

# 人工智能语义分析技术在用户研究中的应用

刘再行, 刘毅, 郜洵  
(广州美术学院, 广州 510260)

**摘要:** **目的** 实现用户研究中语言材料分析的半自动化。**方法** 首先通过文献研究, 明确了访谈类语言材料以性质分析为主要方法, 问卷调查和网络评价类语言材料以量化统计分析为主要方法。然后通过方法比较, 总结人工智能语义分析技术对于用户研究语言材料的八种常用能力: 中文分词、命名实体识别、相似度分析、词频统计、情感倾向分析、文本分类、信息抽取、自动文摘, 以及总结了以此为基础的语料分析处理流程。最后以“某品牌微蒸烤一体机交互设计创新”项目为例, 介绍了应用人工智能语义分析技术对用户研究中的语言材料进行分析的具体执行方法。**结果** 测试验证人工智能语义分析技术在用户研究中的应用的有效性。**结论** 在设计领域应用人工智能技术具有巨大的潜力, 通过总结针对性的算法, 可以引入人工智能技术替代部分需要人智分析的工作, 从而提高工作效率。

**关键词:** 用户研究; 人工智能; 语义分析

**中图分类号:** TB472 **文献标识码:** A **文章编号:** 1001-3563(2020)18-0053-07

**DOI:** 10.19554/j.cnki.1001-3563.2020.18.007

## Application of Artificial Intelligence Semantic Analysis Techniques in User Research

LIU Zai-xing, LIU Yi, GAO Xun  
(Guangzhou Academy of Fine Arts, Guangzhou 510260, China)

**ABSTRACT:** The paper aims to semi-automate the analysis of linguistic material in user research. Through the literature study, it was clarified that qualitative analysis was the main method for interview-type language materials. Quantitative statistical analysis was the main method for language materials from questionnaires and web-based evaluation. Then, by comparing the methods, eight common capabilities of AI semantic analysis techniques for user study language materials were summarized: Chinese words segmentation, identity of named entity, similarity analysis, word frequency statistics, sentiment analysis, text classification, information extraction, automatic abstraction. This paper summarized the semantic data analysis processing flow based on this approach. With the “Interaction design innovation of a brand of micro-steaming machine” as an example, it introduced the specific implementation method of using artificial intelligence semantic analysis technology to analyze language materials in user research. The effectiveness of applying artificial intelligence semantic analysis techniques in user research was tested and verified. The application of AI technology in design has great potential. By summarizing the targeted approach, AI technology can be introduced to replace some of the tasks that require human intelligence analysis and improve efficiency.

**KEY WORDS:** user research; artificial intelligence; semantic analysis

随着体验经济时代的来临, 产品设计更多地开始注重用户的情感需求, 用户体验设计越来越受到了国内外学者和企业界的重视<sup>[1]</sup>。用户体验设计是从产品

用户的角度出发, 关注用户使用产品的全过程<sup>[2]</sup>, 因此以了解用户使用体验为目的的用户研究显得尤为重要。用户研究通过对用户行为、认知特征和满意度

收稿日期: 2020-08-01

基金项目: 广州美术学院学术提升计划项目(20XSC43)

作者简介: 刘再行(1977—), 男, 广东人, 广州美术学院讲师, 主要研究方向为信息与交互设计及工业设计用户体验测试评价。

的调研来了解用户的实际需求,在该过程中会获取大量的研究数据。其中语言材料是用户研究中最常见的基础数据,来源包括访谈、用户反馈、网络评价、街头拦截或焦点小组等。虽然用户研究中收集语料数据的方式较多,但是成熟的分析方法并不多<sup>[3]</sup>,这是因为语言材料数据属于非结构化数据,内容庞杂,使用人工整理和分析的方式耗时耗力,研究者无法快速从中获取有价值的内容。

近年来,随着人工智能和大数据技术的应用发展,深度学习技术已经在图像识别、语音识别和自然语言处理等多个领域取得了突破。其中面向自然语言处理的语义分析技术在研究人类的交流方式方面起着重要的作用,它不仅影响了人工智能技术的发展,而且还影响了人机交互的方式<sup>[4]</sup>。利用人工智能语义分析技术对用户研究所获取的语言材料进行辅助分析,能够显著地提高用户研究结果的分析效率,甚至可以挖掘出人工分析难以发现的重要信息。

## 1 人工智能语义分析方法的引入

### 1.1 自然语言处理以及语义分析

自然语言处理(Natural Language Processing,简称NLP)就是利用电子计算机为工具,对人类特有的自然语言的信息进行处理和加工的技术<sup>[5]</sup>。对语言材料进行语义分析是自然语言处理的基础,涉及语言学、计算语言学、机器学习以及认知语言等多个学科,是人工智能的一个分支。语义分析的目标就是通过建立有效的计算分析模型(包括数学统计模型和深度学习模型),实现词汇、句子和章节的自动语义分析,理解语言文本表达的含义。其中词语的语义分析是基础,常用的知识表示方法有:语义场、语义网络、概念图与本体论<sup>[6]</sup>。

