



开放科学  
(资源服务)  
标识码  
(OSID)

# 对话式搜索下学生群体的信息搜寻行为对搜索结果质量的影响研究

## ——基于 fsQCA 的组态分析

陆伟 张文波 张帆

武汉大学信息管理学院 武汉 430072

**摘要:** [目的/意义] 本文旨在揭示以学生为主的用户群体在对话式搜索环境下, 不同信息搜寻行为特征如何实现高质量的搜索结果。[方法/过程] 基于用户实验, 通过设计搜索任务与开发对话式搜索交互系统, 收集并量化了用户的查询、浏览、点击等信息搜寻行为特征。随后, 采用模糊集定性比较分析(fsQCA)方法, 探讨了在不同任务复杂度下, 信息搜寻行为与搜索结果质量之间的组态关系。[局限] 本文的样本规模有限, 行为变量选取也存在一定约束, 这可能影响研究发现的广泛适用性与覆盖范围, 需在未来研究中进一步拓展。[结果/结论] 在低任务复杂度条件下, 识别出三种能够关联较高搜索结果质量的行为特征组态; 而在高任务复杂度情境中, 仅有两组组态表现出此类效果。综合来看, 这些组态可归纳为快速浏览型、深度探索型、混合探索型和 AIGC 主导型四类信息搜寻行为模式。研究表明, 对话式搜索中信息搜寻行为与结果质量之间存在多重并发因果关系, AIGC 主导型的出现也反映出对话式搜索新范式对用户行为模式的影响与重塑。

**关键词:** 对话式搜索; 信息搜寻行为模式; 搜索结果质量; 任务复杂度; 模糊集定性分析

**中图分类号:** TP18; G252; G35

## The Impact of Students' Information-Seeking Behavior on Search Outcome Quality in Conversational Search: A Configuration Analysis Based on fsQCA

LU Wei ZHANG Wenbo ZHANG Fan

School of Information Management, Wuhan University, Wuhan 430072, China

**Abstract:** [Objective/Significance] This study aims to uncover how different information-seeking behavior patterns among

**基金项目** 国家自然科学基金青年科学基金项目“对话式搜索离线评价方法研究”(62202342)。

**作者简介** 陆伟(1974-), 博士, 教授, 主要研究方向为信息检索、数据智能、AI 治理等; 张文波(1999-), 硕士研究生, 主要研究方向为信息检索和用户行为; 张帆(1994-), 通信作者, 博士, 副研究员, 主要研究方向为信息检索、用户行为等, E-mail: fan.zhang@whu.edu.cn。

**引用格式** 陆伟, 张文波, 张帆. 对话式搜索下学生群体的信息搜寻行为对搜索结果质量的影响研究——基于 fsQCA 的组态分析[J]. 情报工程, 2026, 12(1): 86-96.

student users in conversational search environments lead to high-quality search outcomes. [Methods/Processes] Based on user experiments, we designed search tasks and developed an interactive conversational search system to collect and quantify behavioral features such as querying, browsing, and clicking. We then employed fuzzy set Qualitative Comparative Analysis (fsQCA) to explore the configurational relationships between information-seeking behaviors and search outcome quality under tasks of varying complexity. [Limitations] The limited sample size and the constrained selection of behavioral variables may affect the generalizability and coverage of the findings, which calls for further expansion in future research. [Results/Conclusions] Under low task complexity, three configurations of behavioral features were found to be associated with high search outcome quality, while only two such configurations emerged in high task complexity contexts. These configurations can be categorized into four behavior patterns: rapid-browsing, deep-exploration, hybrid-exploration, and AIGC-dominated. The study reveals multiple concurrent causal relationships between behavioral features and search outcomes, and the emergence of the AIGC-dominated pattern suggests that the new paradigm of conversational search is reshaping users' information-seeking behavior to some extent.

**Keywords:** Conversational Search; Information-Seeking Behavior Patterns; Search Outcome Quality; Task Complexity; Fuzzy-Set Qualitative Comparative Analysis

## 引言

在信息爆炸的时代，搜索引擎已成为人们获取信息的核心工具。然而，随着用户需求日益多样化和个性化，传统搜索引擎难以全面满足用户需求。以 ChatGPT 为代表的大语言模型的出现，加速了对话式搜索引擎的商业化落地，例如微软推出的新型搜索引擎“新必应”（New Bing）<sup>[1]</sup>。对话式搜索引擎作为一种新兴搜索方式，通过模拟人与人之间的自然对话，为用户提供了更为直观、便捷的信息获取体验。

