



(12)发明专利

(10)授权公告号 CN 106919702 B

(45)授权公告日 2020.02.11

(21)申请号 201710137604.4
 (22)申请日 2017.03.09
 (65)同一申请的已公布的文献号
 申请公布号 CN 106919702 A
 (43)申请公布日 2017.07.04
 (66)本国优先权数据
 201710079411.8 2017.02.14 CN
 (73)专利权人 北京时间股份有限公司
 地址 100089 北京市海淀区西三环北路3号
 一区1号楼7层710
 (72)发明人 王立宁 陈劲
 (74)专利代理机构 北京市浩天知识产权代理事
 务所(普通合伙) 11276
 代理人 宋菲 刘兰兰

(51)Int.Cl.
 G06F 16/9535(2019.01)
 G06F 16/33(2019.01)
 G06N 3/04(2006.01)
 G06N 3/08(2006.01)
 (56)对比文件
 CN 106095749 A,2016.11.09,
 CN 105679316 A,2016.06.15,
 CN 101523338 A,2009.09.02,
 CN 106021463 A,2016.10.12,
 CN 104143005 A,2014.11.12,
 US 2005197983 A1,2005.09.08,
 蒙海涛.一种基于神经网络和主动反馈的反
 垃圾邮件技术的研究.《微电子与计算机》.2011,
 第28卷(第6期),
 审查员 李若晨

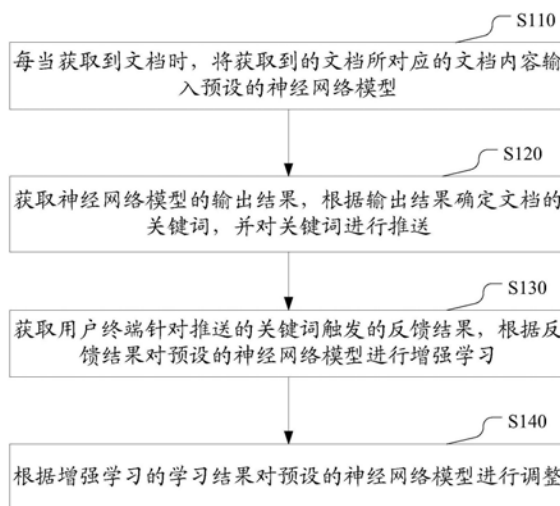
权利要求书3页 说明书14页 附图4页

(54)发明名称

基于文档的关键词推送方法及装置

(57)摘要

本发明公开了一种基于文档的关键词推送方法及装置,至少能够解决现有技术中由于无法根据文本内容中的语义来挖掘关键词而导致推送的关键词过于粗糙,不够准确的技术问题。该方法包括:每当获取到文档时,将获取到的文档所对应的文档内容输入预设的神经网络模型;获取神经网络模型的输出结果,根据输出结果确定文档的关键词,并对关键词进行推送;获取用户终端针对推送的关键词触发的反馈结果,根据反馈结果对预设的神经网络模型进行增强学习;根据增强学习的学习结果对预设的神经网络模型进行调整。



1. 一种基于文档的关键词推送方法,包括:

每当获取到文档时,将获取到的文档所对应的文档内容输入预设的神经网络模型;

获取所述神经网络模型的输出结果,根据所述输出结果确定所述文档的关键词,并对所述关键词进行推送;

获取用户终端针对推送的所述关键词触发的反馈结果,根据所述反馈结果对所述预设的神经网络模型进行增强学习;

根据所述增强学习的学习结果对所述预设的神经网络模型进行调整;

其中,对关键词进行推送时,为各个关键词设置多个不同的点击操作入口;在根据用户终端针对推送的关键词触发的反馈结果对预设的神经网络模型进行增强学习时,进一步将查询到的用户日志文件中包含的与关键词对应的点击事件细分为多种类型的点击操作事件,预先为每种类型的点击操作事件设置不同的分析优先级,并根据不同类型的点击操作事件的分析优先级设置对应的反馈结果的结果分值,该结果分值用于确定对应的样本的样本权重;其中,所述点击操作事件包括:评分点击操作事件、浏览点击操作事件和搜索点击操作事件。

