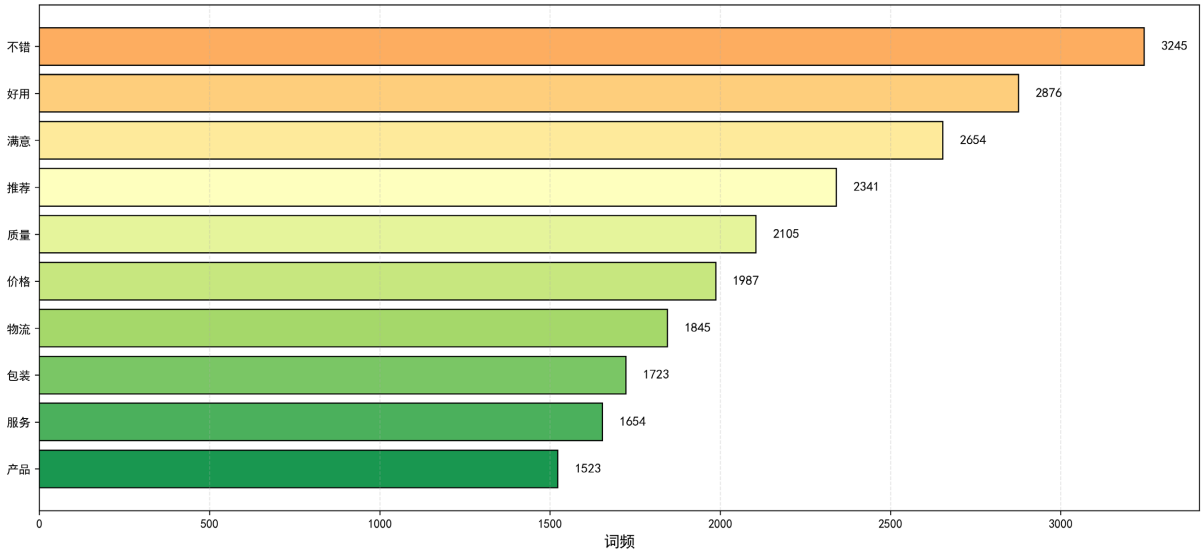


Figure 1. Top 10 high-frequency word distribution chart
图 1. TOP 10 高频词分布统计图

4.2. 模型结构

结合上述电商评论短文本、情感线索分散以及噪声表达较多等特点，本文进一步设计面向该场景的情感分类模型，以提升评论极性识别的准确性和稳定性。设输入评论文本序列为 $X=\{x_1,x_2,\cdots,x_n\}$，在经过 BERT 编码后，可得到对应的上下文语义向量序列 $H=\{h_1,h_2,\cdots,h_n\}$，其中，h_i 表示第 i 个 token 在上下文中的语义表示。不同于仅提取[CLS]向量进行分类的方式，本文保留完整序列输出，并将其送入 BiGRU 层继续建模。BiGRU 在正向和反向两个方向上分别学习文本序列信息，其隐藏状态可以表示为：

$$\overrightarrow{h_t}=GRU\left(x_t,\overrightarrow{h_{t-1}}\right) \tag{1}$$

$$\overleftarrow{h_t}=GRU\left(x_t,\overleftarrow{h_{t+1}}\right) \tag{2}$$

经过 BiGRU 层后，模型能够进一步理解文本前后片段之间的依赖关系，尤其适合处理“虽然……但是……”“本来以为……结果……”等电商评论中常见的语义转折结构。

在此基础上，为突出不同词语在情感判断中的重要程度，本文引入 Attention 机制。设 BiGRU 输出为h_t，则词语的重要性得分为：

$$u_t=\tanh\left(W_w h_t+b_w\right) \tag{3}$$

归一化后的注意力权重为：

$$\alpha_t=\frac{\exp\left(u_t^{\mathrm{T}}u_w\right)}{\sum_t\exp\left(u_t^{\mathrm{T}}u_w\right)} \tag{4}$$