信息搜寻行为是指用户为了实现某种目标，有目的地搜索信息以满足自身需求的行为，作为信息行为的重要组成部分，一直是学术研究的重点<sup>[2]</sup>。由于对话式搜索引擎在信息获取方式上的巨大变革，其深刻影响了用户信息搜索的各个环节。在传统的搜索过程中，信息搜寻行为受到用户自身特征、任务属性、情景要素等因素的影响，不同的搜索习惯与策略对搜索

效果会产生重要影响<sup>[3-4]</sup>。这些因素不仅决定了用户与搜索系统的交互方式，还影响搜索结果的呈现效果及用户的满意度<sup>[5]</sup>。例如，用户的搜索习惯可能影响查询的构建方式<sup>[6]</sup>，而搜索策略则决定了他们如何获取和筛选搜索结果<sup>[7]</sup>。此外，用户的搜索偏好可能影响其对搜索结果的接受度和满意度。

然而，在对话式搜索下，上述的影响机制尚没有过多的研究提及。本文从任务复杂度这一影响因素出发，揭示了对话式搜索环境下不同的信息搜寻行为特征与搜索效果之间存在的因果关联。同时，这些多元化的搜寻行为特征通过相互作用，表现出特定的信息搜寻行为模式<sup>[8]</sup>。本文旨在探究用户搜寻行为特征与搜索结果质量之间的复杂因果关系，通过模糊集定性比较分析的方法分析不同特征对搜索效果的共同影响，并归纳出相应的信息搜寻行为模式，以为对话式搜索系统的发展提供理论支持和实践指导。

## 1 对话式搜索下信息搜寻行为研究概述

现有关于对话式搜索用户行为的研究主要沿着两条路径展开：交互范式对比与新兴行为特征发现。在范式对比方面，学者们通过实证分析揭示了对话式搜索与传统搜索的显著差异。王若佳等<sup>[9]</sup>通过查询日志分析发现，对话式搜索中的检索式平均长度比传统搜索平均增加 5.61 个字符，且构建时间延长 8.92 秒，这源于用户更倾向于使用自然语言表达复合信息需求。范雨田等<sup>[10]</sup>进一步从多维度证实了这种差异，包括检索式重构频次降低、系统初始响应时间延长，但同时也发现在用户满意度和结果质量上无显著差异。卢新元等<sup>[11]</sup>基于使用对话式搜索引擎的问卷访谈数据，阐明了在信息查寻、信息选择、信息利用三个阶段的用户信息行为模型，进一步揭示了对话式搜索引擎环境下信息行为与传统信息行为的差异。在新兴行为特征发现方面，研究重点关注语言风格对用户感知的影响<sup>[12]</sup>。合适的语言风格不仅能增强可信度和拟人化程度，还能提升交互体验的感知质量<sup>[13]</sup>。例如，Xing 等<sup>[14]</sup>发现，老年用户更倾向于使用以“I”开头的请求，较少使用情态动词，并表现出独特的请求阐述方式，如在一个请求中包含多个查询或问题。Xu 等<sup>[15]</sup>开发的情绪支持对话（ESC）系统通过共情互动帮助用户缓解负面情绪，而 Hu 等<sup>[16]</sup>设计的语气感知聊天机器人能够生成富有激情和同理心的回应，甚至被用户认为比人类更具同理心。

总而言之，在对话式搜索下的信息搜寻行为，现有研究侧重于定性分析，通过问卷访谈

的方式进行编码，缺乏对用户信息搜寻行为的深入分析。尽管一些研究采用用户实验量化地收集了用户搜寻行为<sup>[9-10]</sup>，但仅通过传统的检验方法进行了方差分析，没有进一步分析对话式搜索中不同搜寻行为特征之间的关系。为了揭示不同的搜寻行为特征对搜索效果的复杂因果关系，本文将采用模糊集定性比较分析（fuzzy-set Qualitative Comparative Analysis, fsQCA）的方法，从组态分析的视角探究不同搜寻行为特征组合对搜索效果的影响。

## 2 研究方法

### 2.1 研究变量选择与测量

#### 2.1.1 搜寻行为特征

在传统的搜索过程中，查询、浏览与点击是三种重要的信息搜寻行为<sup>[17]</sup>。由于检索增强生成技术（Retrieval-Augmented Generation, RAG）的出现，对话系统的生成内容及其相关链接成为信息呈现的主要组成部分。尽管对话式搜索带来了交互范式变革，但用户基础认知需求（如信息修正、结果评估）仍会映射到查询调整（重构）、结果浏览与点击等行为上。在对话式搜索过程中，用户输入查询，浏览检查生成内容并点击内容中的相关链接进行信息溯源或补充，这一搜索流程仍然与传统的搜索场景类似。因此，借鉴近期对话式搜索相关工作<sup>[9-10]</sup>，我们仍然主要关注查询重构、浏览、点击这三种行为，旨在清晰地揭示技术演进对用户行为模式的影响。