2. 根据权利要求1所述的方法,其中,所述将获取到的文档所对应的文档内容输入预设的神经网络模型的步骤之前,进一步包括:

对所述获取到的文档进行预处理,将预处理结果作为所述文档对应的文档内容;其中,所述预处理包括以下中的至少一个:分词处理、以及词向量转换处理。

3. 根据权利要求1所述的方法,其中,所述根据所述输出结果确定所述文档的关键词的步骤具体包括:根据预设的筛选策略对所述输出结果进行筛选,将筛选结果确定为所述文档的关键词;

其中,所述筛选策略包括以下中的至少一个:根据词性进行筛选的策略;以及,根据词频进行筛选的策略。

4. 根据权利要求1所述的方法,其中,所述对所述关键词进行推送的步骤之后进一步包括:当监测到用户终端针对所述关键词触发的点击事件时,将所述点击事件记录到用户日志文件中;

则所述获取用户终端针对推送的所述关键词触发的反馈结果的步骤具体包括:

查询所述用户日志文件中是否包含与所述关键词对应的点击事件;

当查询结果为否时,确定所述用户终端针对推送的所述关键词触发的反馈结果为第一类反馈结果;其中,所述第一类反馈结果对应的关键词用于作为负样本提供给所述神经网络模型进行增强学习;

当查询结果为是时,确定所述用户终端针对推送的所述关键词触发的反馈结果为第二类反馈结果;其中,所述第二类反馈结果对应的关键词用于作为正样本提供给所述神经网络模型进行增强学习。

5. 根据权利要求4所述的方法,其中,当查询结果为是时,进一步包括:确定所述用户日志文件中包含的与所述关键词对应的点击事件的次数,根据次数确定第二类反馈结果对应的结果分值;其中,所述结果分值用于确定对应的负样本的样本权重。

6. 根据权利要求4或5所述的方法,其中,所述获取用户终端针对推送的所述关键词触发的反馈结果的步骤具体包括:

每隔预设的时间间隔获取一次用户终端针对推送的所述关键词触发的反馈结果；和/或，

每当所述用户日志文件中录入点击事件时触发与所述点击事件对应的通知消息，通过订阅所述通知消息获取用户终端针对推送的所述关键词触发的反馈结果。

7. 根据权利要求1-5任一所述的方法，其中，所述神经网络模型包括以下中的至少一个：卷积神经网络、以及循环神经网络；

其中，所述卷积神经网络包括：至少一个卷积核，用于对所述文档内容进行特征提取；

所述循环神经网络包括：长短期记忆网络，其中，所述长短期记忆网络中的当前时间点的输入数据为上一个时间点的输出数据。

8. 根据权利要求1-5任一所述的方法，其中，所述增强学习的算法包括反向传播算法。

9. 根据权利要求1-5任一所述的方法，其中，所述方法执行之前，进一步包括步骤：从预设的分布式消息队列中获取文档。

10. 一种基于文档的关键词推送装置，包括：

输入模块，适于每当获取到文档时，将获取到的文档所对应的文档内容输入预设的神经网络模型；

第一获取模块，适于获取所述神经网络模型的输出结果；

确定模块，适于根据所述输出结果确定所述文档的关键词，并对所述关键词进行推送；

第二获取模块，适于获取用户终端针对推送的所述关键词触发的反馈结果；

增强学习模块，适于根据所述反馈结果对所述预设的神经网络模型进行增强学习；

调整模块，适于根据所述增强学习的学习结果对所述预设的神经网络模型进行调整；

其中，对关键词进行推送时，为各个关键词设置多个不同的点击操作入口；在根据用户终端针对推送的关键词触发的反馈结果对预设的神经网络模型进行增强学习时，进一步将查询到的用户日志文件中包含的与关键词对应的点击事件细分为多种类型的点击操作事件，预先为每种类型的点击操作事件设置不同的分析优先级，并根据不同类型的点击操作事件的分析优先级设置对应的反馈结果的结果分值，该结果分值用于确定对应的样本的样本权重；其中，所述点击操作事件包括：评分点击操作事件、浏览点击操作事件和搜索点击操作事件。