自然语言中存在着大量无法用简单逻辑关系进行演算的情感和思想,这种复杂性决定了计算机识别和理解自然语言的目标任重而道远。从自然语言处理技术的发展历史来看,相关的研究可分成符号派(Symbolic)和随机派(Stochastic)两个阵营。符号派采用基于语言学定义规则的分析方法,着重研究推理和逻辑语言理解模型问题,被称为“理性主义”。随机派主要研究随机和统计算法对大规模的语料库进行识别模型计算,被称为“经验主义”<sup>[7]</sup>。

早期符号派所作出的贡献较大。在20世纪70年代初,美国计算语言学家香克和德国学者特雷尔分别提出描述句义和语义的方法以及语义场的概念。在此基础上国内外研究者提出了基于词语间相互关系的义场的描述,基本奠定了后续研究的方向。在汉语的语法结构中,构成句子的基础单元是字而不是词,对计算机识别来说汉语句子中没有表示词语边界的分隔符号,因此要还原句意必须先提取句子中的词语,

这被称为汉语自动分词<sup>[8]</sup>。然后在此基础上进一步根据自然语言的三个层级——词语、句子和篇章来分别进行分析方法研究,即首先从词语的层级上进行语义分析,包括词语的词性分析和词语的语义相似度分析等,然后再进一步进行句子分析和篇章语义分析的研究<sup>[9]</sup>。

理性主义的技术方法虽然可以完成对单个句子的基本分析,但是对于复杂句式的理解,乃至对于整个段落或篇章的分析质量不高。随机派引入的统计学方法为自然语言理解的研究方法提供了新的思路,但是早期效果未如人意,直到最近几年深度学习技术的出现使得情况大有改观。依靠统计学方法和深度学习算法,机器能够从大量数据中进行自我学习,得出语言使用规律,相比传统方法在效率和准确性上均有很大的提升。特别是在汉语语义分析的应用中效果提升尤为突出。具体来说就是利用深度学习模型的自动提取特征取代传统机器学习中需要人工构建特征的过程,大大地提高了特征模型的有效性。

### 1.2 语义分析技术以及应用

通过理解上述自然语言处理技术发展的历史沿革,可以看到语义分析技术在以深度学习为代表的机器学习方法的介入下得到了跨越式的发展。通过研究文献发现,已有不少学者对智能语义识别技术的应用方法做了研究,而且根据不同的应用目的,其方法和分析步骤也有差异。例如,赵京胜等人从获取识别模型的角度对自然语言的处理流程作出总结,大致可分为:获取语言材料、对语料进行预处理、把分词后的字和词特征化、模型识别训练和对建模后的效果进行评价<sup>[10]</sup>。刘挺等人提出的一种以基础语言技术支持为目的的“语言技术平台”,其处理流程依次为分词、词性标注、命名实体识别、句法分析、词义消歧和语义角色标注<sup>[11]</sup>。赵文清等人以微博热点话题等情感分析为目的进行话题聚类,其处理流程依次为内容筛选、分词预处理、抽取词频特征矩阵、矩阵降维、聚类分析和提取话题<sup>[12]</sup>。由此可以把应用智能语义识别技术的整体思路归结为:先分解后分析,并从字词开始,向句子、段落和篇章逐层递进进行语义识别。

参照上述概括智能语义分析技术的应用与分析方法,对涉及的技术进行拓展研究,可以进一步总结出主要分析方法的技术原理、局限性与使用方式,智能语义分析方法以及用途,见表1。

### 1.3 用户研究语言材料的分析方法

目前用户研究的社会科学实证研究方法包括定量研究和定性研究两大类,胡飞引入了日记法、问卷法、观察法和访谈法等常见的社会科学调查研究方法,从而将“民族学”加以简化应用于设计的用户研究<sup>[18]</sup>。关于用户研究语言材料的分析方法,胡飞等人也进一步进行了系统性综述:(1)从概括总结性的角

