Modern Linguistics 现代语言学, 2026, 14(11), 19-28
Published Online November 2026 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2026.14111042

Hans汉斯

基于语料库的人机学术英语摘要语言特征多维对比研究

解佳颖

西安电子科技大学语言智能学部，陕西 西安

收稿日期：2026年7月31日；录用日期：2026年10月22日；发布日期：2026年11月2日

摘　要

为探究人机学术写作差异，本研究构建平行语料库，提取多维计量指标，运用随机森林算法对比人类专家与大模型的英文摘要。结果表明，大模型虽能拟合宏观连贯，但在微观语用策略上模型倾向堆砌复杂词汇与过度嵌套或扁平化的极端句式以刻意模拟学术感，难以像人类般自然平衡文本信息。研究证实深层句法与词汇特征能甄别人机文体英文摘要，为揭示机器生成机制与优化语篇甄别提供了实证支持。

关键词

语料库，大语言模型，学术英语摘要，句法复杂度，随机森林

A Multidimensional Comparative Study on Linguistic Features of Human-Machine Academic English Abstracts Based on Corpus

Jiaying Xie

Faculty of Linguistic Intelligence, Xidian University, Xi'an Shaanxi

Received: July 31, 2026; accepted: October 22, 2026; published: November 2, 2026

Abstract

To explore the discrepancies between human and AI academic writing, this study constructs a parallel corpus and extracts multi-dimensional quantitative metrics, and adopts the random forest algorithm to compare English abstracts written by human scholars and large language models. The results reveal that although large language models can reproduce macro textual coherence, they adopt biased micro-pragmatic strategies: they tend to overuse sophisticated vocabulary and employ extreme

文章引用: 解佳颖. 基于语料库的人机学术英语摘要语言特征多维对比研究[J]. 现代语言学, 2026, 14(11): 19-28.
DOI: 10.12677/ml.2026.14111042

sentence structures featuring excessive embedding or flat syntax to artificially simulate an academic tone, and fail to balance textual information naturally as human writers do. This study verifies that deep syntactic and lexical features can distinguish human-written abstracts from AI-generated ones, and provides empirical evidence for uncovering the generation mechanism of large language models and improving the discrimination of machine-produced texts.

Keywords

Corpus, Large Language Model, Academic English Abstract, Syntactic Complexity, Random Forest

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

近年来，大语言模型在自然语言处理领域取得了显著进展[1]。基于大规模语料库的预训练与上下文表征，豆包、DeepSeek 等模型在文本生成和句法润色方面表现突出，并逐渐应用于学术英语写作。然而，现有语言学领域研究仍不充分且多停留在比较人机摘要在词汇复杂度与句法复杂度的指标差异[2] [3]，缺乏进一步探讨如何利用这些特征来有效识别文本来源并提升判别准确度。

传统的学术英语语篇研究多集中于宏观修辞结构或元话语分布，而对机器生成文本的微观语言学特征探讨仍处于起步阶段。在语料库语言学的框架下，学术语篇的体裁特征往往依赖于特定语法结构的分布。句法复杂度，尤其是短语层面的信息压缩与从句嵌套，是衡量学术文本信息密度与作者认知加工偏好的关键指标[4] [5]。与此同时，语篇连贯性直接影响摘要的逻辑推进与信息传递效率[6]。现有研究较少将微观句法构型与宏观语篇连贯结合起来寻找核心特征做判别分析，因而难以揭示大模型在生成学术摘要时的潜在语用策略与文体偏误。

基于此，本研究构建人机平行语料库，以国际外语语言学期刊的专家摘要为基准，生成对应的大模型摘要。研究采用语料库工具提取句法复杂度与语篇连贯性指标，并结合随机森林分类算法，分析人机文本的差异与特征贡献度。研究目标在于识别最具区分度的语言学特征，从而揭示人机学术摘要在语言学层面的本质差异，并为学术英语写作教学与人工智能文本甄别提供实证支持[7]。

2. 文献综述

2.1. 生成式 AI 与学术英语语篇表征研究

近年来，大语言模型在学术英语写作中的应用引发了广泛关注。已有研究指出，机器生成文本在语法正确性和流利度方面表现突出，但在修辞策略和语篇衔接上往往呈现程式化倾向[8]。Biber 和 Gray [4]强调学术英语的句法复杂度主要体现在短语压缩与从句嵌套，这也是衡量学术文本信息密度的重要指标。论文摘要是论文关键内容、研究方法、研究结果及结论等的概括，它对读者获取论文信息和学术交流至关重要[9]。然而，针对学术摘要这一高度规约化体裁的研究仍然不足，这为进一步探讨人机文本的语言学差异提供了研究空间。

