



(12) 发明专利申请

(10) 申请公布号 CN 120124009 A

(43) 申请公布日 2025.06.10

(21) 申请号 202510197399.5

G06N 20/00 (2019.01)

(22) 申请日 2025.02.21

(71) 申请人 广东水木清宇科技有限公司

地址 528000 广东省佛山市南海区大沥镇
盐步穗盐西路92号厂房A栋2楼14室

(72) 发明人 李昊 张頔 赵顺 傅建勋
周汪研

(74) 专利代理机构 广州市一新专利商标事务所
有限公司 44220

专利代理师 何健施

(51) Int. Cl.

G06F 18/27 (2023.01)

G06F 18/10 (2023.01)

G06F 40/30 (2020.01)

G06F 40/279 (2020.01)

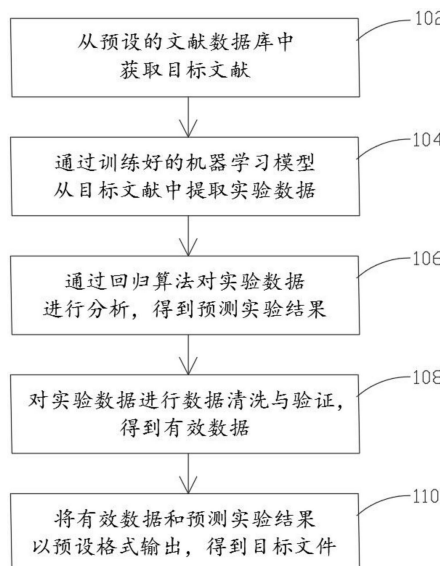
权利要求书2页 说明书8页 附图4页

(54) 发明名称

科学文献数据提取方法、装置、计算机设备和存储介质

(57) 摘要

本申请涉及一种科学文献数据提取方法、装置、计算机设备和存储介质。该方法包括：从预设的文献数据库中获取目标文献；通过训练好的机器学习模型从目标文献中提取实验数据；通过回归算法对实验数据进行分析，得到预测实验结果；对实验数据进行数据清洗与验证，得到有效数据；将有效数据和预测实验结果以预设格式输出，得到目标文件。采用本方法能够高效准确地提取实验数据，并预测未来实验结果。



1. 一种科学文献数据提取方法,所述方法包括:
 - 从预设的文献数据库中获取目标文献;
 - 通过训练好的机器学习模型从所述目标文献中提取实验数据;
 - 通过回归算法对所述实验数据进行分析,得到预测实验结果;
 - 对所述实验数据进行数据清洗与验证,得到有效数据;
 - 将所述有效数据和所述预测实验结果以预设格式输出,得到目标文件。
2. 根据权利要求1所述的方法,其特征在于,所述从预设的文献数据库中获取目标文献,包括:
 - 通过脚本代码在预设的文献数据库中检索目标领域内的相关文献;
 - 通过自然语言处理技术对所述相关文献进行预处理,得到目标文献,所述目标文献为所述机器学习模型可处理的格式。
3. 根据权利要求2所述的方法,其特征在于,所述通过训练好的机器学习模型从所述目标文献中提取实验数据,包括:
 - 通过训练好的机器学习模型对所述目标文献进行语义分析,并提取所述目标文献中的实验数据;所述实验数据包括实验条件、实验参数及实验结果中的至少一项。
4. 根据权利要求1所述的方法,其特征在于,在所述通过训练好的机器学习模型从所述目标文献中提取实验数据之前,所述方法还包括:
 - 获取训练数据集,所述训练数据集包含多个已标注的训练数据;
 - 基于所述训练数据集,对预设的机器学习模型进行训练,得到所述训练好的机器学习模型。
5. 根据权利要求1至4任意一项所述的方法,其特征在于,所述通过回归算法对所述实验数据进行分析,得到预测实验结果,包括:
 - 将所述实验数据输入回归算法,构建回归预测模型;
 - 通过所述回归预测模型基于所述实验数据和历史实验数据进行预测,得到所述实验数据对应的预测实验结果。
6. 根据权利要求5所述的方法,其特征在于,所述对所述实验数据进行数据清洗与验证,得到有效数据,包括:
 - 对所述实验数据进行数据清洗;
 - 以及,通过所述回归预测模型对所述实验数据进行数据校验,通过异常检验算法识别和修正所述实验数据中的异常数据点。
7. 根据权利要求6所述的方法,其特征在于,所述方法还包括:
 - 对所述目标文件进行可视化处理,生成可视化结果,并展示所述可视化结果。
8. 一种科学文献数据提取装置,其特征在于,所述装置包括:
 - 文献获取模块,用于从预设的文献数据库中获取目标文献;
 - 数据提取模块,用于通过训练好的机器学习模型从所述目标文献中提取实验数据;
 - 实验结果预测模块,用于通过回归算法对所述实验数据进行分析,得到预测实验结果;
 - 数据清洗验证模块,用于对所述实验数据进行数据清洗与验证,得到有效数据;
 - 目标文件输出模块,用于将所述有效数据和所述预测实验结果以预设格式输出,得到目标文件。