表 1 智能语义分析方法以及用途  
Tab.1 Methods and uses of intelligent semantic analysis

分析方法	描述	处理方法	局限性	使用方式
中文分词	自动分词是中文自然语言处理中最基础的环节	基于词典的分词算法，基于统计的机器学习分词算法	对歧义词和新词的识别错误率高	建立特定的应用词库，建立停用词字典
命名实体识别 <sup>[8]</sup>	识别文本中具有特定意义的实体	基于规则的方法，基于统计机器学习的方法	要先完成分词，中英文交替使用	识别特定范围的实体，如人名、地名等
相似度分析 <sup>[13]</sup>	通过计算短文本相似度数值来判断文本相似度	基于语料库，基于世界知识，基于字符串，基于混合方法	不适用于长文本，对属性集和算法有一定依赖	针对短句进行相似度分析
词频统计 <sup>[14]</sup>	用以评估词语在特定文件集中的重要程度	全量统计，指定统计，字频统计	属于非智能化的机械统计	在剔除停用词基础上对特定词汇进行词频统计
情感倾向分析 <sup>[15]</sup>	识别主观文本的情感极性，分为正面、负面和中性并计算极性强度	基于情感词库和机器学习方法	多观点篇章的细分、聚类和消歧有待提高	对文本进行情感倾向性聚类分析
文本分类 <sup>[13]</sup>	将给定文档分类为若干类别中的一个或多个	传统机器学习方法，深度学习方法	不同文档对应不同模型效果差异较大	要先通过测试来筛选合适的分类算法
信息抽取 <sup>[16]</sup>	从自然语言文本中抽取指定类型的实体等事实信息，并形成结构化数据	知识工程方法和自动训练方法，基于深度学习的方法	信息抽取技术并不试图全面理解整篇文档	格式化后的文本将更容易统计以及进行文本挖掘
自动文摘 <sup>[17]</sup>	利用计算机按照某一规则自动地对文本信息进行提取，集成简短摘要	机械文摘和理解文摘	文摘质量不高	作为语料分析的辅助性工具

表 2 用户研究语言材料的智能语义分析  
Tab.2 Intelligent semantic analysis of user research language material

分析方法	分析能力	应用对象	应用目的
中文分词	建立应用词库	检查补充业务词库	提高语义分析正确率
命名实体识别/相似度分析/词频统计/情感倾向/文本分类	聚类分析/情感分析	用户评价/问卷调查/头脑风暴/用户评价/话语分析	亲和图/辅助编码
信息抽取	观点抽取	用户访谈	质性内容分析/诠释标注语料/撰写访谈概要
自动文摘	摘要抽取	用户访谈	叙事分析/会话分析/总体分析

度分析总结了设计学与用户研究中的常用文本分析方法，例如对事实进行有机组合与归纳的亲和图法等；（2）归纳了用户访谈的语言材料分析的基本流程，其中关键步骤为“提取标记语料并编码”、“诠释标注语料”、“无关语的剔除”和“撰写访谈概要”<sup>[3]</sup>。此外，胡岷也介绍了当前在国内外研究人员在进行用户研究时惯常使用的研究方法，包括问卷调查法、访谈法、小组讨论法、实验法、观察法、出声思维法和记录分析法七种研究方法<sup>[19]</sup>。

本文通过对这些方法的研究，结合人工智能语义进行综合分析，得出在分析方法上可以采用语义分析技术的应用关系，用户研究语言材料的智能语义分析见表 2。

2 引入智能语义分析技术的用户研究文字材料的分析流程

2.1 分析流程和相关处理方法的原理

基于对引入智能语义分析技术对用户研究语言材料进行分析的各种处理方法的研究，可以把相关过程归结为两个主要步骤：简化原始语料数据；抽取核心信息进行归类。智能语义分析技术对用户研究语言材料进行分析的基本处理流程，见图 1。

1) 预操作。为了提高后续人工智能文本处理的准确性和效率，针对语言材料的业务内容建立更准确的业务词库与停用词库。根据研究对象的业务性质，

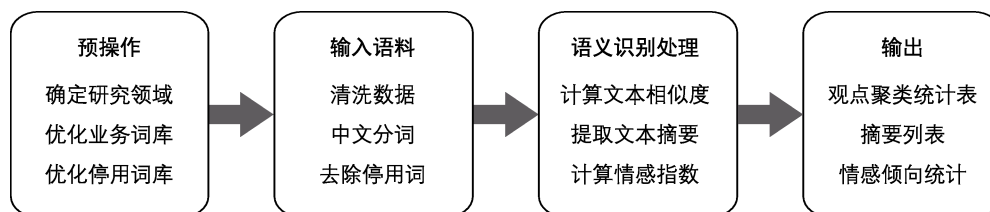


图 1 智能语义分析技术对用户研究语言材料进行分析的基本处理流程

Fig.1 Basic processing flow of analyzing user research language material by intelligent semantic analysis techniques

可以使用现成的中文褒贬义词和按行业的分类词库等现有业务词库与停用词库。如果测试发现现有词库的分词准确性较差,还可以根据初步的分词统计概率加入新词,或者在有条件的情况下训练自有的业务词库。停用词库主要包括副词、语气助词和连接词等不代表实质含义的用词。

2) 清洗数据。用于语义识别的数据清洗是指对文本数据集中的无用字符数据进行检测并去除的过程。清洗数据通常分为两个层级的操作:去除对语义识别造成干扰的非法字符;去除对本次研究无用的词语。例如引起歧义的用词和停用词库内的词语等。

3) 中文分词。中文分词的方法主要分为基于词典和基于理解两类方法。基于词典的方法核心是将句子内的文字与预先建立的词典表进行迭代对照,把能完成匹配的词语定为分词。预操作中确立的分词字典、切分规则和匹配顺序是提高分词准确性的关键。基于理解的分词方法就是在基于词典分词基础上的同时判断句法和语义是否存在歧义,进而进行准确分词。