在查询行为方面，查询式的构造会影响用

户的信息获取，查询重构类型也会影响用户搜索的有效性。查询相关的指标主要在于其表征查询式的复杂度和构造来源。本文借鉴了相关工作<sup>[18-19]</sup>，将查询复杂度（Query Complexity, *QC*）表征为用户在搜索过程中使用的词条数目。查询构造来源（Query Construction Source, *QS*）则用查询包含词条来自任务描述的占比来衡量。

在浏览行为方面，对话式搜索中，用户接触到的信息来源主要有两方面：系统回复的文本信息和文本信息下方的相关链接。用户既可以通过浏览并检查系统回复的文本内容来获取相关信息，也可以通过点击下方相关链接来满足信息需求。我们借鉴浏览行为的相关工作<sup>[20-21]</sup>，关注用户在浏览这两方面信息上所花费的时长，分别是SECP浏览时长（Browsing Duration of Search Engine Conversation Page, *BS*）和链接浏览时长（Browsing Duration of Links, *BL*）。浏览时长的统计单位为秒。

在点击行为方面，用户的点击行为主要是针对下方的相关链接而言的，点击下方的链接，从而可以浏览链接的全文。我们借鉴了点击行为的相关工作<sup>[22]</sup>，关注点击数量（Number of Clicks, *CN*）和点击有用性（Usefulness of Clicks, *CU*）两个重要方面，一个代表用户对链接数目的依赖程度，另一个则反映了用户对相关链接有用性的判别能力。

### 2.1.2 搜索效果

评估信息搜索效果的指标可以从系统和用户两个角度来衡量。系统角度的评估指标基本为离线指标，包括准确率、NDCG、BLEU、BERTScore等，反映检索结果的连贯性、相关

性与准确性<sup>[23]</sup>。用户角度的指标一般为在线指标，侧重于用户参与度、用户满意度、任务完成率以及认知负荷等指标。

本文关注用户角度的搜索效果，侧重以客观的方式来评价用户任务完成的质量来反映搜索是否成功<sup>[8, 24]</sup>，我们组织了与搜索任务相关的专家对用户的任务回答进行评分。具体地，首先，由于我们基于新闻和法律两个领域进行任务设计，所以要求专家需要具备相应的领域知识。我们通过线上招募的方式，将是否拥有新闻或法律相关的学士学位作为遴选标准，为新闻与法律两个领域分别邀请至少三名专家。为保证评价的客观性，专家们共同制定了标准任务回答，接着，专家对用户的任务回答进行公正、客观地评价，评价过程中，专家需要根据任务回答中知识点的个数和相关性进行赋分。之后，我们对专家的评分进行了内部一致性系数检验，绝对一致性水平超过0.75且显著（ $p < 0.001$ ）表明专家评分具有良好的可靠性。最终，将专家的评分取平均值表示为用户在该搜索任务下的搜索结果质量（Search Outcome Quality, *SQ*）。若用户最终提交的任务回答中知识点更完整且更准确，则其搜索效果好，任务回答质量得分高；若知识点的相关性和数量的表现都很差，则其搜索效果差，任务回答质量得分低。

综上所述，本文基于查询、浏览和点击行为来探究不同特征组合影响信息搜索效果的多重并发因果机制，即上述三个维度的行为特征会表现出何种组态，能够带来搜索效果的提升。搜寻行为特征和搜索效果分别构成条件变量和结果变量，如表1所示。

表 1 研究变量汇总

研究变量	测度
条件变量	
查询复杂度 (QC)	用户在搜索过程中输入查询所包含的词条数量
查询构造来源 (QS)	用户查询所包含的词条来自任务描述集合的占比
SECP 浏览时长 (BS)	在浏览 SECP 页面的系统回复上花费的时长
链接浏览时长 (BL)	在浏览链接上花费的时长
点击数量 (CN)	用户在搜索过程中点击引用链接的数量
点击有用性 (CU)	用户在搜索过程中点击链接的平均有用性
结果变量	
搜索结果质量 (SQ)	用户完成搜索任务的回答质量