9.一种计算机设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至7中任一项所述方法的步骤。

10.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至7中任一项所述方法的步骤。

科学文献数据提取方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及科学文献数据处理技术领域,特别是涉及一种科学文献数据提取方法、装置、计算机设备和存储介质。

背景技术

[0002] 随着科学技术的不断发展,文献数量急剧增加,科研人员面临着海量文献数据的筛选与分析压力。现有的文献数据提取方法主要依赖于人工筛选或基于规则的自动化处理。

[0003] 然而,人工处理文献数据不仅消耗大量时间,还容易遗漏重要信息,存在效率低下、出错率高的问题;而基于规则的自动化处理方法则难以处理复杂的自然语言文本,提取的结果容易出现误差,导致数据提取的准确性差;同时,上述两种方法对于不同领域或新兴研究领域的适应性均较差,难以满足不同领域的文献数据提取需求。

[0004] 综上所述,现有的文献数据提取方法存在效率低、准确性差以及灵活性差等问题。

发明内容

[0005] 基于此,有必要针对上述技术问题,提供一种科学文献数据提取方法、装置、计算机设备和存储介质,能够高效准确地提取实验数据,并预测未来实验结果。

[0006] 一种科学文献数据提取方法,该方法包括:

[0007] 从预设的文献数据库中获取目标文献;

[0008] 通过训练好的机器学习模型从目标文献中提取实验数据;

[0009] 通过回归算法对实验数据进行分析,得到预测实验结果;

[0010] 对实验数据进行数据清洗与验证,得到有效数据;

[0011] 将有效数据和预测实验结果以预设格式输出,得到目标文件。

[0012] 在其中一个实施例中,从预设的文献数据库中获取目标文献,包括:通过脚本代码在预设的文献数据库中检索目标领域内的相关文献;通过自然语言处理技术对相关文献进行预处理,得到目标文献,目标文献为机器学习模型可处理的格式。

[0013] 在其中一个实施例中,通过训练好的机器学习模型从目标文献中提取实验数据,包括:通过训练好的机器学习模型对目标文献进行语义分析,并提取目标文献中的实验数据;实验数据包括实验条件、实验参数及实验结果中的至少一项。

[0014] 在其中一个实施例中,在通过训练好的机器学习模型从目标文献中提取实验数据之前,该方法还包括:获取训练数据集,训练数据集包含多个已标注的训练数据;基于训练数据集,对预设的机器学习模型进行训练,得到训练好的机器学习模型。

[0015] 在其中一个实施例中,通过回归算法对实验数据进行分析,得到预测实验结果,包括:将实验数据输入回归算法,构建回归预测模型;通过回归预测模型基于实验数据和历史实验数据进行预测,得到实验数据对应的预测实验结果。

[0016] 在其中一个实施例中,对实验数据进行数据清洗与验证,得到有效数据,包括:对

实验数据进行数据清洗;以及,通过回归预测模型对实验数据进行数据校验,通过异常检验算法识别和修正实验数据中的异常数据点。

[0017] 在其中一个实施例中,该方法还包括:

[0018] 对目标文件进行可视化处理,生成可视化结果,并展示可视化结果。

[0019] 一种科学文献数据提取装置,该装置包括:

[0020] 文献获取模块,用于从预设的文献数据库中获取目标文献;

[0021] 数据提取模块,用于通过训练好的机器学习模型从目标文献中提取实验数据;

[0022] 实验结果预测模块,用于通过回归算法对实验数据进行分析,得到预测实验结果;

[0023] 数据清洗验证模块,用于对实验数据进行数据清洗与验证,得到有效数据;

[0024] 目标文件输出模块,用于将有效数据和预测实验结果以预设格式输出,得到目标文件。

[0025] 一种计算机设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现以下步骤:从预设的文献数据库中获取目标文献;通过训练好的机器学习模型从目标文献中提取实验数据;通过回归算法对实验数据进行分析,得到预测实验结果;对实验数据进行数据清洗与验证,得到有效数据;将有效数据和预测实验结果以预设格式输出,得到目标文件。

[0026] 一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现以下步骤:从预设的文献数据库中获取目标文献;通过训练好的机器学习模型从目标文献中提取实验数据;通过回归算法对实验数据进行分析,得到预测实验结果;对实验数据进行数据清洗与验证,得到有效数据;将有效数据和预测实验结果以预设格式输出,得到目标文件。