4) 文本相似度的计算。语料经过正确的分词之后才可以进行文本相似度的计算。其主要原理是通过统计语料中 TF (词频) 和 IDF (逆文档频率) 两个参数的特征向量值来提取文本特征模型。TF-IDF 特征值的表达式为:  $TF-IDF(i, j) = TF_{i,j} \times IDF_i$ , 两篇语料的相似度为 1, 特征向量夹角为  $0^\circ$ 。基于该模型定义,使用传统机器学习算法中诸如 NB、RF、SVM 和 KNN 等分类的模型都可以用于语料的聚类分析。此外, FastText、TextCNN、TextRNN 和 Attention 等深度学习文本分类模型在不同的任务上会有不同的表现效果,在实践上需要根据试验情况来进行选择。

5) 情感分类。情感分类是语义分析技术应用中较为成熟的场景,例如商品评价等,可以对优化产品设计和功能更新迭代起指导作用。通过评价语料的情感倾向,可以统计出产品在各个维度的用户情感指数,从而明确如何改进产品。情感倾向计算可以采用基于词典的传统方法,也可以采用基于人工智能的方法。近年来基于深度学习的情感倾向分析技术获得了较大的进展,具有精度高、通用性强和不需要情感词典等优点。

6) 提取文本摘要。文档摘要提取算法通过寻找包含信息最多的词句来创建整个文档的内容摘要。通

通过对句子进行分词计算加权直方图,统计句子的得分并提取分数最高的句子作为摘要。

根据上述语义计算的原理可以把相关技术分为数学模型法和深度学习法两类。数学模型法主要是通过现有词典和数学模型算法来完成不同文本的特征计算;而深度学习法则是通过预先对大量数据应用神经网络算法进行学习训练,形成识别模型。两种方法各有优点,词典法应用较为成熟并且可以通过扩充词库提高准确率,而深度学习法则针对特定的应用领域有较好的性能表现。

## 2.2 人工智能语义的分析应用开发简介

针对特定目的的人工智能语义分析技术的应用,一般通过开发基于深度学习引擎的系统应用或开发基于自然语言处理开放平台接口的功能应用来实现。前者主动性较强,使用较为灵活,可以根据研究对象语料的特殊性选择不同类型的算法或创造新算法,具有一定的技术研发性质。后者则可以应用于相对通用性的语料分析,开发成本较低,但是只能适用于常规的业务领域。目前,国内提供语义识别开发接口的平台主要有百度 AI 开放平台、阿里云平台、腾讯 AI 开放平台和讯飞开放平台等。这些平台各自提供了常用的自然语言处理方法的接口,通过调用这些接口可直接实现中文文本分词、情感倾向分析、评论观点抽取和新闻摘要等功能。

这些技术平台为人工智能语义分析的应用提供了成熟且便利的条件。应当注意,各技术平台以及各开发接口对于不同业务场景的准确率有差异,需要对所研究的语料对象进行深入分析,根据业务特征甄选和创新效果更优的算法,并通过增加业务词库和停用词库的方式来提高效能。

## 3 应用实例

采用上述总结的方法,笔者对某品牌的微蒸烤一体机交互设计优化项目的用户研究语言材料数据进行了辅助分析。该项目旨在针对一款已经上市的微蒸烤一体机产品的交互界面设计做优化,用户研究基于具有实际使用经验的用户进行,并获取了大量的语言材料,包括“报修反馈”(1002 行,5 列)、“问卷调查记录”(305 行,266 列)、“用户访谈”(1230 行,2 列)和“在线用户评价”(4741 行,17 列)四个数

据表。这些语言材料数据量大和非结构化，无法直接进行统计分析，采用人工方式逐条阅读理解较为耗时，使用人工智能语义分析技术对其进行处理能够加快相关进程。

3.1 预处理

初始数据库包含了四个数据表合计 169 197 个语料记录，数据来源复杂，存在不完整和冗余的数据，并且不同来源的数据也需要分别做有针对性的预处理。（1）删除空记录、仅有语气助词与仅有无意义符号（所有数据表）。（2）删除标记为“故障”等的与本研究内容无关的技术问题记录（“报修反馈”表）。（3）删除无需参与分析的单选和多选字段（“调查问卷记录”表）。（4）空评论、空用户评分或“默认反馈”记录（用户评价记录）。

本文通过使用正则表达式配合停用词库匹配的模式，分别对四个数据表进行了数据清理。清理提高了数据质量，使数据分析的结果更加可靠。最终数据库的清理版本减少到 113 281 个语料数据记录字段。

3.2 补充专业词库并检验分词准确率

从前文分析可知，引入适用的业务词库对语义分析技术的效果有较明显的提升。本文通过抽样验证的方法，对所引入的词库的业务匹配程度进行检验，在 Jieba 分词库基础上逐步补充自定义词，获得针对本文研究内容效果最优的“分词器”。对已获取的语料数据随机抽取十句进行分词处理，经人工验证准确率为 96.7%，该“分词器”可以有效地处理本次研究的内容。优化后的“分词器”测试结果见表 3。