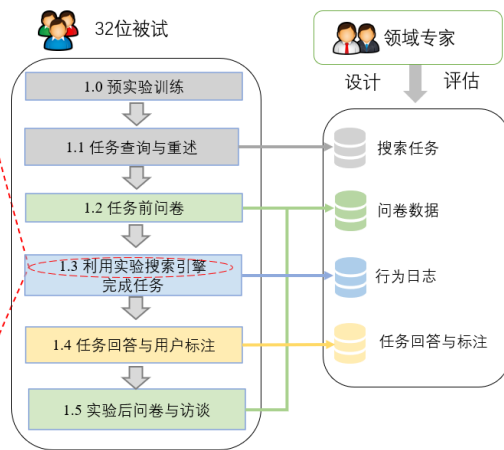
## 2.2 用户实验

为了揭示信息搜寻行为与搜索效果之间的

因果关系，我们按照图 1 所示的实验框架设计并进行了用户实验。实验问卷中的问题均采用 5 点的李克特量表设计。



(a) 实验交互界面



(b) 用户实验流程

图 1 实验框架

### 2.2.1 实验任务

信息搜寻行为的相关研究<sup>[25-26]</sup>表明，任务复杂度是影响用户行为及其搜索效果的重要因素。任务复杂度的一个重要维度是认知复杂性，而任务的认知复杂性与完成任务所需要的心理活动的类型（和多样性）有关。因此，本文根据 Urgo 等<sup>[27]</sup>对搜索任务所提出的设计原则，通

过控制知识维度不变，改变认知过程维度来改变任务复杂度。本文将实验任务根据任务复杂度分为两类，具体内容如表 2 所示。

在实验过程中，用户需要根据自身的专业背景选择两个领域知识较低的任务完成。为了避免任务分配顺序的影响，我们采取拉丁方的方式来为用户分配任务。

表2 用户实验任务

任务复杂度	任务情景	任务要求
低任务复杂度	假设你现在是一名新闻编辑，打算撰写一篇数据新闻，需要了解数据新闻的相关知识。	请深入理解数据新闻的形成过程并给出一个例子来阐明这一过程。找到相关内容，理解完成后，形成最终的任务答案。
	假设你现在是一名刚刚入职的律师，你想梳理一下判例法的相关知识。	请解释在法庭上如何有效地使用判例法并给出一个例子来论证这一过程。找到相关内容，解释完成后，形成最终的任务答案。
高任务复杂度	假设你现在是一名海洋环境保护者，现在手上有各种类型的数据资源，如文稿、音频、视频等，内容聚焦于海洋野生动植物的保护和可持续利用。	请设计一个合理的传播策略来呼吁大众保护海洋濒危物种，包括目标受众、内容类型、传播渠道等，然后描述实施该策略的步骤和方法，完成后，形成最终展示的策划案。
	假设你现在是一名活动筹划的负责人，近期需要举办一场模拟法庭辩论活动，帮助参与者锻炼法律推理、辩论和口头表达能力。	请设计一个模拟法庭辩论活动，包括角色分配、案件材料准备和模拟审判流程等，然后描述举办该活动的步骤和流程，设计完成后，形成最终的活动策划案。

### 2.2.2 实验对象

本次用户实验的被试主要是来自武汉大学的本科生。为了避免用户的领域知识，信息检索能力对搜索效果的影响，我们通过线上问卷发布招募信息，筛选出38位本科生作为被试。后对被试实验数据进行检查后，符合要求的数据中涉及32位被试者，其中男女人数比例约为2:3，年龄分布在19~23岁。每名被试需要完成不同复杂度的任务各一个，因而有效数据数量为64份。为了避免领域知识对搜索效果的影响，我们要求被试在每个任务开始前填写任务前问卷，问卷中包含被试对搜索任务的相关知识的熟悉程度。我们对被试报告的各个搜索任务的领域知识进行了检验，检验结果显示其中位数为0，均值约为0.7，方差约为1。这表明参与实验的被试所具备的领域知识水平较低。

### 2.2.3 实验交互界面

在我们的用户实验中，我们以New Bing的对话搜索页面为基础设计了对话式搜索交互界面，如图1(a)所示。我们删除了原有页面上与实验无关的组件或元素，添加了右侧的笔记区域用于用户在搜索过程中进行相关内容的保

存。同时，我们添加了对话完成按钮来退出搜索，方便其进行下一个搜索任务。

### 2.2.4 搜寻行为日志

为了记录被试的上述搜寻行为，我们开发了一个基于Chrome浏览器的扩展程序，该扩展程序将JavaScript注入参与者在搜索过程中访问的每个网页。我们进一步验证了Chrome扩展程序在实验过程中没有造成明显的网页延迟。这种记录方式对于用户行为的收集更加灵活，同时也不会干扰被试在搜索过程中的行为。