[0027] 上述科学文献数据提取方法、装置、计算机设备和存储介质,通过从预设的文献数据库中获取目标文献,可以实现目标文献的自动化检索与获取,提高文献检索和获取的效率;然后,通过训练好的机器学习模型从目标文献中提取实验数据,能够避免人工提取的误差,提高数据提取的准确性和效率;通过回归算法对实验数据进行分析,得到预测实验结果,能够为科研人员提供实验趋势预测和数据优化支持;对实验数据进行数据清洗与验证,得到有效数据,确保提取数据的准确性与一致性;将有效数据和预测实验结果以预设格式输出,得到目标文件,便于存储与后续分析。

[0028] 综上,上述方法通过结合机器学习与回归预测模型,能够从海量科学文献中自动提取实验数据,并对提取到的实验数据进行回归分析,预测未来实验结果,既能提高数据提取的效率、准确性及适应性,又能提供预测实验结果,从而为科研人员提供更加高效、精准的数据提取与分析工具。

附图说明

[0029] 图1为一个实施例中科学文献数据提取方法的流程示意图;

[0030] 图2为另一个实施例中科学文献数据提取方法的流程示意图;

[0031] 图3为一个实施例中科学文献数据提取装置的结构框图;

[0032] 图4为另一个实施例中科学文献数据提取装置的结构框图;

[0033] 图5为一个实施例中计算机设备的内部结构图。

具体实施方式

[0034] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0035] 在一个实施例中,如图1-2所示,提供了一种科学文献数据提取方法,本实施例以该方法应用于终端为例进行说明,可以理解的是,该方法也可以应用于服务器,还可以应用于包括终端和服务器的系统,并通过终端和服务器的交互实现。其中,终端可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备,服务器可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

[0036] 在本实施例中,该方法可以包括以下步骤:

[0037] 步骤102,从预设的文献数据库中获取目标文献。

[0038] 其中,根据用户的实际需求,文献数据库的数量可以为一个或多个,为了提高可获取的目标文献数量,一般预设有多个文献数据库,如PubMed、IEEE Xplore、Web of Science等数据库,即终端可支持多种文献数据库接口;目标文献可以是用户所需目标领域内相关的科学文献,其格式为适合输入机器学习模型的格式,如JSON、CSV等格式。

[0039] 具体地,终端自动从一个或多个文献数据库中获取目标文献;需要注意的是,文献数据库内的文献可能为机器学习模型可处理的格式,也可能是不适合输入机器学习模型的格式。对于前者,终端可直接从文献数据库内获取到目标文献;但对于后者,终端需要先对检索到的文献进行格式转化,才能得到目标文献。

[0040] 上述步骤102可以实现目标文献的自动化检索与获取,提高文献检索和获取的效率。

[0041] 步骤104,通过训练好的机器学习模型从目标文献中提取实验数据。

[0042] 其中,机器学习模型为一种能够将输入数据转化为特定输出的算法或数学函数。它们是机器学习的核心组件,通过训练数据,机器学习模型可以自动学习和改进自己,使得其在处理新数据时更加准确和有效。在步骤104中,训练好的机器学习模型可以根据输入的目标文献,输出提取的实验数据。

[0043] 具体地,终端调用训练好的深度学习模型,自动对目标文献的内容进行分析与数据提取,得到目标文献中的实验数据;实验数据可以包括实验条件、实验参数、实验结果等关键信息。在实际应用场景下,目标文献的数量通常不止一篇,机器学习模型既可以是逐篇提取实验数据,也可以是同时从多篇目标文献中提取实验数据。

[0044] 上述步骤104利用训练好的机器学习模型来提取实验数据,能够避免人工提取的误差,提高数据提取的准确性和效率。

[0045] 步骤106,通过回归算法对实验数据进行分析,得到预测实验结果。

[0046] 其中,回归算法是一种机器学习算法,用于根据输入数据预测数值,可用于对新的、看不见的数据进行准确预测;具体实施时,上述回归算法可采用线性回归、支持向量回归、随机森林回归、决策树回归等;

[0047] 具体地,终端采用回归算法,对步骤104中提取到的实验数据进行分析 and 预测,得到预测实验结果,能够为科研人员提供实验趋势预测和数据优化支持。

[0048] 步骤108,对实验数据进行数据清洗与验证,得到有效数据。

[0049] 其中,数据清洗可去除实验数据中冗余、重复或无效的数据,并确保实验数据的统一性和一致性;数据验证可利用回归算法和异常检验算法实现,采用回归算法对实验数据进行校验,采用异常检验算法识别和修正实验数据中的异常数据点;有效数据即经过数据清洗与验证后的实验数据。