2.2. 句法复杂度与语篇连贯性的摘要计量研究

句法复杂度与语篇连贯性是学术摘要的重要语言学特征。Kyle 和 Crossley [5]提出，句法复杂度反映

了作者的认知加工偏好。McNamara 等[6]则通过 Coh-Metrix 系统量化了文本的局部连贯性与词汇多样性。卜小伟和杨云峰[10]对比英汉摘要的衔接方式，发现英文摘要更依赖人称照应与连接词，而中文摘要更倾向于词汇衔接；朱君辉等[11]对比了大模型生成回答与人类回答的语言特征，发现机器文本在词汇多样性和句法复杂度上存在显著差异。这些研究为比较人机摘要在句法与语篇特征上的差异提供了理论基础。

2.3. 基于机器学习的文本来源判别研究

随着机器生成文本在学术界的普及，如何鉴别文本来源成为研究热点。早期方法依赖规则或统计特征，但难以应对大模型生成的复杂文本[12] [13]。近年来，研究者开始引入机器学习方法，其中随机森林模型因能输出语言学特征的重要性而受到关注[6]。崔中良[14]从语言生成机制出发，指出机器生成文本在本质上是基于词语规则预测的横向加工，缺乏人类自然语言所具备的“具身经验”与不规则性，这种特征差异为通过语言学指标判别文本来源提供了理论依据。

鉴于此，本研究构建了一个包含外语语言学顶刊人类专家基准文本、豆包大模型与 DeepSeek 大模型生成文本的三轨影子平行语料库。

研究首先通过单变量显著性检验揭示三组语料在单一语言学指标上的系统性异同；在此基础上，进一步构建随机森林机器学习模型，通过评估多维语言学特征矩阵的分类准确率并输出特征重要性排行，旨在以可解释的方式挖掘拦截机器语篇的核心特征，从而为人工智能时代的学术英语语篇表征及学术盲审实践提供实证依据。为此，本研究旨在解决以下三个问题：

第一，豆包与 DeepSeek 大模型生成的语言学学术摘要，与人类专家文本在小句层面的句法复杂度指标上表现出怎样的差异？

第二，豆包与 DeepSeek 大模型生成的语言学学术摘要与人类专家文本，在语篇连贯性与可读性度量维度上展现出怎样的差异？

第三，基于微观句法与高阶语篇特征构建的随机森林分类模型，能否有效切分语言学学术英语摘要的人机文体边界？哪些语言学特征指标在此分类中做出了主要贡献？

3. 研究方法

3.1. 研究设计

本研究分别采用 Doubao-2.0 Pro、DeepSeek-v4-pro (API 版本)构建 AI 生成组语料，并使用零样本提示生成学术英语摘要。选择该模型是因为其基于超大规模双语预训练，生成的学术文本语法成熟，能有效代表当前生成式人工智能的高英语拟合水平。

Table 1. Model parameter configuration
表 1. 模型参数设置

参数	设置
Temperature	0.3
Top-p	1
Frequency penalty	0
Presence penalty	0

本研究采用基于语料库的定量对比研究设计。整个实验流程包含三个核心部分。首先，研究构建一个

由语言学顶级期刊人类文本与大语言模型生成文本构成的影子平行语料库。其次，利用特定文本计量工具与自然语言处理库，批量提取人机语篇的句法复杂度指标、语篇连贯指标与可读性指数。最后，通过单变量统计检验分析三语料在单一指标上的系统性异同，并在此基础上构建随机森林算法模型，以测试多维语言学指标的全局分类准确率。在 API 批量调用时，对核心解码参数进行了严格锁定如表 1，确保学术摘要文体的严谨与稳定。温度参数与 top-p 参数的设置依据为 Moslem 等(2023) [15]、Kwok 等(2025) [16]。

3.2. 语料库构建与样本选择

本研究构建的影子平行语料库总计包含 150 篇学术摘要样本，由作为对比基准的人类专家学术摘要语料与作为实验对照的大语言模型生成学术摘要语料两部分共同构成。其中，基准组的人类专家语料精选自外国语言学及应用语言学领域的 5 本核心期刊，具体包括 System、Journal of Pragmatics、Journal of English Linguistics、Cognitive Linguistics 以及 Bilingualism: Language and Cognition。研究从每本期刊中各随机抽取 10 篇实证研究论文的摘要，共计 50 篇，用以代表高水平人类学者的学术英语特征。为了避免生成式人工智能对基准文本的隐性污染，所有入选的人类文献发表时间均被严格控制在 2015 年至 2020 年之间。