### 2.2.5 实验流程

用户实验流程如图1(b)所示。首先，在实验开始前(I.0阶段)，用户需要阅读相关的任务指导说明，同时，我们提供一个示例任务帮助用户熟悉实验操作流程。其次，在任务查询与重述阶段(I.1阶段)，用户需要阅读本次搜索任务并且记忆任务后进行重述，这一步骤的目的是确保用户理解任务。实验开始后(I.2阶段)，用户在开始搜索任务前需要填写任务前问卷，该问卷收集了用户对该搜索任务的熟悉情况、困难以及兴趣程度。接着，用户利用对话式搜索引擎完成搜索任务(I.3阶段)。完

成搜索后，用户需要对搜索任务进行回答，回答完成后，对对话内容进行标注，反馈在搜索过程中的点击链接的有用性（I.4 阶段）。最后，实验结束后（I.5 阶段），用户需要填写实验后问卷和进行访谈。

本文收集了任务前问卷，主要用于排除领域知识对搜索效果的影响。任务前问卷包含了用户对某一搜索任务的领域知识水平，我们需要根据任务前问卷保证本文使用的样本中用户的领域知识水平较低。实验后问卷及访谈收集

了用户对搜索任务的复杂度评价，用于验证本文的搜索任务设计是否合理。因此，后续实验中，我们使用了客观测量的搜寻行为特征和搜索效果数据。

### 2.3 变量统计与校准

#### 2.3.1 描述性统计

通过用户实验，不同任务复杂度下收集的搜寻行为特征及搜索效果的统计结果如表 3 所示。

表 3 研究变量的描述性统计

任务复杂度	统计量	搜索结果质量 (SQ)	查询复杂度 (QC)	查询构造来源 (QS)	SECP 浏览时长 (BS)	链接浏览时长 (BL)	点击数量 (CN)	点击有用性 (CU)
低复杂度任务	最大值	15.50	91.00	0.82	692.23	1329.43	22.00	3.67
	最小值	1.50	5.00	0.19	50.57	0.00	0.00	0.00
	均值	9.39	23.56	0.51	260.24	194.72	4.66	1.81
	标准差	4.14	22.38	0.18	171.69	265.10	4.96	1.17
高复杂度任务	最大值	18.50	155.00	1.00	886.52	906.31	33.00	4.00
	最小值	6.00	12.00	0.23	105.08	0.00	0.00	0.00
	均值	12.27	40.91	0.58	375.07	125.69	5.13	1.45
	标准差	2.64	27.41	0.19	198.07	178.49	7.33	1.08

#### 2.3.2 变量校准

本文采用集合隶属分数赋值进行变量校准，即将原始数据转化为集合论中的隶属分数。本文中的条件变量与结果变量均为连续型数据，需要进行校准，将数据转化为对应条件的模糊集隶属

分数。我们采用上四分位、中位数和下四分位数这 3 个临界点作为完全隶属、交叉点以及完全不隶属数值。两种任务复杂度下的案例（Low Task Complexity, LC; High Task Complexity, HC）中各研究变量的校准锚点，如表 4 所示。

表 4 研究变量的校准锚点

类型	变量	LC 案例	HC 案例
条件变量	QC	(11.000, 14.000, 27.250)	(25.750, 32.500, 44.750)
	QS	(0.381, 0.500, 0.714)	(0.454, 0.576, 0.690)
	BS	(118.857, 232.300, 321.378)	(227.806, 340.123, 457.681)
	BL	(39.526, 96.964, 252.031)	(20.047, 60.170, 171.532)
	CN	(1.750, 3.000, 6.000)	(1.000, 2.000, 5.250)
	CU	(1.200, 1.933, 2.812)	(0.491, 1.513, 2.175)
结果变量	SQ	(5.875, 10.500, 12.000)	(11.000, 12.000, 13.000)

### 3 实验结果与发现

#### 3.1 不同任务复杂度下的搜寻行为组态分析

与通常的fsQCA探究类似，在进行组态分析之前，我们需要对单一条件进行必要性判断，即检验某个单一条件及其非集是否为高搜索结果质量的必要条件。对必要性的判断一般为一致性指标。一般认为，当条件的一致性水平大于0.9时，该条件为结果的必要条件。本研究对各变量及其非集是否为高搜索结果质量的必要条件进行了检验，发现6个条件变量作为必要条件的一致性均小于0.9，说明了这些前因变量均不是搜索效果提高的必要条件。