[0050] 具体地,终端对实验数据进行数据清洗、去重及异常值检测,能够确保提取到的数据质量。

[0051] 步骤110,将有效数据和预测实验结果以预设格式输出,得到目标文件。

[0052] 其中,预设格式为结构化格式,如CSV、Excel等;目标文件为结构化格式的数据文件。

[0053] 具体地,终端将有效数据和预测实验结果输出为结构化数据文件,便于存储与后续分析。

[0054] 上述科学文献数据提取方法中,通过从预设的文献数据库中获取目标文献,可以实现目标文献的自动化检索与获取,提高文献检索和获取的效率;然后,通过训练好的机器学习模型从目标文献中提取实验数据,能够避免人工提取的误差,提高数据提取的准确性和效率;通过回归算法对实验数据进行分析,得到预测实验结果,能够为科研人员提供实验趋势预测和数据优化支持;对实验数据进行数据清洗与验证,得到有效数据,确保提取数据的准确性与一致性;将有效数据和预测实验结果以预设格式输出,得到目标文件,便于存储与后续分析。综上,本发明通过结合机器学习与回归预测模型,能够从海量科学文献中自动提取实验数据,并对提取到的实验数据进行回归分析,预测未来实验结果,既能提高数据提取的效率、准确性及适应性,又能提供预测实验结果,从而为科研人员提供更加高效、精准的数据提取与分析工具。

[0055] 在一个实施例中,步骤102包括:

[0056] 通过脚本代码在预设的文献数据库中检索目标领域内的相关文献;

[0057] 通过自然语言处理技术对相关文献进行预处理,得到目标文献,目标文献为机器学习模型可处理的格式。

[0058] 其中,目标领域可由用户输入的预设关键词确定,脚本代码可以是使用Python语言编写而成的自动化脚本,用于实现文献的检索和下载功能;自然语言处理技术(Natural Language Processing, NLP)是用计算机来处理、理解及生成人类语言的科学和应用技术,可以用计算机代替人工来处理大规模的自然语言信息,这里用于相关文献的预处理,即将相关文献自动转化为适合输入机器学习模型的格式,如JSON、CSV等格式,即目标文献。

[0059] 具体地,终端根据用户输入的预设关键词确定目标领域,并使用自动化的脚本代码,在预设的各个文献数据库中检索并下载目标领域内的相关文献;然后,通过自然语言处理技术对相关文献进行格式转换,得到机器学习模型可处理的格式的目标文献。

[0060] 在上述实施例中,通过脚本代码检索和获取相关文献,在实现文献检索自动化的同时,还提高了文献检索效率和准确性;然后利用自然语言处理技术进行预处理,得到可输入机器学习模型的目标文献,更加便于机器学习模型进行实验数据提取,有助于提高数据提取效率。

[0061] 在一个实施例中,步骤104包括:通过训练好的机器学习模型对目标文献进行语义分析,并提取目标文献中的实验数据;实验数据包括实验条件、实验参数及实验结果中的至

少一项。

[0062] 其中,机器学习模型可以采用具有自动提取特征能力的深度学习模型,从而更加高效、精准地提取实验数据;深度学习模型可采用现有技术中常见的深度学习模型,例如BERT、GPT、T5等。

[0063] 在一个实施例中,在步骤104之前,该方法还包括:

[0064] 获取训练数据集,训练数据集包含多个已标注的训练数据;

[0065] 基于训练数据集,对预设的机器学习模型进行训练,得到训练好的机器学习模型。

[0066] 其中,训练数据集包含大量已标注的训练数据,这些训练数据可以来自相同领域或不同领域。

[0067] 具体实施时,通过向训练数据集加入不同领域的训练数据,再进一步对机器学习模型进行训练,可以使训练好的机器学习模型在实验数据的提取过程中,能够处理不同领域的文献和复杂的自然语言描述,从而更准确地提取实验数据,并满足不同领域的文献数据提取需求,提高灵活性。

[0068] 在一个实施例中,步骤106,包括:

[0069] 将实验数据输入回归算法,构建回归预测模型;

[0070] 通过回归预测模型基于实验数据和历史实验数据进行预测,得到实验数据对应的预测实验结果。

[0071] 其中,历史实验数据是在之前已提取的实验数据。具体地,终端将实验数据输入预设的回归算法,构建回归预测模型;并提取实验数据中的实验条件,回归预测模型根据该实验条件和历史实验数据进行预测,得到实验条件对应的预测实验结果;根据预测实验结果,科研人员可以优化实验设计或对未来实验结果进行预判。

[0072] 在一个实施例中,步骤108,包括:

[0073] 对实验数据进行数据清洗;

[0074] 以及,通过回归预测模型对实验数据进行数据校验,通过异常检验算法识别和修正实验数据中的异常数据点。

[0075] 其中,异常数据点可以是实验数据中的错误值或不一致之处。

[0076] 上述实施例通过对实验数据进行数据清洗,能够去除实验数据中冗余、重复或无效的数据,并确保实验数据的统一性和一致性;在数据清洗的过程中,通过回归预测模型对实验数据进行数据校验,可实现更准确的校验,同时,通过异常检验算法识别和修正实验数据中的异常数据点,能够得到高质量的数据。