3.3 对“问卷调查”数据表的开放式问题回答语料进行分析

该数据表的语料数据语言以短句子为主，但存在一定的口语化倾向，含义相似的内容在用词、表达顺序和用词多寡等方面都会有一定的差异。针对这一语料特征，笔者采用基于 Word2vec 构建的中文词向量

库<sup>[19]</sup>TF-IDF 加权余弦相似度计算的算法获得两个句子的相似度。主要分析过程描述如下。

1）对需要比较相似度的句子进行分词，并过滤停用词。

2）获取所有分词的词向量，并以每个词的 TF-IDF 为权重，对句子中的所有词的词向量进行加权平均，获取句子向量。其向量加权平均公式为：

$$V = \sum_{i=0}^n tf-idf(d_i, v) W_i$$
，其中  $d_i$  为为第  $i$  个分词， $v$  是当前句子， $W_i$  为第  $i$  个分词的词向量。

3）求出两个句子的句子向量余弦值来比较两个句子相似度。其余弦公式为：Similarity $V_1, V_2 = \cos \theta = \frac{V_1 \cdot V_2}{\|V_1\| \|V_2\|}$ ，其中  $V_1, V_2$  分别为上述第二条中计算的两个句子的句子向量。

4）交叉计算文本的相似度得分，对所有回答进行聚类分析。

本文选取其中一个简短的案例来展示该处理过程，本算法对“问卷调查”语料进行迭代分类的过程见图 2（数值为算法返回的短文本相似度计算结果，取值[0,1]，分数越高说明相似度越高）。可以看到经过三次程序自动筛选之后，把“购买本产品用户的家庭结构（其他）”这个开放式问题的回答分为了五类情况：一家四口、独居、四代同堂、会所和其他。该方法可以快速地把问卷调查内原本无法参与统计的开放性问题转化成为单选题数据，作为后续统计分析的新选项。

3.4 对“在线用户评价”数据表的语料进行分析

在线用户评价的数据来源于各大电商平台上的购买用户评价，这类语料内容庞杂，大部分数据价值不高，如果工人逐条检阅效率较低。考虑到本研究的目的是从评价中获取用户关于产品体验的描述，采取语义分析的方法过滤出有研究价值的语料。采用对比用户评分和文本情感倾向差异的策略来发掘有研究

表 3 优化后的“分词器”测试结果（部分）  
Tab.3 Test results of the optimized “word divider” (partial)

序号	输入语料	分析结果
1	显示屏有水蒸气，多的时候凝结成水，要求尽快跟进处理	显示屏/有/水蒸气/，/多/的/时候/凝结/成/水/，/要求/尽快/跟进/处理/
2	蒸汽功能使用 30 s 就显示缺水，但水箱是有水的，具体请电联核实，谢谢	蒸汽/功能/使用/半小时/就/显示/缺水/，/但/水箱/是/有/水/的/，/具体/请/电联/核实/，/谢谢/。
3	我们家橱柜设计得有一点问题，然后我把门把手给卸掉了，在取出食物的过程中不会被烫到，玻璃是防烫的不会烫着，都戴着手套	我们/家/橱柜/设计/的/有/一点/问题/，/然后/我/把/门把手/给/卸/掉/了/，/在/取出/食物/的/过程/中/不会/被/烫/到/，/玻璃/是/防/烫/的/不会/烫/着/，/都/戴着/手套/。
4	插头在后面，不方便拔掉，待机是亮的，会有耗电量。建议不用时可以自动断电	插头/在/后面/，/不方便/拔掉/，/待机/是/亮/的/，/会/有/耗电量/。/建议/不用/时/可以/自动/断电/。
5	相对竞品价格低，品质	相对/竞品/价格/低/，/品质/高/