对照组的机器影子语料则通过提取前述 50 篇人类文献的标题与关键词作为唯一的语义输入源。研究将这些文本种子输入豆包、DeepSeek 大模型，并设定了 150 至 250 词的字数硬性限制，采用零样本英文提示词，未添加任何关于文体倾向、语法高级度等诱导性指令。

3.3. 语言学特征提取

3.3.1. 句法复杂度维度指标

采用 TAASSC 2.0 (Tool for the Automatic Analysis of Syntactic Sophistication and Complexity)的句法复杂度与成分分析模块，从宏观句体量、小句嵌套、动词使用及名词修饰四个层面提取 8 项微观指标。

3.3.2. 语篇衔接与可读性维度指标

采用 Coh-Metrix 3.0 系统，从局部连贯、显性衔接、词汇变异以及整体加工负载四个层面提取 4 项高阶宏观指标。

3.4. 数据分析与统计检验

本研究的数据分析包含单变量差异检验与随机森林分类两个部分。

在单变量检验中，首先对 12 项语言学指标进行 Shapiro-Wilk 正态性检验。针对符合正态分布且方差齐性的指标，采用单因素方差分析(One-way ANOVA)及 Tukey HSD 事后检验；不符合条件的指标则采用 Kruskal-Wallis H 检验及 Dunn’s 检验。显著性水平设定为 $\alpha = 0.05$。

在分类分析中，构建随机森林三分类模型，以 8 项 TAASSC 句法指标与 4 项 Coh-Metrix 语篇指标作为自变量，以文本来源(人类专家、豆包、DeepSeek)为因变量。模型采用五折交叉验证以评估泛化性能，并定量判定不同维度特征对人机文体边界的区分贡献度。

4. 结果与讨论

4.1. 基于 TAASSC 的句法复杂度特征差异分析

在句法构型维度上，本研究利用 TAASSC 工具提取了 8 个核心句法复杂度指标。三组语料的描述性统计、单因素方差分析及事后检验结果见表 2[1]。

[1]nouns_as_modifiers 指标提取自 TAASSC 工具的成分分析模块，其输出均值为经过常模参照标准化处理后的成分得分(Z-score)，用以表征各组别相对于基准语料库的偏离倾向，故数值呈现正负分布，非绝对频次计数。

Table 2. Difference test results of syntactic complexity dimension
表 2. 句法复杂度维度差异检验结果

指标(Metric)	人类专家组均值	豆包生成组均值	DeepSeek 生成组均值	F 值	偏 η^2	p 值	显著性	事后检验显著两两对比
MLS (平均句长)	27.201	25.080	28.737	6.334	0.079	0.0023	极显著	DeepSeek－豆包
MLT (平均 T 单位长度)	25.663	24.623	27.637	5.522	0.070	0.0049	极显著	DeepSeek－豆包
MLC (平均从句长度)	15.424	16.910	13.966	11.923	0.140	<0.001	极显著	DeepSeek－豆包、DeepSeek－人类、豆包－人类
C_S (每句从句数)	1.829	1.517	2.106	21.777	0.229	<0.001	极显著	DeepSeek－豆包、DeepSeek－人类、豆包－人类
C_T (T 单位复杂比)	1.724	1.492	2.020	20.739	0.220	<0.001	极显著	DeepSeek－豆包、DeepSeek－人类、豆包－人类
VP_T (每 T 单位动词短语数)	2.394	2.416	2.926	12.104	0.141	<0.001	极显著	DeepSeek－豆包、DeepSeek－人类
av_nominal_deps (平均名词性依存数)	1.510	1.662	1.571	9.476	0.114	<0.001	极显著	DeepSeek－豆包、豆包－人类
nouns_as_modifiers (名词作修饰语得分)	0.456	−0.354	−0.101	1.358	0.018	0.2605	不显著	无显著两两差异

根据表 2 的统计数据，单因素方差分析显示，除名词作修饰语频次(nouns_as_modifiers)外，其余 7 项句法复杂度指标在三组语料间均存在极显著的组间差异。