于是评论文本的加权表示可以写为：

$$s=\sum_t\alpha_t h_t \tag{5}$$

该表示向量 s 聚合了模型认为更重要的情感特征。最终，分类层通过全连接映射与 Softmax 函数输出类别概率：

$$\hat{y} = \text{Softmax}(Ws + b) \tag{6}$$

训练中采用交叉熵损失函数优化模型参数，其形式为：

$$L = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log(\hat{y}_{ic}) \tag{7}$$

其中，N 表示样本数，C 表示类别数，y_{ic} 表示真实标签，$\hat{y}_{ic}$ 表示预测概率。

为提升模型鲁棒性，本文在训练阶段引入 FGM 对抗训练。其基本思想是在嵌入层上添加沿梯度方向的微小扰动，使模型在面对轻微输入变化时仍能保持稳定预测。设梯度为 g，则扰动向量表示为：

$$r_{adv} = \epsilon \frac{g}{\|g\|} \tag{8}$$

其中，ϵ 为扰动系数。通过对抗样本的引入，模型能够学习更平滑的决策边界，从而增强其对口语噪声、符号替代和轻微拼写变异的适应能力。

实验设计方面，本文将数据集划分为训练集、验证集和测试集。设样本总量为 N，则有：

$$N = N_{train} + N_{val} + N_{test} \tag{9}$$

本实验在正式划分数据集时采用分层抽样方法，将清洗后的 62,713 条样本按训练集、验证集和测试集划分为 43,899、9407 和 9407 条，比例约为 70%:15%:15%。分层抽样的目的在于尽量保持各子集中正向与负向样本的分布一致，从而避免由于随机划分导致的类别偏移，保证模型训练、调参与最终评估的公平性和代表性。这样的划分方式既能充分利用训练数据进行模型学习，又能在验证集和测试集上较客观地反映模型的泛化性能。

4.3. 评价指标与参数设置

模型评价采用文本分类中较常用的 Accuracy、Precision、Recall 和 F1-score 指标。其计算公式如下：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{10}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{11}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

其中，TP、TN、FP、FN 分别表示真正例、真负例、假正例和假负例。通过这些指标，可以从整体正确率、正类识别准确性和召回能力等多个角度对模型表现进行综合评价。

模型训练参数根据计算资源与数据规模进行设置。为保证后续实验可复现，对主要参数进行统一记录见表 2。

5. 实验结果分析

5.1. 模型对比结果

本文选取 SVM、BiGRU、TextCNN 以及 BERT + BiGRU + Attention + FGM 模型进行对比，结果见

表 3。考虑到各模型在 Accuracy、Precision 和 Recall 上的差异较为接近，本文以 F1-score 作为主要比较指标，并结合混淆矩阵中的 TP、TN、FP 和 FN 分析分类效果。

Table 2. Model parameter settings
表 2. 模型参数设置

参数	值	说明
预训练模型	bert-base-chinese	12 层 Transformer，768 维隐层
最大序列长度	128	最大输入长度
隐藏维度	768	BERT 输出维度
BiGRU 隐藏维度	128	BiGRU 双向隐层大小
Batch Size	16	训练批次大小
学习率	2e−5	AdamW 优化器学习率
Epoch 数	3	训练轮数
Dropout	0.1	防止过拟合
FGM 扰动系数	1.0	对抗训练扰动系数

Table 3. Comparison of classification results for different models
表 3. 不同模型分类结果对比

模型	F1-score	TP	TN	FP	FN
SVM	0.7821	3678	3679	1025	1025
BiGRU	0.9493	4465	4465	239	238
TextCNN	0.9509	4472	4473	231	231
BERT + BiGRU + Attention + FGM	0.9525	4480	4480	224	223

从结果看，传统 SVM 明显落后于深度学习模型，这说明电商评论中的上下文关系和隐含态度信息并不能被简单的词频特征充分表达。对于业务场景而言，这意味着若仍依赖传统方法进行评论监测，平台可能会遗漏大量带有真实负面倾向的表达。

BiGRU 和 TextCNN 已经能够取得较高的 F1-score，说明电商评论短文本任务更适合由深度学习模型处理。TextCNN 略优于 BiGRU，也符合短文本中局部关键词决定情感极性的特点。对于平台来说，这意味着很多评论的判断重点并不在长距离依赖，而在少量关键短语上。