问题：您的家庭结构

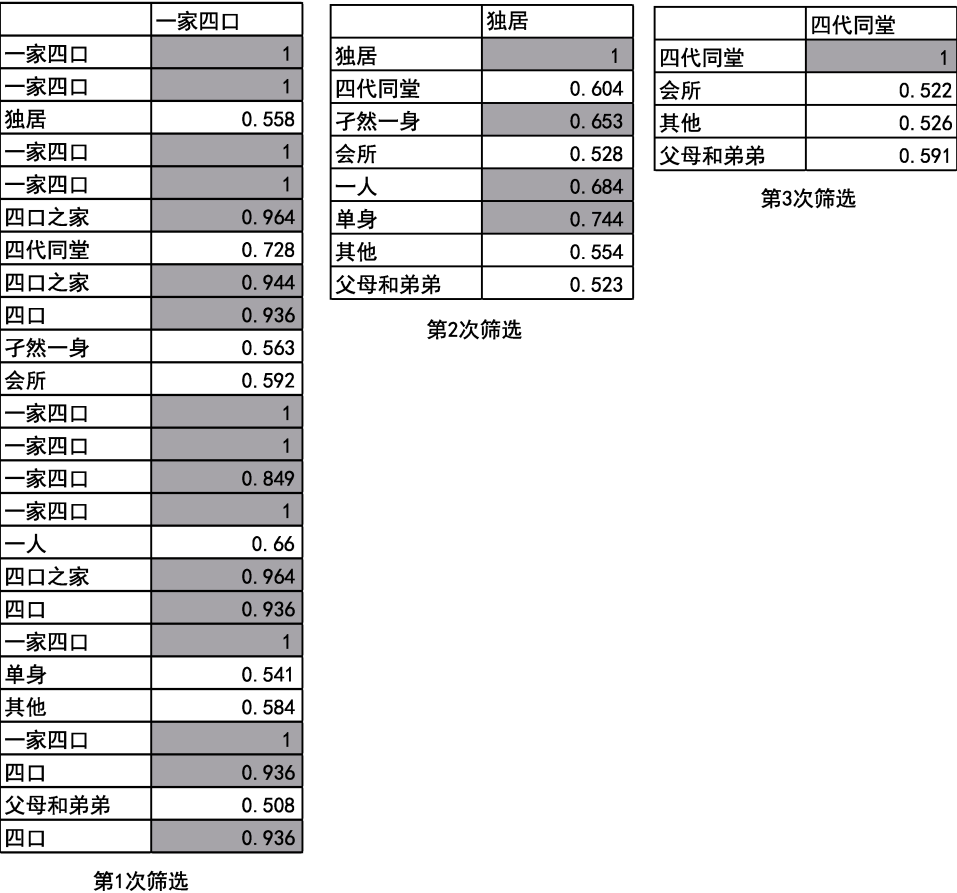


图 2 本算法对“问卷调查”语料进行迭代分类过程  
Fig.2 Iterative classification of “questionnaire survey”corpus by this algorithm

价值的语料，具体逻辑如下。

1) 用户评分为 3~4 颗星，可以视为理性评价，如情感倾向为中性和负向，可以认为有较高的研究价值。

2) 除以上语料外，如情感倾向为中性和负向都可认为有一定研究价值。

3) 用户评分为 1~2 颗星，情感倾向为正向，可以视为数据异常，作标注。

4) 除以上语料外，情感倾向为正向，可以视为无研究价值。

按上述逻辑筛选出有研究价值的语料并按优先级进行区分，可以大幅提高人工检阅效率。使用“情感倾向分析”算法对所有“网络评价”语料进行分析，结果为：54.1%，中性 18.9%，负向 27.0%，并且平均置信度在 0.9 以上。这部分的语料经过智能语义分析技术处理后，人工检阅的工作量仅为原来的 27%，通过评分对比之后还可以优先对其中研究价值较高的语料进行检阅。

**3.5 对“维修反馈”和“用户访谈”数据表的语料进行分析**

这部分数据的语料口语化严重，数据内容也差异

较大。经过初步试验发现，由于对部分的数据无法提取有效观点，使用“评论观点抽取”和“观点/文章摘要”等算法分析效果都不理想。根据“维修反馈”数据表的数据语中每字段的内容会局限在同一个较小范围的主题内的特点，统计涉及的关键动词和名词聚类可以获得有意义的效果。与本文所描述的方法不同，该算法不进行 TF-IDF 加权合并向量，而是需要加入人工监督进行分类模型训练，具体步骤如下：

(1) 对数据表内的语料进行中文分词，并过滤用词；

(2) 对数据进行二次处理，仅保留词性为动词和名词的分词实体；

(3) 获取所有分词的词向量，为每个向量记录其所属反馈记录；

(4) 以记录为单位交叉计算分词的相似度，对分词集合做自动归类；

(5) 从上一步的计算结果数据提取部分结果作人工标注；

(6) 用上一步中得到的标注数据，训练分类模型；

(7) 评估模型效果，适当引入新的标注数据，重复步骤 4~6 直到可输出有意义的效果为止。

通过以上机器学习算法，可以统计出用户反馈的主要关注点统计表，“维修反馈”语料分析结果见表 4。可以看到，原本处理难度较大的语料数据，经过处理后可以得出维修反馈的问题主要集中在“使用困难”、“清理积水”、“水盒使用”和“WiFi 连接”四个方

表 4 “维修反馈” 语料分析结果 ( 部分 )  
Tab.4 Results of “maintenance feedback” corpus analysis (partial)

实体分词提取并进行统计划分	问题归纳	计数 ( 占比 )
不通电/使用/调试/不工作/失灵/ 不启动/按键/无法/不能/ 声音异常/开灯/面板	使用困难	393 ( 39.2% )
积水/漏水	清理积水	98 ( 9.7% )
水箱/水盒/缺水/水位	水盒使用	48 ( 4.7% )
网络/联网	WiFi 连接	18 ( 1.7% )