在宏观句长及 T 单位长度维度，MLS 与 MLT 均达到了极显著差异水平。然而，事后检验结果表明，这一显著性完全源于 DeepSeek 生成组与豆包生成组之间的两两差异(DeepSeek 的句长均值显著高于豆包)，而两类大模型组与人类专家组之间并无统计学显著差异。这说明从外在的篇幅控制来看，无论是豆包还是 DeepSeek，其生成的单句规模均与人类学者的学术摘要高度契合。

然而，深入到句子内部的微观构型时，数据则揭示了明显的结构分化。在衡量小句数量和主从嵌套倾向的每句从句数(C_S)、T 单位复杂比(C_T)以及每 T 单位动词短语数(VP_T)指标上，DeepSeek 组的均值均显著高于人类专家组与豆包组。尤其是 DeepSeek 的 C_S 均值达到了 2.106，明显超出了人类专家水平(1.829)，每句从句数(C_S)不仅呈现出极显著的组间差异，且贡献了极大的效应量(偏 $\eta^2 = 0.229$)，这构成了区分人机句法策略的核心断层。

这意味着 DeepSeek 在撰写学术文本时，更倾向于使用多层从小句或并列小句的纵向叠加来承载复杂语义。与之相反，豆包生成组则表现出明显的扁平化特征，其 C_S (1.517)与 C_T (1.492)值在三组中均为最低，说明其在生成学术文本时主动规避了多层从句嵌套。

与小句数量的发展路径形成互补的是单个小句内部的信息密度。平均从句长度(MLC)与平均名词性

依存数(av_nominal_deps)在三组间同样存在极显著分化。豆包生成组的 MLC 均值(16.910)显著高于人类专家组(15.424)与 DeepSeek 组(13.966)，且其 av_nominal_deps 均值(1.662)也处于最高水平。这说明豆包虽然减少了从句的嵌套数量，但极大地扩展了单个小句内部的成分长度，通过高频使用名词性依存成分在短语层面上进行信息打包。相反，DeepSeek 由于将字数分散在更多的从句中，其平均每个从句的长度反而显著偏低。人类专家组则在小句数量(C_S = 1.829)和小句长度(MLC = 15.424)上保持了相对均衡的状态，避免了过于冗长的小句或过度的从句叠加。

4.2. 基于 Coh-Metrix 的语篇衔接与可读性特征差异分析

本研究进一步引入语篇度量工具 Coh-Metrix，对人机文本的上下文局部连贯性、显性衔接标记、词汇多样性以及可读性等级进行了多维考察，差异性检验结果见表 3。

Table 3. Difference test results of discourse and vocabulary dimension
表 3. 语篇与词汇维度差异检验结果

指标(Metric)	人类专家组均值	豆包生成组均值	DeepSeek 生成组均值	F 值	偏 η^2	p 值	事后检验显著两两对比
CRFNO1 (邻近句名词重叠率)	0.709	0.568	0.613	3.957	0.051	0.0212	豆包 - 人类
CNCAll (显性连词总频次)	83.417	79.518	74.696	2.504	0.033	0.0852	无显著两两差异
LDMTLD (MTLD 词汇多样性)	88.661	114.379	132.015	31.668	0.301	<0.001	DeepSeek - 豆包、DeepSeek - 人类、豆包 - 人类
RDFKGL (Flesch-Kincaid 可读性等级)	17.204	20.281	21.318	65.080	0.470	<0.001	DeepSeek - 豆包、DeepSeek - 人类、豆包 - 人类

在语篇与词汇维度上，单因素方差分析显示，除显性连词总频次(CNCAll)不显著外，其余三项指标在三组语料间均存在统计学上的显著或极显著差异。

在词汇丰富度方面，LDMTLD 指标呈现出极显著的组间分化，且事后检验表明三组之间存在两两极显著差异。这表明两类大语言模型生成的文本在短篇幅内表现出远高于人类学者的词汇变异率。在文本生成机制中，这种极高的词汇多样性通常与语言模型的概率采样解码策略有关，如避免高频重复词汇的惩罚机制，大模型在生成时倾向于频繁更换相同概念的同义词表述。

在语篇连贯维度，邻近句名词重叠率(CRFNO1)展现出分化的特征($F = 3.957$, $p = 0.0212$)。事后检验(Tukey HSD)表明，显著的组间差异仅存在于豆包生成组与人类专家组之间($p < 0.05$)，豆包的重叠率(0.568)显著低于人类(0.709)；而 DeepSeek 生成组(0.613)虽较人类有所下降，但未达到统计学显著。这表明豆包在生成学术摘要时，存在更为明显的刻意规避词汇重复、频繁更换同义词的去同质化倾向，从而对局部语篇的连贯链条造成了破坏。