11. 根据权利要求10所述的装置，其中，所述装置进一步包括：

预处理模块，适于对所述获取到的文档进行预处理，将预处理结果作为所述文档对应的文档内容；其中，所述预处理包括以下中的至少一个：分词处理、以及词向量转换处理。

12. 根据权利要求10所述的装置，其中，所述确定模块具体用于：

根据预设的筛选策略对所述输出结果进行筛选，将筛选结果确定为所述文档的关键词；

其中，所述筛选策略包括以下中的至少一个：根据词性进行筛选的策略；以及，根据词频进行筛选的策略。

13. 根据权利要求10所述的装置，其中，所述装置进一步包括：

监测模块，适于当监测到用户终端针对所述关键词触发的点击事件时，将所述点击事件记录到用户日志文件中；

则所述第二获取模块具体包括：

查询单元,适于查询所述用户日志文件中是否包含与所述关键词对应的点击事件;

第一确定单元,适于当查询结果为否时,确定所述用户终端针对推送的所述关键词触发的反馈结果为第一类反馈结果;其中,所述第一类反馈结果对应的关键词用于作为负样本提供给所述神经网络模型进行增强学习;

第二确定单元,适于当查询结果为是时,确定所述用户终端针对推送的所述关键词触发的反馈结果为第二类反馈结果;其中,所述第二类反馈结果对应的关键词用于作为正样本提供给所述神经网络模型进行增强学习。

14. 根据权利要求13所述的装置,其中,所述第二确定单元进一步用于:

确定所述用户日志文件中包含的与所述关键词对应的点击事件的次数,根据次数确定第二类反馈结果对应的结果分值;其中,所述结果分值用于确定对应的负样本的样本权重。

15. 根据权利要求13或14所述的装置,其中,所述第二获取模块进一步用于:

每隔预设的时间间隔获取一次用户终端针对推送的所述关键词触发的反馈结果;和/或,

每当所述用户日志文件中录入点击事件时触发与所述点击事件对应的通知消息,通过订阅所述通知消息获取用户终端针对推送的所述关键词触发的反馈结果。

16. 根据权利要求10-14任一所述的装置,其中,所述神经网络模型包括以下中的至少一个:卷积神经网络、以及循环神经网络;

其中,所述卷积神经网络包括:至少一个卷积核,用于对所述文档内容进行特征提取;

所述循环神经网络包括:长短期记忆网络,其中,所述长短期记忆网络中的当前时间点的输入数据为上一个时间点的输出数据。

17. 根据权利要求10-14任一所述的装置,其中,所述增强学习的算法包括反向传播算法。

18. 根据权利要求10-14任一所述的装置,其中,所述装置进一步包括:第三获取模块,适于从预设的分布式消息队列中获取文档。

基于文档的关键词推送方法及装置

技术领域

[0001] 本发明涉及通信技术领域,具体涉及一种基于文档的关键词推送方法及装置。

背景技术

[0002] 随着互联网的飞速发展,越来越多的用户习惯于通过网络获取各类信息。其中,在用户获取文档信息时,为了方便用户的搜索和了解,会针对该文档信息推送对应的关键词,以方便用户根据关键词快速了解文档内容,从而为用户的阅读、搜索提供便利。

[0003] 在现有的关键词推送方式中,通常采用如下方式实现:首先,对文档中的内容进行分词处理;然后,从分词处理的结果中提取出现频率较高、或重要程度较高的词汇作为关键词进行推送。

[0004] 但是,发明人在实现本发明的过程中发现现有技术中的上述方案至少存在下述缺陷:通过简单的分词提取的方式只能将文档中明确出现的词汇确定为关键词,因此,无法从语义层面进行分析,从而无法对文档的内容进行准确地总结和概括,进而导致的推送的关键词无法准确地反映文档的核心内容。