[0077] 在一个实施例中,如图2所示,该方法还包括:

[0078] 步骤112,对目标文件进行可视化处理,生成可视化结果,并展示可视化结果。

[0079] 其中,可视化工具可为Matplotlib、Seaborn等;可视化结果可以为图表形式,如预测趋势图、回归曲线图等。

[0080] 具体地,终端通过可视化工具对目标文件进行可视化处理,生成预测趋势图、回归曲线图等可视化结果,能够帮助科研人员快速理解数据及其趋势。

[0081] 上述实施例中,通过目标文件的可视化,展示有效数据和预测实验结果的趋势图,方便科研人员更直观地理解数据和预测实验结果。

[0082] 应该理解的是,虽然图1-2的流程图中的各个步骤按照箭头的指示依次显示,但是

这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,这些步骤可以以其它的顺序执行。而且,图1-2中的至少一部分步骤可以包括多个子步骤或者多个阶段,这些子步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,这些子步骤或者阶段的执行顺序也不必然是依次进行,而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

[0083] 在一个实施例中,如图3所示,提供了一种科学文献数据提取装置,包括:文献获取模块202、数据提取模块204、实验结果预测模块206、数据清洗验证模块208和目标文件输出模块210,其中:

[0084] 文献获取模块202,用于从预设的文献数据库中获取目标文献;

[0085] 数据提取模块204,用于通过训练好的机器学习模型从目标文献中提取实验数据;

[0086] 实验结果预测模块206,用于通过回归算法对实验数据进行分析,得到预测实验结果;

[0087] 数据清洗验证模块208,用于对实验数据进行数据清洗与验证,得到有效数据;

[0088] 目标文件输出模块210,用于将有效数据和预测实验结果以预设格式输出,得到目标文件。

[0089] 在一个实施例中,文献获取模块202包括:检索单元,用于通过脚本代码在预设的文献数据库中检索目标领域内的相关文献;预处理单元,用于通过自然语言处理技术对相关文献进行预处理,得到目标文献,目标文献为机器学习模型可处理的格式。

[0090] 在一个实施例中,数据提取模块204,具体用于通过训练好的机器学习模型对目标文献进行语义分析,并提取目标文献中的实验数据;实验数据包括实验条件、实验参数及实验结果中的至少一项。

[0091] 在一个实施例中,该装置还包括:模型训练模块,用于获取训练数据集,训练数据集包含多个已标注的训练数据;基于训练数据集,对预设的机器学习模型进行训练,得到训练好的机器学习模型。

[0092] 在一个实施例中,实验结果预测模块206包括:模型构建单元,用于将实验数据输入回归算法,构建回归预测模型;结果预测单元,用于通过回归预测模型基于实验数据和历史实验数据进行预测,得到实验数据对应的预测实验结果。

[0093] 在一个实施例中,数据清洗验证模块208包括:数据清洗单元,用于对实验数据进行数据清洗;以及,数据校验单元,用于通过回归预测模型对实验数据进行数据校验,通过异常检验算法识别和修正实验数据中的异常数据点。

[0094] 在一个实施例中,如图4所示,该装置还包括:可视化模块212,用于对目标文件进行可视化处理,生成可视化结果,并展示可视化结果。

[0095] 关于科学文献数据提取装置的具体限定可以参见上文中对于科学文献数据提取方法的限定,在此不再赘述。上述科学文献数据提取装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中,也可以以软件形式存储于计算机设备中的存储器中,以便于处理器调用执行以上各个模块对应的操作。

[0096] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是终端,其内部结构

图可以如图5所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口、显示屏和输入装置。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种科学文献数据提取方法。该计算机设备的显示屏可以是液晶显示屏或者电子墨水显示屏,该计算机设备的输入装置可以是显示屏上覆盖的触摸层,也可以是计算机设备外壳上设置的按键、轨迹球或触控板,还可以是外接的键盘、触控板或鼠标等。

[0097] 本领域技术人员可以理解,图5中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0098] 在一个实施例中,提供了一种计算机设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,处理器执行计算机程序时实现以下步骤:从预设的文献数据库中获取目标文献;通过训练好的机器学习模型从目标文献中提取实验数据;通过回归算法对实验数据进行分析,得到预测实验结果;对实验数据进行数据清洗与验证,得到有效数据;将有效数据和预测实验结果以预设格式输出,得到目标文件。

[0099] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:通过脚本代码在预设的文献数据库中检索目标领域内的相关文献;通过自然语言处理技术对相关文献进行预处理,得到目标文献,目标文献为机器学习模型可处理的格式。

[0100] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:通过训练好的机器学习模型对目标文献进行语义分析,并提取目标文献中的实验数据;实验数据包括实验条件、实验参数及实验结果中的至少一项。