然而，从 Flesch-Kincaid 可读性等级(RDFKGL)来看，三组语料存在极显著的两两分化，其阅读难度表现为：DeepSeek 生成组(21.318) > 豆包生成组(20.281) > 人类专家组(17.204)。大模型生成的文本由于大量调用多音节学术衍生词或长难词，推高了可读性公式的得分，使其在外观形式上表现出极高的技术难度。但结合低名词重叠率来看，这种由词汇和局部句法堆砌带来的高可读性等级，并未转化为更紧密

的语篇连贯链条，反而体现了机器文本表层词汇难度偏高、深层显性衔接相对松散的结构特征。

4.3. 基于随机森林的人机学术摘要特征重要性分析

本研究使用随机森林模型，对 12 个关键语言学指标进行了特征重要性排序(见图 1)，以评估它们在区分人类、豆包和 DeepSeek 文本时的贡献度。结果显示，各项指标的区分能力差异明显，这反映了不同生成源在文本构建和语用策略上的本质区别。

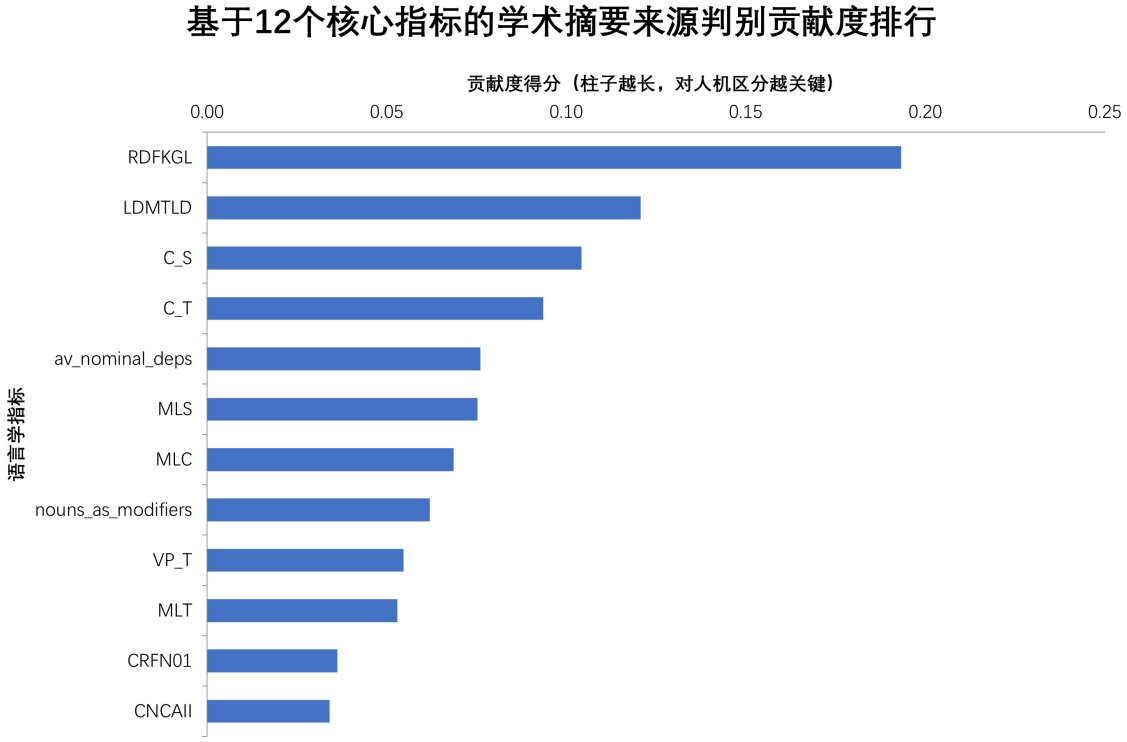


Figure 1. Contribution ranking of academic abstract source discrimination based on 12 core indicators
图 1. 基于 12 个核心指标的学术摘要来源判别贡献度排行

首先，文本的阅读难度和词汇丰富度是区分人机身份的最核心指标。模型结果表明，文本可读性(RDFKGL, 25.22%)和词汇多样性(LDMTLD, 11.64%)的贡献度排名前两位，合计占比近 37%。结合前文的均值对比可以看出，人工智能(尤其是 DeepSeek)在生成学术摘要时，采用了一种“词汇过度堆砌”的语用策略。它们倾向于使用高难度词汇和复杂的书面表达来刻意模仿学术风格；相比之下，人类学者则更遵循“精准且适度简洁”的交际原则。这种词汇使用上的刻意感和不自然感，成为了鉴别 AI 文本的有效特征。