发明内容

[0005] 鉴于上述问题,提出了本发明以便提供一种克服上述问题或者至少部分地解决上述问题的基于文档的关键词推送方法及装置。

[0006] 根据本发明的一个方面,提供了一种基于文档的关键词推送方法,包括:每当获取到文档时,将获取到的文档所对应的文档内容输入预设的神经网络模型;获取所述神经网络模型的输出结果,根据所述输出结果确定所述文档的关键词,并对所述关键词进行推送;获取用户终端针对推送的所述关键词触发的反馈结果,根据所述反馈结果对所述预设的神经网络模型进行增强学习;根据所述增强学习的学习结果对所述预设的神经网络模型进行调整。

[0007] 可选地,所述将获取到的文档所对应的文档内容输入预设的神经网络模型的步骤之前,进一步包括:对所述获取到的文档进行预处理,将预处理结果作为所述文档对应的文档内容;其中,所述预处理包括以下中的至少一个:分词处理、以及词向量转换处理。

[0008] 可选地,所述根据所述输出结果确定所述文档的关键词的步骤具体包括:根据预设的筛选策略对所述输出结果进行筛选,将筛选结果确定为所述文档的关键词;其中,所述筛选策略包括以下中的至少一个:根据词性进行筛选的策略;以及,根据词频进行筛选的策略。

[0009] 可选地,所述对所述关键词进行推送的步骤之后进一步包括:当监测到用户终端针对所述关键词触发的点击事件时,将所述点击事件记录到用户日志文件中;则所述获取用户终端针对推送的所述关键词触发的反馈结果的步骤具体包括:查询所述用户日志文件中是否包含与所述关键词对应的点击事件;当查询结果为否时,确定所述用户终端针对推送的所述关键词触发的反馈结果为第一类反馈结果;其中,所述第一类反馈结果对应的关

关键词用于作为负样本提供给所述神经网络模型进行增强学习；当查询结果为是时，确定所述用户终端针对推送的所述关键词触发的反馈结果为第二类反馈结果；其中，所述第二类反馈结果对应的关键词用于作为正样本提供给所述神经网络模型进行增强学习。

[0010] 可选地，当查询结果为是时，进一步包括：确定所述用户日志文件中包含的与所述关键词对应的点击事件的次数，根据次数确定第二类反馈结果对应的结果分值；其中，所述结果分值用于确定对应的负样本的样本权重。

[0011] 可选地，所述获取用户终端针对推送的所述关键词触发的反馈结果的步骤具体包括：每隔预设的时间间隔获取一次用户终端针对推送的所述关键词触发的反馈结果；和/或，每当所述用户日志文件中录入点击事件时触发与所述点击事件对应的通知消息，通过订阅所述通知消息获取用户终端针对推送的所述关键词触发的反馈结果。

[0012] 可选地，所述神经网络模型包括以下中的至少一个：卷积神经网络、以及循环神经网络；其中，所述卷积神经网络包括：至少一个卷积核，用于对所述文档内容进行特征提取；所述循环神经网络包括：长短期记忆网络，其中，所述长短期记忆网络中的当前时间点的输入数据为上一个时间点的输出数据。

[0013] 可选地，所述增强学习的算法包括反向传播算法。

[0014] 可选地，所述方法执行之前，进一步包括步骤：从预设的分布式消息队列中获取文档。

[0015] 根据本发明的另一个方面，提供了一种基于文档的关键词推送装置，包括：输入模块，适于每当获取到文档时，将获取到的文档所对应的文档内容输入预设的神经网络模型；第一获取模块，适于获取所述神经网络模型的输出结果；确定模块，适于根据所述输出结果确定所述文档的关键词，并对所述关键词进行推送；第二获取模块，适于获取用户终端针对推送的所述关键词触发的反馈结果；增强学习模块，适于根据所述反馈结果对所述预设的神经网络模型进行增强学习；调整模块，适于根据所述增强学习的学习结果对所述预设的神经网络模型进行调整。

[0016] 可选地，所述装置进一步包括：预处理模块，适于对所述获取到的文档进行预处理，将预处理结果作为所述文档对应的文档内容；其中，所述预处理包括以下中的至少一个：分词处理、以及词向量转换处理。

[0017] 可选地，所述确定模块具体用于：根据预设的筛选策略对所述输出结果进行筛选，将筛选结果确定为所述文档的关键词；其中，所述筛选策略包括以下中的至少一个：根据词性进行筛选的策略；以及，根据词频进行筛选的策略。