在fsQCA研究中，一般要求一致性阈值高于0.75，案例频数阈值为总案例的1.5%。本研究参照已有研究，将一致性阈值设置为0.90，阈值频数设置为1，PRI（Proportional Reduction in Inconsistency）一致性设置为0.7，以确保数据的有效性。根据fsQCA通用的研究流程，本文以中间解为基准，对比简约解找到对应的相似组态，进行变量的逐一比对。若变量存在于简约解中，则该变量为核心条件；若该变量不存在于简约解中，则该变量为辅助条件。

##### 3.1.1 低任务复杂度

在低任务复杂度下，带来高搜索结果质量的搜寻行为组态包含LC-I、LC-II、LC-III三种，三个组态的一致性水平和总体一致性水平均高于0.85，如表5所示。

组态LC-I ( $\sim QC^*QS^*\sim BS^*BL^*CN$ )（组态路径中大写字母表示核心条件，小写字母表示辅助条件； $\sim$ 表示条件不出现，下同）：该行

为组态以低查询复杂度、高查询来源任务依赖度、低浏览SECP时长、高浏览链接时长、高点击数目为核心条件，点击有用性为无关条件，在低任务复杂度上取得较好的搜索效果。这一组态与传统搜索下的快速浏览型用户行为类似，这类用户倾向于通过提高查询来源任务依赖度来获取更加丰富的链接信息，从而通过大量的点击这些链接来完成任务。

表5 低任务复杂度下带来高搜索结果质量的行为组态

组态	LC-I	LC-II	LC-III
QC	⊗	⊗	•
QS	●	•	⊗
BS	⊗	•	●
BL	●	●	•
CN	●	⊗	●
CU		●	●
原始覆盖度	0.137	0.114	0.098
唯一覆盖度	0.086	0.056	0.062
一致性	0.911	0.905	0.867
解的覆盖度		0.269	
解的一致性		0.906	

注：●表示某条件出现为核心条件，•表示某条件出现为辅助条件；⊗表示某条件不出现为核心条件，⊙表示某条件不出现为辅助条件；空白表示组态路径中该变量可有可无。

组态LC-II ( $\sim QC^*qs^*bs^*BL^*\sim cn^*CU$ )表明在低复杂度下，带来高搜索结果质量的组态以低查询复杂度、高浏览链接时长、高点击有用性为核心条件，以查询来源任务依赖度较高、浏览SECP时长较多、点击数目较少为辅助条件。这一组态与传统的搜索下的保守型用户行为类似，他们并不会依赖点击链接的数量，而是更加关注其质量（相关性），他们往往会在少量有用链接内停留大量时间来获取信息，他们也会检查系统回复的文本信息，将其作为一种辅

助的信息获取手段。

组态 LC-III ( $qc^* \sim QS^* BS^* bl^* CN^* CU$ ) 表明, 以低查询来源任务依赖度、长浏览 SECP 时长, 高点击数目和高点击有用性为核心条件, 辅以较高查询复杂度、较长浏览链接时长时, 能够获得较好的搜索结果。这一组态与传统搜索不同, 用户不关注查询来源任务依赖度, 在浏览时长上更加依赖对话系统的生成内容。同时, 他们也会点击较多的链接, 但这些链接的有用性也较高。他们更倾向于将点击链接看作一种辅助的信息获取手段。当用户将系统生成的内容作为获取信息的主要方式时, 他们需要思考如何将上下文信息融入查询中来有效地激发系统的生成潜力, 从而导致了该组态下查询构造来源和浏览时长分布的变化。

### 3.1.2 高任务复杂度

在高任务复杂度下, 带来高搜索结果质量的搜索行为组态有 HC-I、HC-II 两种, 其一致性水平及总统一致性水平均超过 0.90, 如表 6 所示。

表 6 高任务复杂度下带来高搜索结果质量的行为组态

组态	HC-I	HC-II
$QC$	⊗	●
$QS$	●	⊗
$BS$		●
$BL$	●	
$CN$	•	●
$CU$	⊗	⊗
原始覆盖度	0.190	0.131
唯一覆盖度	0.149	0.090
一致性	0.955	0.924
解的覆盖度	0.280	
解的一致性	0.950	

注: ● 表示某条件出现为核心条件, • 表示某条件出现为辅助条件; ⊗ 表示某条件不出现为核心条件。

组态 HC-I ( $\sim QC^* QS^* BL^* cn^* \sim CU$ ) 表明, 在进行高任务复杂度的搜索任务时, 以低查询复杂度、高查询来源任务依赖度、长浏览链接时长和低点击有用性为核心条件, 辅以较高点击数目的组态, 能够带来高搜索结果质量。该组态介于 LC-I 与 LC-II 之间, 与 SECP 浏览时长无关, 但主要区别在于其点击行为。在高任务复杂度下, 一方面用户减少了链接的点击数量, 另一方面, 点击链接的平均有用性也下降了。