其次，句法结构的复杂度是区分不同大模型的重要依据。每句从句数(C_S, 10.03%)、T 单位复杂比(C_T, 9.00%)和平均句长(MLS, 7.19%)紧随其后。数据表明，不同 AI 模型在句法包装上采取了截然不同的策略：DeepSeek 倾向于通过层层嵌套的从句来浓缩信息，句子结构复杂；而豆包则偏向于使用扁平、简短的句型。这种底层算法导致的句法分化，为辨认具体是哪一款 AI 提供了关键线索。

最后，传统的语篇连贯性指标已难以有效区分人机差异。值得注意的是，邻近句名词重叠率(CRFNO1, 3.44%)和连词使用频次(CNCAll, 3.24%)的贡献度排名垫底。这说明在上下文衔接和表面逻辑连贯方面，当前的大模型已经达到了与人类高度相似的伪装水平。因此，仅依靠连接词数量或名词重复率等表面衔接手段，已经无法准确鉴别人机文本。

总体而言，人类与大模型在学术摘要写作上的根本差异，已经不再是表面的语篇连贯，而是深层词汇选择的刻意性与句法嵌套策略的分化。

4.4. 人机学术摘要语用策略的定性分析

4.4.1. 句法嵌套(C_S)与语用深度的分化策略

正如前文定量结果所示，DeepSeek 的每句从句数(C_S)极显著高于豆包。通过对平行语料的定性检视，我们发现这一统计学分化源于不同模型对“学术深度”的语用策略差异。

豆包倾向于采用“句法扁平化”策略来降低读者的认知负荷，其生成的句子多以简单的并列或单一主从复合句为主。例如，在探讨语言学习策略的语料中，豆包生成了典型的扁平化结构：“The results reveal that memory and social strategies are the most frequently used, while metacognitive strategies see relatively less adoption.”

相反，DeepSeek 则高频调用“深层句法嵌套”策略，试图在单一句法树中穷尽复杂的逻辑关系。例如，在分析网络身份建构的语料中，DeepSeek 生成了高度嵌套的长难句：“We identify a paradox at the heart of these performances: users simultaneously disclaim the literal intent of their threats while relying on the visceral force of rape as a symbol of gendered power to cement homosocial bonds.”这种分化表明，DeepSeek 将高密度的从句叠加作为提升学术语体正式度的一种显性语用策略。

4.4.2. 词汇可读性(RDFKGL)与“过度学术包装”

随机森林模型表明，Flesch-Kincaid 可读性等级(RDFKGL)是切分人机边界的核心特征。定性分析进一步揭示，AI 文本极高的阅读难度实际上源于其“过度学术包装”的词汇语用倾向。

在人类学者的原始摘要中，关键概念的传递往往遵循格莱斯的“量准则”，用词自然且克制：“Noticing the hole (NTH) occurs when speakers want to say something, but realise they do not know the right word(s). Such awareness of lacking knowledge supposedly facilitates the acquisition of the unknown word(s) from later input”(Swain, 1993) [17]。

相比之下，大模型(尤其是 DeepSeek)倾向于无节制地调用低频、抽象的学术衍生词，形成了一种非自然的词汇饱和感，是一种刻意迎合学术语体概率分布的“词汇炫技”：“Drawing on dynamic systems theory, we argue that early phonological development proceeds not through the gradual accumulation of individual phonemes but through the abstraction of schemas—structured, holistic representations extracted from whole-word exemplars.”

4.5. 基于随机森林的多维特征分类预测

为系统评估所筛选的 12 项语言特征在区分人机学术摘要中的效力，本研究采用随机森林算法构建分类模型，并实施 5 折交叉验证。同时，为探究特征集的信息纯度与冗余度，本研究基于特征贡献度排序，进行了递增式特征验证[18] [19]。交叉验证结果(见图 2)显示，模型分类准确率随特征数量的增加呈现显著的阶段性变化规律。

首先，在仅引入可读性(RDFKGL)与词汇多样性(LDMTLD)时，模型准确率为 58.67%，表明单一的词汇层面特征不足以构成稳健的判别边界。然而，当引入关键句法指标“每句从句数(C_S)”时，模型准确率实现关键跃升，直达 78.00%。这一数据断层从实证角度证实，句法嵌套的复杂性是区隔人机学术文体的决定性维度。人工智能在生成摘要时，倾向于调用更为密集的从句结构来模拟学术语体。