[0101] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:获取训练数据集,训练数据集包含多个已标注的训练数据;基于训练数据集,对预设的机器学习模型进行训练,得到训练好的机器学习模型。

[0102] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:将实验数据输入回归算法,构建回归预测模型;通过回归预测模型基于实验数据和历史实验数据进行预测,得到实验数据对应的预测实验结果。

[0103] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:对实验数据进行数据清洗;以及,通过回归预测模型对实验数据进行数据校验,通过异常检验算法识别和修正实验数据中的异常数据点。

[0104] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:对目标文件进行可视化处理,生成可视化结果,并展示可视化结果。

[0105] 在一个实施例中,提供了一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现以下步骤:从预设的文献数据库中获取目标文献;通过训练好的机器学习模型从目标文献中提取实验数据;通过回归算法对实验数据进行分析,得到预测实验结果;对实验数据进行数据清洗与验证,得到有效数据;将有效数据和预测实验结果以预设格式输出,得到目标文件。

[0106] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:通过脚本代码在

预设的文献数据库中检索目标领域内的相关文献;通过自然语言处理技术对相关文献进行预处理,得到目标文献,目标文献为机器学习模型可处理的格式。

[0107] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:通过训练好的机器学习模型对目标文献进行语义分析,并提取目标文献中的实验数据;实验数据包括实验条件、实验参数及实验结果中的至少一项。

[0108] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:获取训练数据集,训练数据集包含多个已标注的训练数据;基于训练数据集,对预设的机器学习模型进行训练,得到训练好的机器学习模型。

[0109] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:将实验数据输入回归算法,构建回归预测模型;通过回归预测模型基于实验数据和历史实验数据进行预测,得到实验数据对应的预测实验结果。

[0110] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:对实验数据进行数据清洗;以及,通过回归预测模型对实验数据进行数据校验,通过异常检验算法识别和修正实验数据中的异常数据点。

[0111] 在一个实施例中,计算机程序被处理器执行时还实现以下步骤:对目标文件进行可视化处理,生成可视化结果,并展示可视化结果。

[0112] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以 通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机 可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中, 本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可 包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM (PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括 随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限,RAM以多种形式可得, 诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强 型SDRAM(ESDRAM)、同步链路(Synch l i nk) DRAM(SLDRAM)、存储器总线(Rambus)直接RAM (RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0113] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例 中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾, 都应当认为是本说明书记载的范围。

[0114] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不 能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护 范围。因此,本申请专利的保护范围应以所附权利要求为准。