组态 HC-II ( $QC^* \sim QS^* BS^* CN^* \sim CU$ ) 表明, 以高查询复杂度、低查询来源任务依赖度、长浏览 SECP 时长、高点击数目和高点击有用性为核心条件的组态在高任务复杂度下能够带来较好的搜索结果。该组态与 LC-III 类似, 与链接浏览时长无关, 主要区别在于链接的有用性。由于任务复杂度增加, 用户很难判断哪些链接有用, 往往需要点击链接后才能做出判断。但点击有用性较低, 这表明当前的对话式搜索提供的链接对于用户而言仍存在不足。

### 3.1.3 稳健性检验

为了进一步验证上述研究结果的稳定性, 将一致性门槛值提高到 0.9, 案例频数门槛值依旧以总案例数的 1.5% 为标准, 设为 1。在进行阈值设置后, 同样为了保证数据的有效性, 对于组态中原始一致性高于 0.9, 但 PRI 一致性明显低于 0.8 的条件组合结果手动设置为 0, 重新计算在不同任务复杂度下的简约解和中间解。根据上述稳健性检验的解决方案, 可获得新组态。稳健性检验的组态结果中, 仅在低复杂度下少了一种组态, 其余并未出现不同。这并不影响原组态的解释。

## 4 讨论

本文揭示了对话式搜索中用户信息搜寻行为特征及其与搜索结果质量的复杂关系。研究发现,高质量搜索结果由特定的行为组态驱动,这些组态与任务复杂度密切相关:在低复杂度任务中,用户行为呈现多元化特征,包括传统链接驱动的快速浏览型(LC-I)和深度探索型(LC-II),以及新兴的AIGC主导型(LC-III);而在高复杂度任务中,行为模式则收敛为更精细化的混合探索型(HC-I)和AIGC主导型(HC-II)。其中,混合探索型指HC-I由LC-I与LC-II混合形成。这种从“多元并存”到“策略收敛”的转变,本质上反映了信息加工深度与认知负荷的平衡机制——当任务复杂度提升时,用户会自发淘汰低效的粗放式策略(如泛化链接浏览),转而采用更集约化的信息处理方式,或通过生成内容优化链接筛选效率,或直接依托AIGC完成知识整合。值得注意的是,传统链接驱动模式(LC-I/II、HC-I)与新兴AIGC驱动模式(LC-III、HC-II)的共存,表明对话式搜索并非简单替代而是拓展了用户的搜寻行为模式。

## 5 结论与不足

本文基于组态分析视角,深入探讨了信息搜寻行为特征与搜索结果质量之间的复杂因果关系。结果表明,搜索效果的提升并非依赖单一行为特征,而是多种特征间相互协同的结果。在低任务复杂度下,用户表现出快速浏览型、深度探索型和AIGC主导型三类行为模式;而在高任务复杂度下,则主要呈现混合探索型和AIGC主导

型两类模式。尤其值得关注的是,AIGC主导型在不同任务复杂度下均表现出较好的搜索效果,显示出对话式搜索在一定程度上对用户行为产生了重塑作用。这些发现为理解用户在对话式搜索中的行为模式提供了新的视角。

尽管本文识别了获得高质量搜索结果的行为模式,但仍存在一定局限。首先,样本规模有限,可能影响研究发现的普适性。其次,行为特征指标的选取虽具代表性,但难以全面覆盖信息搜寻行为的多样性,可能导致行为组态的覆盖度不足。最后,实验任务类型与主题相对集中,可能限制了行为差异的全面呈现。未来研究可通过扩大样本、丰富任务设计及引入更多行为特征来增强研究的稳健性与外部适用性。

## 参考文献

- [1] 李爱琳. 微软上线新版对话式必应搜索引擎新时代已到来 [N]. 第一财经日报, 2023-02-09(A05).
- [2] WILSON T D. Human information behavior[J]. Information Science, 2000, 3(2): 49-55.
- [3] XIE I, JOO S. Factors affecting the selection of search tactics: tasks, knowledge, process, and systems[J]. Information Processing & Management, 2012, 48(2): 254-270.
- [4] LU K, JOO S, LEE T, et al. Factors that influence query reformulations and search performance in health information retrieval: A multilevel modeling approach[J]. Journal of the Association for Information Science and Technology, 2017, 68(8): 1886-1898.
- [5] ANAND A, CAVEDON L, JOHO H, et al. Conversational search (Dagstuhl Seminar 19461)[J]. Dagstuhl Reports, 2020, 9(11): 34-83.
- [6] 刘畅, 宋筱璇. 信息搜索用户检索式行为研究综述 [J]. 图书情报工作, 2017, 61(16): 122-134.
- [7] DOSSO C, TAMINE L, PAUBEL P V, et al. Navigational and thematic exploration-exploitation trade-offs during web search: effects of prior domain