其次，随着修饰语名词比例(nouns_as_modifiers)等中段细节特征的补充，模型准确率稳步攀升，并在纳入前 8 至 9 项特征时达到全局峰值 82.67%。

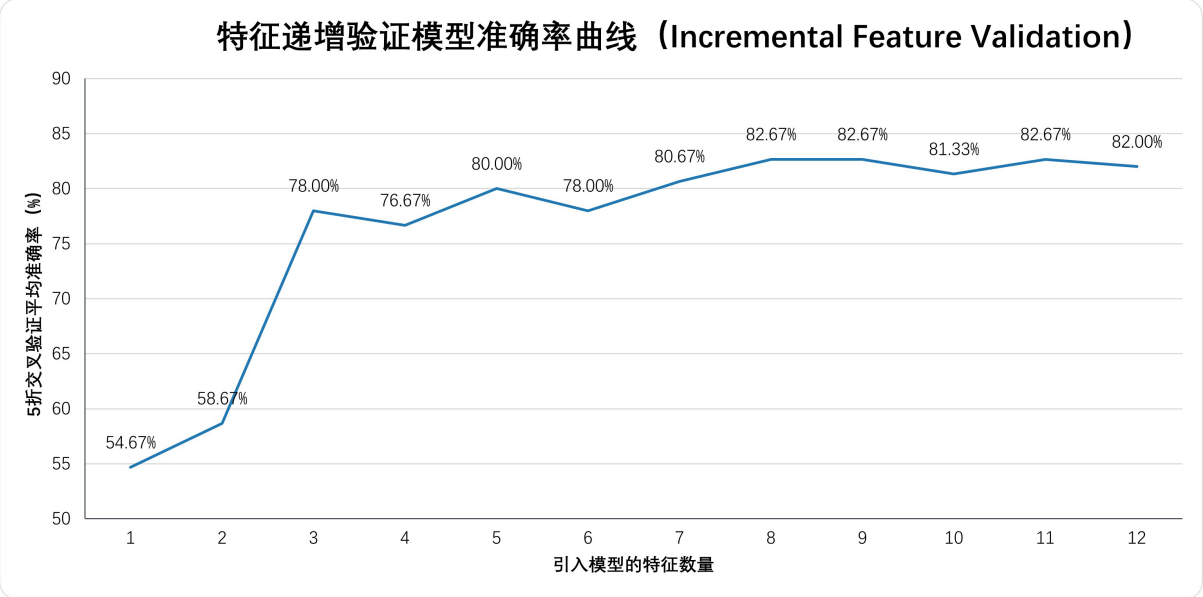


Figure 2. Incremental feature validation curve of model accuracy
图 2. 特征递增验证模型准确率曲线

最后，当纳入末端特征(如显性连词频次 CNCAll)时，准确率微调至 82.00%并趋于稳定。这一微弱的回落反证了特征筛选的必要性，表明表面连词等后置指标所携带的文体信息，已被前置的高权重指标(词汇与句法)所覆盖。综合而言，模型最终稳定在 82.00%的准确率，证实本研究构建的 12 项特征集已达到判别效力与维度的最优平衡，有效避免了信息冗余。更重要的是，词汇和句法指标在此次判别中展现出的绝对主导地位，揭示了人机写作机制的本质差异：大模型受限于逐词生成的算法逻辑，只能通过在局部刻意堆砌复杂词汇和多层从句来展现“学术感”，却难以像人类学者那样自然、从容地把控整篇摘要的宏观连贯性。

5. 结论

本研究基于人机平行语料库与随机森林算法，对比了人类与大模型在学术摘要中的计量语言特征。研究证实，大模型虽能拟合宏观衔接，但在微观语用策略上存在显著偏差：模型倾向于通过过度堆砌词汇与极端句式，包括过度嵌套或扁平化，来模拟学术语体，难以实现人类学者在信息密度与交际经济性间的平衡。基于词汇多样性、可读性及句法嵌套构建的分类模型准确率达 82.00%，确立了深层句法与词汇指标作为人机文体边界的核心地位，并为机器语篇甄别提供了量化依据。

基金项目

西安电子科技大学研究生创新基金项目(YJSJ26010)。

参考文献

[1] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P. and Amodei, D. (2020) Language Models Are Few-Shot Learners. *Advances in Neural Information Processing Systems* 33: *Annual Conference on Neural Information Processing Systems* 2020, *NeurIPS* 2020, 6-12 December 2020, 1877-1901.