面, 这为后续的设计优化策略的制定提供了具体的数据支撑。

鉴于自然语言处理各种分析方法的性能差异, 在本项目的实际操作中笔者针对不同类型的语料测试了各类型算法的实际效果, 并对相关的应用算法进行优化调整。对于可用于用户研究的相关语料来说, “中文分词”、“短文本相似度”和“情感倾向”等不需要根据上下文的语境来进行判断的智能语义分析算法准确率较高, 在实际应用中可以主要围绕这几种分析方法来设计相匹配的分析方法。

4 结语

从上述研究和实践可以看到, 在用户研究过程中应用人工智能技术具有巨大的潜力。引入人工智能技术替代部分需要人智分析的工作, 通过总结相关技术的特点和适用范围, 可以设计出有效的流程与方法, 提高工作效率。在目前的探索中, 部分语义分析方法仍有准确性不高的缺陷, 对一些具有上下文语境的描述欠缺前后逻辑关系的语义分析。本文验证了“中文分词”、“短文本相似度”和“情感倾向”这几种智能语义分析算法的有效性。应用这些算法, 可以减轻用户研究语料分析的工作量, 还可以进一步挖掘新的分析方法和创新手段。例如本文提出的短文本语义聚类分析方法, 可以对更大规模的头脑风暴亲和图进行自动聚类划分, 提高创意产出的能力。

可以预见在用户研究和产品设计流程中引入人工智能语义分析技术是未来的重要发展趋势, 是跟上大数据时代不可或缺的重要手段。

参考文献:

[1] 丁一, 郭伏, 胡名彩, 等. 用户体验国内外研究综述[J]. 工业工程与管理, 2014, 19(4): 92-97.  
DING Yi, GUO Fu, HU Ming-cai, et al. Domestic and Foreign Research Overview for User Experience[J]. Industrial Engineering and Management, 2014, 19(4): 92-97.

[2] 李小青. 基于用户心理研究的用户体验设计[J]. 情报

科学, 2010, (5): 125-129.

LI Xiao-qing. Design of User Experience Based on Research on User Psychology[J]. Information Science, 2010, (5): 125-129.

[3] 胡飞, 彭瑶玲. 用户研究中的语言材料分[J]. 包装工程, 2018, 39(10): 220-225.

HU Fei, PENG Yao-ling. Linguistic Material Analysis in User Research[J].Packaging Engineering, 2018, 39(10): 220-225.

[4] 饶元, 吴连伟, 王一鸣, 等. 基于语义分析的情感计算技术研究进展[J]. 软件学报, 2018, (8): 2397-2426.

RAO Yuan, WU Lian-wei, WANG Yi-ming, et al. Progress of Affective Computing Research Based on Semantic Analysis[J]. Journal of Software, 2018, (8): 2397-2426.

[5] 冯志伟. 自然语言的计算机处理[J]. 中文信息, 1997, (4): 26-27.

FENG Zhi-wei. Computer Processing of Natural Language[J]. Chinese Information, 1997, (4): 26-27.

[6] 秦春秀, 祝婷, 赵捧未, 等. 自然语言语义分析研究进展[J]. 图书情报工作, 2014, 58(22): 130-137.

QIN Chun-xiu, ZHU Ting, ZHAO Peng-wei, et al. Progress of Semantic Analysis Research of Natural Language[J]. Library and Information Service, 2014, 58(22): 130-137.

[7] 刘小冬. 自然语言理解综述[J]. 统计与信息论坛, 2007, 2(2): 5-12.

LIU Xiao-dong. Overview of Natural Language Understanding[J]. Statistics & Information Forum, 2007, 2(2): 5-12.

[8] 刘浏, 王东波. 命名实体识别研究综述[J]. 情报学报, 2018, (3): 79.

LIU Liu, WANG Dong-bo. Research Overview of Identity of Named Entity[J]. Journal of the China Society for Scientific and Technical Information, 2018, (3): 79.

[9] 秦春秀, 祝婷, 赵捧未, 等. 自然语言语义分析研究进展[J]. 图书情报工作, 2015, 58(22): 130-137.

QIN Chun-xiu, ZHU Ting, ZHAO Peng-wei, et al. Progress of Semantic Analysis Research of Natural Language[J]. Library And Information Service, 2015, 58(22): 130-137.

[10] 赵京胜, 宋梦雪, 高祥. 自然语言处理发展及应用综述[J]. 信息技术与信息化, 2019, (7): 142-145.

ZHAO Jing-sheng, SONG Meng-xue, GAO Xiang. Development and Application of Natural Language Processing[J]. Information Technology and Informatization, 2019, (7): 142-145.

[11] 刘挺, 车万翔, 李正华. 语言技术平台[J]. 中文信息学报, 2011, 25(6): 53-62.

LIU Ting, CHE Wan-xiang, LI Zheng-hua. Language Technology Platform[J]. Journal of Chinese Information Processing, 2011, 25(6): 53-62.