[0018] 可选地，所述装置进一步包括：监测模块，适于当监测到用户终端针对所述关键词触发的点击事件时，将所述点击事件记录到用户日志文件中；则所述第二获取模块具体包括：查询单元，适于查询所述用户日志文件中是否包含与所述关键词对应的点击事件；第一确定单元，适于当查询结果为否时，确定所述用户终端针对推送的所述关键词触发的反馈结果为第一类反馈结果；其中，所述第一类反馈结果对应的关键词用于作为负样本提供给所述神经网络模型进行增强学习；第二确定单元，适于当查询结果为是时，确定所述用户终端针对推送的所述关键词触发的反馈结果为第二类反馈结果；其中，所述第二类反馈结果对应的关键词用于作为正样本提供给所述神经网络模型进行增强学习。

[0019] 可选地，所述第二确定单元进一步用于：确定所述用户日志文件中包含的与所述

关键词对应的点击事件的次数,根据次数确定第二类反馈结果对应的结果分值;其中,所述结果分值用于确定对应的负样本的样本权重。

[0020] 可选地,所述第二获取模块进一步用于:每隔预设的时间间隔获取一次用户终端针对推送的所述关键词触发的反馈结果;和/或,每当所述用户日志文件中录入点击事件时触发与所述点击事件对应的通知消息,通过订阅所述通知消息获取用户终端针对推送的所述关键词触发的反馈结果。

[0021] 可选地,所述神经网络模型包括以下中的至少一个:卷积神经网络、以及循环神经网络;其中,所述卷积神经网络包括:至少一个卷积核,用于对所述文档内容进行特征提取;所述循环神经网络包括:长短期记忆网络,其中,所述长短期记忆网络中的当前时间点的输入数据为上一个时间点的输出数据。

[0022] 可选地,所述增强学习的算法包括反向传播算法。

[0023] 可选地,所述装置进一步包括:第三获取模块,适于从预设的分布式消息队列中获取文档。

[0024] 在本发明提供的一种基于文档的关键词推送方法及装置中,一方面,通过将获取到的文档所对应的文档内容输入预设的神经网络模型,并根据神经网络模型的输出结果确定并推送文档的关键词,能够利用神经网络模型的机器学习功能实现对文本语义关系的挖掘和概括,从而提取出更加准确的关键词。另一方面,通过获取用户终端针对推送的关键词触发的反馈结果,并根据反馈结果对预设的神经网络模型进行调整,能够根据用户反馈对神经网络模型进行更新,从而使神经网络模型不断优化,进而在后续过程中进一步提升关键词推送的准确性。由此可见,本发明中的方案能够有效提升关键词的推送质量,准确反映文档的核心内容。

[0025] 上述说明仅是本发明技术方案的概述,为了能够更清楚了解本发明的技术手段,而可依照说明书的内容予以实施,并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂,以下特举本发明的具体实施方式。

附图说明

[0026] 通过阅读下文优选实施方式的详细描述,各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的,而并不认为是对本发明的限制。而且在整个附图中,用相同的参考符号表示相同的部件。在附图中:

[0027] 图1示出了根据本发明实施例一提供的一种基于文档的关键词推送方法的流程图;

[0028] 图2示出了根据本发明实施例二提供的一种基于文档的关键词推送方法的流程图;

[0029] 图3示出了根据本发明实施例三提供的一种基于文档的关键词推送装置的结构框图。

[0030] 图4示出了根据本发明实施例四提供的一种基于文档的关键词推送装置的结构框图。

具体实施方式

[0031] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例,然而应当理解,可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反,提供这些实施例是为了能够更透彻地理解本公开,并且能够将本公开的范围完整的传达给本领域的技术人员。

[0032] 本发明提供了一种基于文档的关键词推送方法及装置,至少能够解决现有技术中由于无法根据文本内容中的语义来挖掘关键词而导致推送的关键词过于粗糙,不够准确的技术问题。