本文融合模型在对比中取得最佳结果，F1-score 达到 0.9525。虽然提升幅度不算大，但在已有较高基线的情况下，这种改进仍具有实际意义。更重要的是，该模型在 FP 和 FN 上都略低于其他方法，说明其正负样本识别更平衡。对于电商平台而言，减少将负面评论误判为正面评论的情况，意味着更有利于及时发现潜在投诉和商品问题。

5.2. 消融实验与模块作用分析

表 4 给出了消融实验结果。

基础 BERT 模型已经具备较强的分类能力，说明预训练语言模型对中文电商评论具有较好的适配性。加入 BiGRU 后性能略有回落，表明在当前数据集上，BERT 本身对上下文的编码已较充分，额外增加循

环结构并不一定带来收益。

Table 4. Comparison of ablation experiment results
表 4. 消融实验结果对比

模型变体	F1-score	TP	TN	FP	FN
BERT	0.9505	4470	4471	233	233
BERT + BiGRU	0.9499	4468	4468	236	235
BERT + BiGRU + Attention	0.9517	4476	4477	227	227
BERT + BiGRU + Attention + FGM	0.9525	4480	4480	224	223

Attention 机制的加入带来了更稳定的提升，这与电商评论的任务特点较为一致。因为评论往往很短，真正影响极性的词语并不多，模型能否抓住这些关键位置，直接影响分类结果。对平台来说，这意味着 Attention 机制有助于更准确地识别“差评核心点”，例如物流、质量或售后问题。

FGM 对抗训练进一步提升了模型效果，说明在包含错别字、缩写和噪声符号的真实评论中，模型鲁棒性仍然很重要。对电商业务而言，这种稳定性有助于减少因非规范表达导致的误判，使评论监测结果更可靠。

5.3. 电商评论分析的应用示意

在电商平台的实际运行中，评论情感分类的作用并不局限于正负极性的自动判断，还可以作为评论治理与用户反馈整理的前置环节。基于模型输出结果，平台可进一步结合评论文本中的高频评价词，对用户关注的价格、质量、物流、包装和服务等问题进行归纳，从而为商品优化和售后处理提供参考。图 2 给出了电商评论分析的一个应用示意。