( 下转第 117 页 )

- Course[J]. Education Modernization, 2018, 5(4): 96-98.
- [13] 陈琼, 王俊勃, 姚旭, 等. 产品改良设计的策略研究[J]. 工业设计, 2016(3): 93-95.
- CHEN Qiong, WANG Jun-bo, YAO Xu, et al. Strategy Research on Product Improvement Design[J]. Industrial Design, 2016(3): 93-95.
- [14] 韩卫国. 基于经济性原则的产品改良设计[J]. 机电产品开发与创新, 2009, 22(4): 50-52.
- HAN Wei-guo. The Reform Design of Product Based on the Principle of Economy[J]. Development and Innovation of Mechanical and Electrical Products, 2009, 22(4): 50-52.
- [15] DARRELL M. Physical Contradictions and Evaporating Clouds[J/OL]. Triz-journal, 2000(4): <http://www.systematic-innovation.com/assets/200004-physicalcontradictionsandevaporatingclouds.pdf>.
- [16] XU Xiao-Yan, SOLOMANI C. Product Design Problem-Solving by Integrating TRIZ and TOC[J]. Journal of University of Ence & Technology of China, 2005, 35.
- [17] 吕楠, 尚清, 马彩云, 等. 康复机器人对痉挛型脑性瘫痪患儿的康复效果[J]. 中国实用神经疾病杂志, 2017, 20(7): 45-47.
- LYU Nan, SHANG Qing, MA Cai-yun, et al. Rehabilitation Effect of Rehabilitation Robot on Children with Spastic Cerebral Palsy[J]. Chinese Journal of Practical Nervous Diseases, 2017, 20(7): 45-47.
- [18] 杜佳音, 范艳萍, 李鑫, 等. 自制脑性瘫痪儿童下肢康复器的研制及临床效果[J]. 中国康复理论与实践, 2017, 23(4): 430-432.
- DU Jia-yin, FAN Yan-ping, LI Xin, et al. Development and Clinical Effects of Self-Made Cerebral Palsy Children's Lower Limb Rehabilitation Device[J]. Chinese Journal of Rehabilitation Theory and Practice, 2017, 23(4): 430-432.
- 
- (上接第 59 页)
- [12] 赵文清, 侯小可, 沙海虹. 语义规则在微博热点话题情感分析中的应用[J]. 智能系统学报, 2014, 9(1): 121-125.
- ZHAO Wen-qing, HOU Xiao-ke, SHA Hai-hong. Application of Semantic Rules in the Emotion Analysis of Microblog Hot Topics[J]. Transactions on Intelligent Systems, 2014, 9(1): 121-125.
- [13] 陈二静, 姜恩波. 文本相似度计算方法研究综述[J]. 数据分析与知识发现, 2017, 1(6): 1-11.
- CHEN Er-jing, JIANG En-bo. Research Overview of Text Similarity Computing Methods[J]. Data Analysis and Knowledge Discovery, 2017, 1(6): 1-11.
- [14] 费洪晓, 康松林, 朱小娟, 等. 基于词频统计的中文分词的研究[J]. 计算机工程与应用, 2005, (7): 67-68.
- FEI Hong-xiao, KANG Song-lin, ZHU Xiao-juan, et al. Research on Chinese Word Segmentation Based on Word Frequency Statistics[J]. Computer Engineering and Applications, 2005, (7): 67-68.
- [15] 杨立公, 朱俭, 汤世平. 文本情感分析综述[J]. 计算机应用, 2013, 33(6): 1574-1578.
- YANG Li-gong, ZHU Jian, TANG Shi-ping. Overview of Text Emotion Analysis[J]. Computer Application, 2013, 33(6): 1574-1578.
- [16] 郭喜跃, 何婷婷, 胡小华, 等. 基于句法语义特征的中文实体关系抽取[J]. 中文信息学报, 2014, 28(6): 183-189.
- GUO Xi-yue, HE Ting-ting, HU Xiao-hua, et al. Extraction of Chinese Entity Relationship Based on Syntactic and Semantic Characteristics[J]. Journal of Chinese Information Processing, 2014, 28(6): 183-189.
- [17] Yogesh Sankarasu Bramaniam, Krishnan Ramanathan, Subhankar Ghosh. Text Summarization Using Wikipedia[J]. Information Processing and Management, 2014, (3): 75.
- [18] 祝帅. 设计学的社会科学化倾向与实证研究的兴起——兼论当代艺术设计研究的理论前沿与发展趋势[J]. 艺术设计研究, 2009, (4): 85-88.
- ZHU Shuai. Social Scientization Tendency of Design and Rise of Empirical Research: Theoretical Frontiers and Development Tendency of Contemporary Art Design[J]. Art Design Research, 2009, (4): 85-88.
- [19] 胡岷. 用户行为研究常用调查方法述略[J]. 成都理工大学学报, 2005, (3): 54-57.
- HU Min. Common Investigation Methods for User Behaviors Research[J]. Journal of Chengdu University of Technology, 2005, (3): 54-57.