- knowledge, search contexts and strategies on search outcome[J]. *Behaviour & Information Technology*, 2024, 43(10): 2232-2258.
- [8] 赵一鸣, 李倩, 邱雨蒙, 等. 用户搜索路径特征对信息搜索效果的影响研究——基于 fsQCA 的方法 [J]. *情报学报*, 2023, 42(1): 103-112.
- [9] 王若佳, 范科鸣, 刘智锋, 等. 生成式人工智能环境下用户信息检索式行为研究 [J]. *数据分析与知识发现*, 2024, 8(Z1): 20-30.
- [10] 范雨田, 王晓慧. 用户视角下大语言模型搜索引擎与传统搜索引擎信息检索行为及成效对比研究 [J]. *信息与管理研究*, 2024, 9(5): 24-35.
- [11] 卢新元, 张进澳, 雷晓鹏. 人工智能生成内容环境下用户信息行为研究——以对话式搜索引擎为例 [J]. *情报理论与实践*. 2023, 46(12): 84-92.
- [12] JAKIC A, WAGNER M O, MEYER A. The impact of language style accommodation during social media interactions on brand trust[J]. *Journal of Service Management*, 2017, 28(3): 418-441.
- [13] HUANG S H, LIN Y F, HE Z, et al. How does conversation length impact user's satisfaction? A case study of length-controlled conversations with LLM-powered Chatbots[C]// *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 2024: 1-13.
- [14] XING Z, YUAN X, MOSTAFA J. Age-related difference in conversational search behavior: Preliminary findings[C]// *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*. 2022: 259-265.
- [15] XU Z, CHEN D, KUANG J, et al. Dynamic demonstration retrieval and cognitive understanding for emotional support conversation[C]// *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2024: 774-784.
- [16] HU T, XU A, LIU Z, et al. Touch your heart: a tone-aware chatbot for customer care on social media[C]// *Proceedings of the 2018 CHI conference on human factors in computing systems*. 2018: 1-12.
- [17] MAO J, LIU Y, KANDO N, et al. How does domain expertise affect users' search interaction and outcome in exploratory search?[J]. *ACM Transactions on Information Systems*, 2018, 36(4): 1-30.
- [18] LIU C, ZHANG X, HUANG W. The exploration of objective task difficulty and domain knowledge effects on users' query formulation[J]. *Proceedings of the Association for Information Science and Technology*, 2016, 53(1): 1-9.
- [19] 刘畅, 宋筱璇. 信息搜索用户检索式行为研究综述 [J]. *图书情报工作*, 2017, 61(16): 122-134.
- [20] MAO J, LIU Y, ZHOU K, et al. When does relevance mean usefulness and user satisfaction in web search?[C]// *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2016: 463-472.
- [21] JOACHIMS T, GRANKA L, PAN B, et al. Accurately interpreting clickthrough data as implicit feedback[C]// *ACM SIGIR Forum*. 2017, 51(1): 4-11.
- [22] JUNG S, HERLOCKER J L, WEBSTER J. Click data as implicit relevance feedback in web search[J]. *Information Processing & Management*, 2007, 43(3): 791-807.
- [23] DERIU J, RODRIGO A, OTEGI A, et al. Survey on evaluation methods for dialogue systems[J]. *Artificial Intelligence Review*, 2021(54): 755-810.
- [24] LIU M, LIU Y, MAO J, et al. "Satisfaction with failure" or "unsatisfied success": Investigating the relationship between search success and user satisfaction[C]// *Proceedings of the 2018 World Wide Web Conference*. 2018: 1533-1542.
- [25] CHOI B, WARD A, LI Y, et al. The effects of task complexity on the use of different types of information in a search assistance tool[J]. *ACM Transactions on Information Systems*, 2019, 38(1): 1-28.
- [26] KELLY D, ARGUELLO J, EDWARDS A, et al. Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework[C]// *Proceedings of the 2015 International Conference on the Theory of Information Retrieval*. 2015: 101-110.
- [27] URGO K, ARGUELLO J, CAPRA R. Anderson and Krathwohl's two-dimensional taxonomy applied to task creation and learning assessment[C]// *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*. 2019: 117-124.

(责任编辑: 浦墨)