[2] Fredrick, D.R. and Craven, L. (2025) Lexical Diversity, Syntactic Complexity, and Readability: A Corpus-Based Analysis of ChatGPT and L2 Student Essays. *Frontiers in Education*, **10**, Article 1616935. https://doi.org/10.3389/feduc.2025.1616935

[3] Shen, C., Guo, J., Shi, P., Qu, S. and Tian, J. (2023) A Corpus-Based Comparison of Syntactic Complexity in Academic Writing of L1 and L2 English Students across Years and Disciplines. *PLOS ONE*, **18**, e0292688. https://doi.org/10.1371/journal.pone.0292688

[4] Biber, D. and Gray, B. (2016) Grammatical Complexity in Academic English: Linguistic Change in Writing. Cambridge University Press. https://doi.org/10.1017/cbo9780511920776

[5] Kyle, K. and Crossley, S. (2017) Assessing Syntactic Sophistication in L2 Writing: A Usage-Based Approach. *Language Testing*, **34**, 513-535. https://doi.org/10.1177/0265532217712554

[6] McNamara, D.S. and Graesser, A.C. (2012) Coh-Metrix: An Automated Tool for Theoretical and Applied Natural Language Processing. In: McCarthy, P.M. and Boonthum-Denecke, C., Eds., *Applied Natural Language Processing*: *Identification*, *Investigation and Resolution*, IGI Global, 188-205. https://doi.org/10.4018/978-1-60960-741-8.ch011

[7] 吴继峰, 胡韧奋, 彭一平. 大语言模型和汉语母语者生成学术汉语语言特征对比研究[J]. 语言文字应用, 2025(4): 27-37.

[8] Casal, J.E. and Kessler, M. (2023) Can Linguists Distinguish between ChatGPT/AI and Human Writing? A Study of Research Ethics and Academic Publishing. *Research Methods in Applied Linguistics*, **2**, Article ID: 100068. https://doi.org/10.1016/j.rmal.2023.100068

[9] 张鹏华, 潘一. 中外学位论文英文摘要语步对比研究[J]. 现代语言学, 2023, 11(5): 2181-2192.

[10] 卜小伟, 杨云峰. 英汉学术论文摘要语篇衔接的对比研究[J]. 长沙铁道学院学报: 社会科学版, 2010, 11(3): 171-172.

[11] 朱君辉, 王梦焰, 杨尔弘, 等. 大模型生成回答与人类回答文本的语言特征比较研究[J]. 中文信息学报, 2024, 38(4): 17-27.

[12] Solaiman, I., Brundage, M., Clark, J., Askell, A., Herbert-Voss, A., Wu, J. and Wang, J. (2019) Release Strategies and the Social Impacts of Language Models. arXiv: 1908.09203.

[13] Uchendu, A., Le, T., Shu, K. and Lee, D. (2020) Authorship Attribution for Neural Text Generation. *Proceedings of the* 2020 *Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), 16-20 November 2020, 8384-8395. https://doi.org/10.18653/v1/2020.emnlp-main.673

[14] 崔中良. 生成式人工智能是否能替代人类生成语言?——从梅洛-庞蒂的语言生成观探讨[J]. 西安交通大学学报(社会科学版), 2025, 45(3): 182-190.

[15] Moslem, Y., Haque, R., Kelleher, J. and Way, A. (2023) Adaptive Machine Translation with Large Language Models. *Proceedings of the* 24*th Annual Conference of the European Association for Machine Translation*, Tampere, 12-15 June 2023, 227-237.

[16] Kwok, H.L., Shi, Y., Xu, H., Li, D. and Liu, K. (2025) GenAI as a Translation Assistant? A Corpus-Based Study on Lexical and Syntactic Complexity of GPT-Post-Edited Learner Translation. *System*, **130**, Article ID: 103618. https://doi.org/10.1016/j.system.2025.103618

[17] Swain, M. (1993) The Output Hypothesis: Its History and Its Future. In: Cook, G. and Seidlhofer, B., Eds., *Principle and Practice in Applied Linguistics*: *Studies in Honour of H.G. Widdowson*, Oxford University Press, 125-144.

[18] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. https://doi.org/10.1023/a:1010933404324

[19] Kuhn, M. (2008) Building Predictive Models in *R* Using the Caret Package. *Journal of Statistical Software*, **28**, 1-26. https://doi.org/10.18637/jss.v028.i05