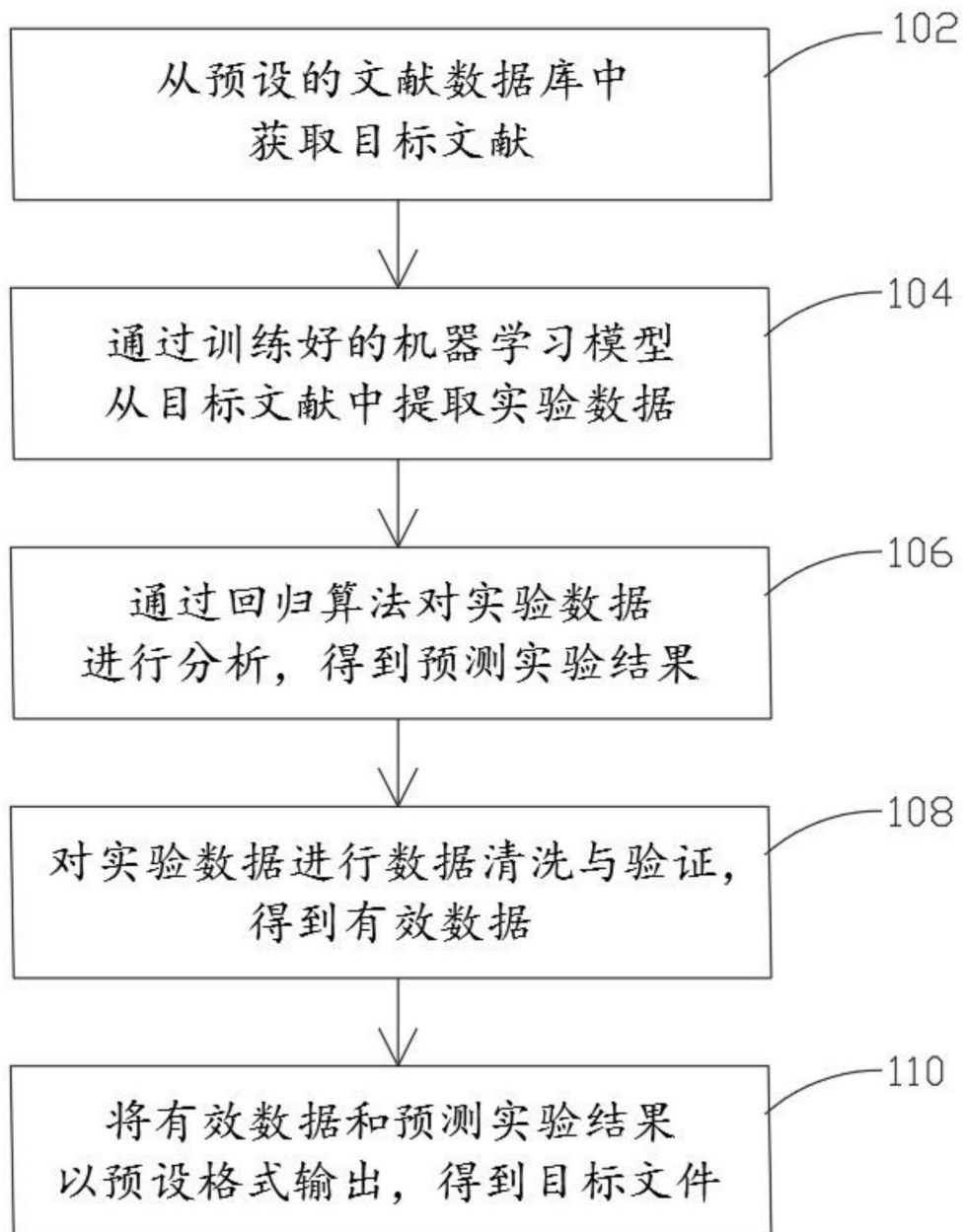


图1

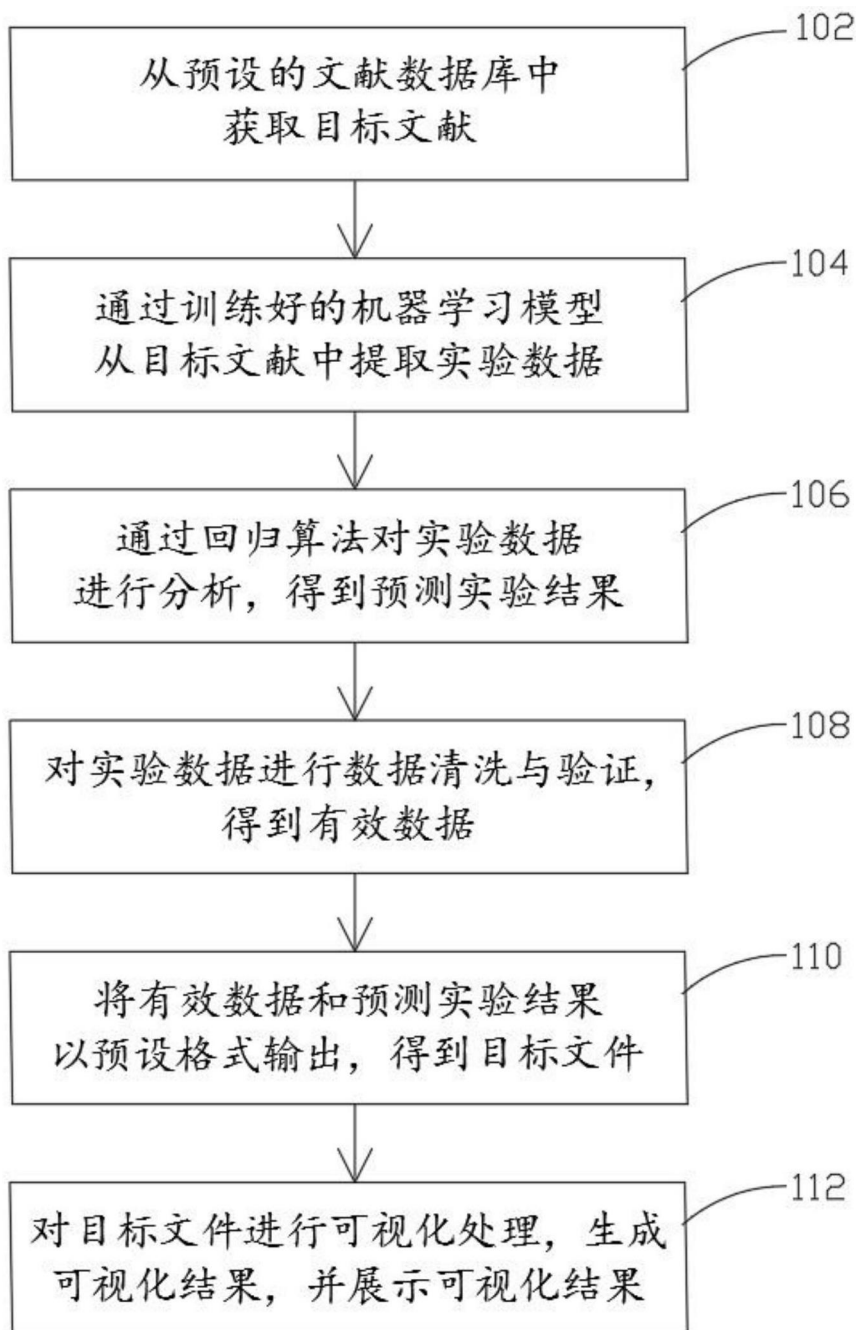


图2

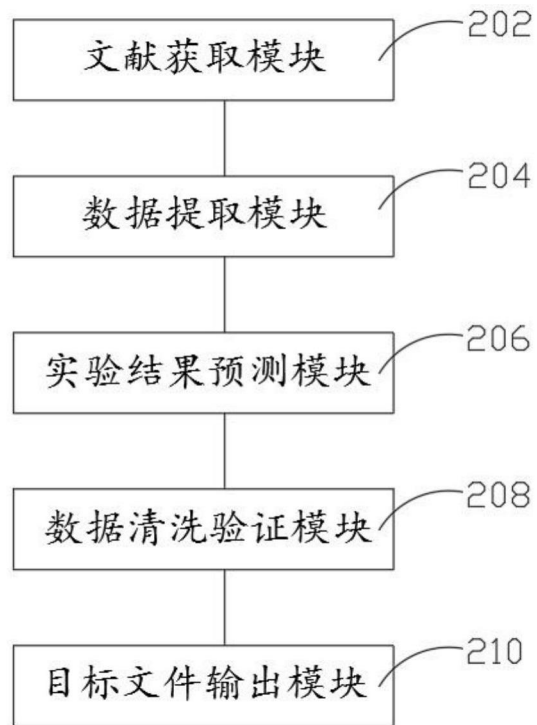