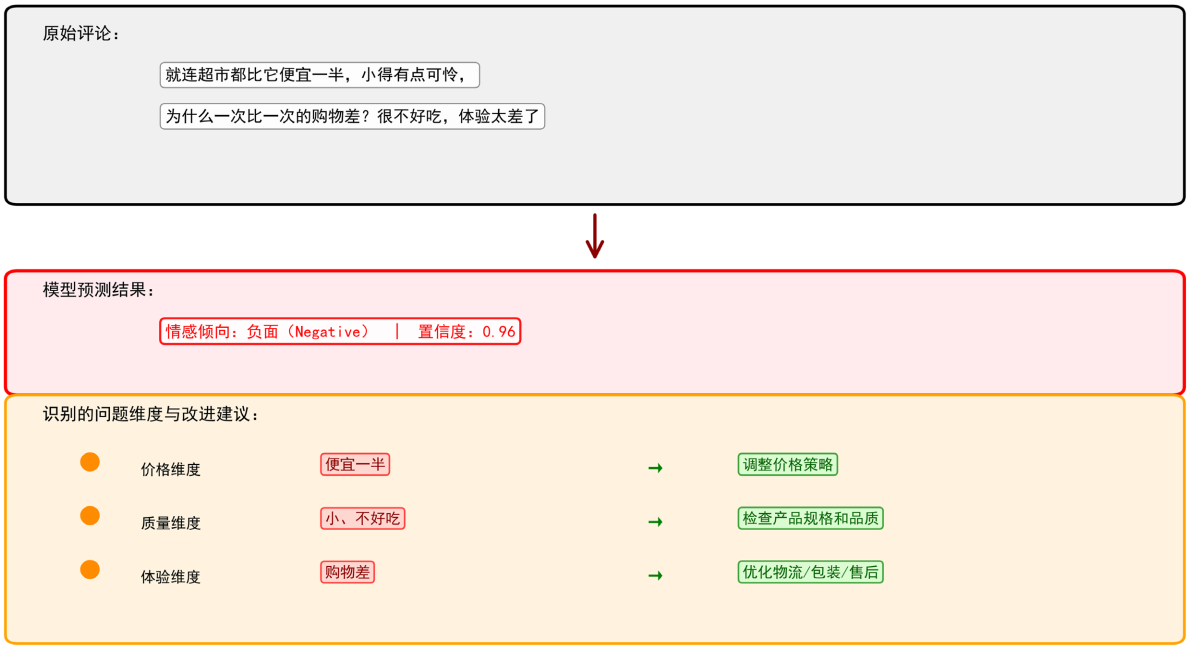


Figure 2. Schematic diagram of sentiment recognition and problem dimension analysis in e-commerce reviews
图 2. 电商评论情感识别及问题维度分析示意图

由图 2 可以看出，当模型识别出评论的负面情绪后，相关结果可以继续用于定位评论中涉及的主要问题维度，并辅助形成后续处理思路。需要说明的是，图中问题维度与改进建议部分主要用于说明情感分类结果在电商业务场景中的潜在使用方式，其目的在于展示评论分析在平台运营中的延伸价值。

6. 结论

本文围绕电商平台用户评论分析需求，基于 Online Shopping 10 Cats 数据集构建了融合 BERT、BiGRU、Attention 和 FGM 的情感分类模型，并对其分类效果进行了实验验证。结果表明，相较于传统方法，深度学习模型在处理电商评论中的口语化表达、短文本结构和局部情感线索方面更具优势。本文所构建的模型在 F1-score 上达到 0.9525，且在正负样本识别上保持了较为均衡的表现。消融实验进一步说明，Attention 机制有助于模型关注评论中的关键评价片段，FGM 对抗训练则在一定程度上提升了模型对噪声表达的适应能力。

从电子商务应用角度看，用户评论情感分类的价值并不局限于文本极性判断本身，更在于为平台评论治理和商家运营提供可利用的信息基础。对于平台而言，该方法能够帮助其在大规模评论数据中快速识别负面反馈，辅助开展差评预警、售后响应和商品风险监测；对于商家而言，分类结果可用于整理用户在质量、物流、包装和服务等方面的集中意见，为商品优化和服务改进提供依据。总体来看，面向电商评论场景的情感分类方法具有较强的实际应用潜力。

参考文献

[1] 韩雨彤, 周季蕾, 任菲. 动态视角下实时评论内容对直播电商商品销量的影响[J]. 管理科学, 2022, 35(1): 17-28.

[2] 成耀, 张铎, 周宇, 等. 基于模糊聚类的电商企业不平衡财务数据风险预测方法[J]. 电子商务评论, 2025, 14(1): 640-647.

[3] 田依林, 刘璐, 莫洁安. 基于在线评论的跨境电商用户需求认知分析[J]. 情报科学, 2025, 43(11): 45-53.

[4] 徐文龙. 电商用户评论的半监督情感分类研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2025.

[5] 王盈, 张文龙, 唐卓然. 基于电商评论情感分析的用户聚类方法研究[J]. 现代信息科技, 2023, 7(16): 24-27, 33.

[6] 高茂娇. 面向电商评论文本的跨领域情感分析研究[D]: [硕士学位论文]. 淮南: 安徽理工大学, 2024.

[7] 洪艺萍. 结合 Bert 和 BiGRU 的电商产品评论情感分析算法研究[J]. 佳木斯大学学报(自然科学版), 2026, 44(2): 91-94.

[8] 纪杨洋, 叶东琛. 跨平台评论信息的融合处理技术与聚类效果研究[J]. 情报理论与实践, 2025, 48(S1): 240-244.

[9] 何金凤, 沈玮, 张圣琪, 等. 基于改进 BP 神经网络的电商财务风险智能预测方法[J]. 电子商务评论, 2024, 13(4): 3022-3028.

[10] 朱子涵, 赵书娴, 程实. 人工智能时代老年数字阅读服务的适老化供给路径研究[J]. 老龄化研究, 2026, 13(9): 129-136. https://doi.org/10.12677/ar.2026.139635