[0033] 实施例一

[0034] 图1示出了根据本发明实施例一提供的一种基于文档的关键词推送方法的流程图。如图1所示,该方法包括以下步骤:

[0035] 步骤S110:每当获取到文档时,将获取到的文档所对应的文档内容输入预设的神经网络模型。

[0036] 具体地,文档包括新闻、网页、邮件等各类文字信息,文档内容具体为针对上述文档进行一定的处理之后的处理结果。其中,文档内容可以为针对上述文档进行分词处理或者词向量转换处理等相关处理之后的处理结果,具体的处理方式可取决于神经网络模型的具体类型。

[0037] 预设的神经网络模型用于根据输入的文档内容来确定对应的关键词。本发明对神经网络模型所使用的具体算法以及训练方式不做限定。例如,在生成预设的神经网络模型时,可以将大量的原始语料作为训练样本集,通过对该训练样本集进行训练来生成上述神经网络模型。其中,上述预设的神经网络模型可以包括卷积神经网络以及循环神经网络等。

[0038] 步骤S120:获取神经网络模型的输出结果,根据输出结果确定文档的关键词,并对关键词进行推送。

[0039] 其中,上述输出结果具体包括对输入的文档内容进行语义分析、抽象以及概括处理等相关处理之后,针对该文档内容生成的一个或多个关键词。在获取输出结果之后,可以直接将输出结果作为对应文档的关键词,也可以先对输出结果进行进一步的筛选或排序等处理后,再根据处理结果确定对应文档的关键词。

[0040] 步骤S130:获取用户终端针对推送的关键词触发的反馈结果,根据反馈结果对预设的神经网络模型进行增强学习。

[0041] 具体地,用户终端在获取并显示推送的关键词后,若接收到用户针对推送的关键词执行预设操作的操作请求,则根据操作请求的次数和类型生成相应的反馈结果,该反馈结果用于提供给神经网络模型进行增强学习。其中,预设操作包括点击操作和/或搜索操作。其中,反馈结果可以通过用户日志的方式进行记录。

[0042] 步骤S140:根据增强学习的学习结果对预设的神经网络模型进行调整。

[0043] 具体地,根据步骤S130中增强学习的学习结果对预设的神经网络模型调整,使预设的神经网络模型能够根据动态的反馈结果来实现对应的动态调整,即:使神经网络模型能够针对用户的需求或者针对用户的兴趣来动态调整输出倾向,以此来进一步提高神经网络模型输出结果的准确性。其中,调整的参数可以包括关键词的权重、逻辑关系以及与其它相关词汇的填补关系等与关键词语义相关的关联关系。

[0044] 由此可见,在本发明提供的一种基于文档的关键词推送方法中,一方面,通过将获取到的文档所对应的文档内容输入预设的神经网络模型,并根据神经网络模型的输出结果确定并推送文档的关键词,能够利用神经网络模型的机器学习功能实现对文本语义关系的挖掘和概括,从而提取出更加准确的关键词。另一方面,通过获取用户终端针对推送的关键词触发的反馈结果,并根据反馈结果对预设的神经网络模型进行调整,能够根据用户对神经网络模型进行更新,从而使神经网络模型不断优化,进而在后续过程中进一步提升关键词推送的准确性。

[0045] 实施例二

[0046] 图2示出了根据本发明实施例二提供的一种基于文档的关键词推送方法流程图。如图2所示,该方法包括以下步骤:

[0047] 步骤S210:对获取到的文档进行预处理,将预处理结果作为文档对应的文档内容。

[0048] 具体地,为了便于存取,可以将文档存储在分布式消息队列中,从而利用分布式消息队列获取文档,并对获取到的文档进行预处理。其中,预处理包括以下中的至少一个:分词处理、以及词向量转换处理。分词处理具体为:在获取到文档之后,将文档语句中包含的词进行词性划分、词义划分等拆分处理,抽取并保留其中的关键词。例如,若文档中包含语句为“原子能应用与发展的重要意义”,则对上述语句进行分词处理之后的结果为:“原子能”、“应用”、“发展”、“重要意义”。词向量转换处理具体为:通过向量形式表示各个词汇之间的关系,从而把文本处理简化为向量运算,通过计算向量空间上的相似度来表示文本语义上的相似度。因此,词向量能够在一定程度上刻画出词与词之间的语义距离。发明人在实现本发明的过程中发现,通过将文档转换为词向量的方式能够显著提升神经网络模型的预测效率。其中,词向量可以通过多种方式得到,例如,可以将大量文章作为训练词向量的训练数据集,通过训练文章中的词汇得到词向量;也可以根据各个词汇的出现频率来确定词向量。具体实施中,在针对获取到的文档进行预处理时,可以根据实际情况来选择上述处理方式中的任意一种方式,或者,也可以将上述两种方式进行结合和优化,从而实现更优的预处理效果。

[0049] 步骤S220:每当获取到文档时,将获取到的文档所对应的文档内容输入预设的神经网络模型。

[0050] 具体地,预设的神经网络模型一般包含三层结构:输入层、输出层以及位于输入层和输出层之间的隐藏层。其中,上述输入层用于接收输入信息,为神经网络模型中的输入端口;上述输出层用于输出处理结果,为神经网络模型中的输出端口;隐藏层位于输入层和输出层之间,具体用于对上述输入信息进行特征提取。具体实施中,在建立预设的神经网络模型时,可以通过对大量的原始语料进行训练来生成上述神经网络模型。

[0051] 在本实施例中,预设的神经网络模型为一种基于深度学习的神经网络模型,其包括以下中的至少一个:卷积神经网络(CNN)、以及循环神经网络(RNN)。其中,卷积神经网络包括:至少一个卷积核,该卷积核用于对文档内容进行特征提取。循环神经网络包括:长短期记忆网络,其中,长短期记忆网络中的当前时间点的输入数据为上一个时间点的输出数据。

[0052] 具体地,卷积神经网络是一种基于深度学习建立的架构。卷积神经网络中可以包含多个不同的卷积核,卷积核可以通过与指定对象进行卷积运算来提取指定对象的不同特

征,其具体过程可以为:将卷积核与文档内容做卷积,其卷积结果即为对文档内容进行特征提取的结果,即:卷积结果为针对文档内容中对应关键字的抽取和概括的结果。具体实施中,本领域技术人员可以根据需要提取的文档内容的特征来设置对应的卷积核。其中,在上述卷积过程中,文档内容中的信息由于与卷积核进行卷积运算而被赋予不同的权重,进而能够针对上述赋予的权重信息来对文档中的内容进行特征提取。并且,由于卷积神经网络中的卷积核具有卷积共享性(即针对于上述权值来讲,具有权重共享性),因此在本实施例中,应用卷积神经网络进行特征提取能够有效降低神经网络模型的复杂度,减少权值的数量,有效避免传统识别算法中复杂的特征提取和数据重建过程。具体地,在将获取到的文档所对应的文档内容输入上述卷积神经网络时,具体为将文档内容以每篇文章为单位输入上述卷积神经网络。

[0053] 循环神经网络是一种用来处理序列数据的网络,其针对的问题为:当两个关联数据之间的距离较远时,如何使这两个关联数据之间产生关联关系。具体地,循环神经网络的具体表现形式为:对前面的信息进行记忆并将上述记忆应用于当前输出的计算中。即:在循环神经网络中,隐藏层之间的节点不再是无连接的节点,而是有连接的节点;并且,隐藏层的输入不仅包括输入层的输出,还进一步包括上一时刻隐藏层的输出。在本实施例中,循环神经网络包括:长短期记忆网络(LSTM)。长短期记忆网络能够针对输入的文档内容进行转码操作,并对文档内容进行拟合。在长短期记忆网络中,当前时间点的输入数据为上一个时间点的输出数据,也就是说,长短期记忆网络能够将上一个时间点输出的文档内容连接至当前时间点的文档内容学习中,增加了神经网络模型学习与预测语言的能力,使深度学习能够适应更加复杂的语言场景。具体地,在将获取到的文档所对应的文档内容输入长短期记忆网络时,具体为将文档内容以句子为单位输入长短期记忆网络。