图3

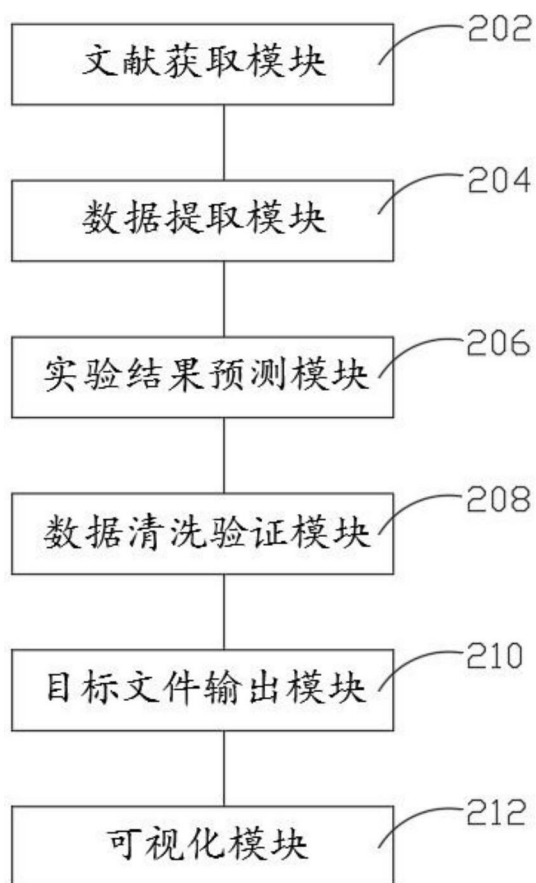


图4

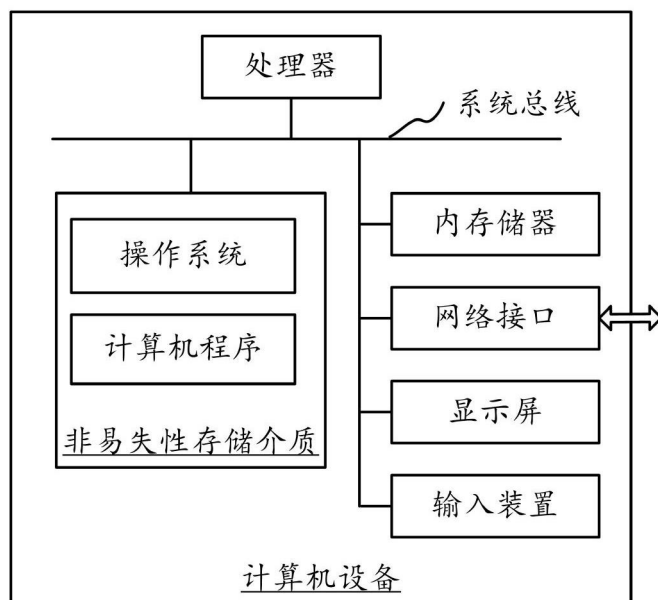


图5