[0054] 步骤S230:根据预设的筛选策略对输出结果进行筛选,将筛选结果确定为文档的关键词,并对上述关键词进行推送。

[0055] 其中,预设的筛选策略包括以下中的至少一个:根据词性进行筛选的策略;以及,根据词频进行筛选的策略。具体地,根据词性的筛选策略具体为:根据每个词在语法上的词性,比如动词、名词以及形容词等词性属性来筛选输出结果,确定输出结果中的关键词。例如,当关键词的词性为副词、连词、助词等词性时,由于这些词性的词一般都不属于实词,因此,可以通过词性筛选策略将属于副词、连词、助词等词性的关键词筛除。根据词频进行筛选的策略为:根据关键词出现的频率来筛除关键词。具体实施中,可以通过 TF-IDF (term frequency-inverse document frequency,词频--反转文件频率) 算法针对输出结果中的关键词的词频进行相关的权重计算和赋值,抽取出权重较大的关键词作为文档的关键词,并对确定的关键词进行推送。在本步骤中,针对输出结果中的关键词进行进一步的筛选处理,进一步提升了针对文档关键词推送的准确性。

[0056] 步骤S240:当监测到用户终端针对关键词触发的点击事件时,将点击事件记录到用户日志文件中。

[0057] 具体地,点击事件具体为用户在用户终端侧针对推送的关键词进行点击的操作事件。当监测到针对关键词触发的点击事件时,将上述点击事件记录在用户日志文件中,并据此来确定用户针对关键词的反馈信息。例如,当用户对关键词1进行点击操作之后,日志中自动记录关键词1的信息并在点击次数信息上进行+1操作。

[0058] 步骤S250:获取用户终端针对推送的关键词触发的反馈结果,根据反馈结果对预设的神经网络模型进行增强学习。

[0059] 其中,反馈结果可以通过步骤S240中的用户日志文件进行获取。具体地,反馈结果可以通过用户针对关键词的点击操作次数和/或点击操作频率得到。在获取反馈结果时,其具体步骤可以为:查询用户日志文件中是否包含与关键词对应的点击事件;当查询结果是否为否时,确定用户终端针对推送的关键词触发的反馈结果为第一类反馈结果;当查询结果为是时,确定用户终端针对推送的关键词触发的反馈结果为第二类反馈结果。其中,当查询结果为是时,进一步包括:确定用户日志文件中包含的与关键词对应的点击事件的次数,根据次数确定第二类反馈结果对应的结果分值。

[0060] 具体实现时,在对应的时间范围内查询用户日志文件,当查询到用户日志文件中并不包含与关键词对应的点击事件时,确定相应的反馈结果为第一类反馈结果。第一类反馈结果对应的关键词作为负样本提供给神经网络模型进行增强学习,神经网络模型通过第一类反馈结果能够确定推送效果不理想的关键词,并在后续的推送过程中改进相关的推送策略。

[0061] 当查询到用户日志文件中包含与关键词对应的点击事件时,确定相应的反馈结果为第二类反馈结果。并且,当反馈结果为第二类反馈结果时,进一步对用户日志文件中包含的与关键词对应的点击事件的记录进行+1操作。第二类反馈结果对应的关键词作为正样本提供给神经网络模型进行增强学习,神经网络模型通过第二类反馈结果能够确定推送效果理想的关键词,并在后续的推送过程中强化相关的推送策略。另外,在具体实施中,为了提升增强学习的学习效果,可以分别为各个正负样本设置不同的样本权重,例如,结果分值越高的第二类反馈结果所对应的正样本的样本权重也越高,从而能够将用户满意度高的关键词的推送策略进行重点强化,进而使后续的输出结果愈发迎合用户需求。

[0062] 进一步地,在获取用户终端针对推送的关键词触发的反馈结果时,可以每隔预设的时间间隔获取一次用户终端针对推送的关键词触发的反馈结果,例如,每隔1分钟自动获取一次用户终端针对推送的关键词触发反馈结果。具体实施时,预设的时间间隔可以根据关键词的数量和/或浏览频率进行设定,而且,该种方式尤其适用于浏览频率较固定的场景中。或者,在获取反馈结果时,每当用户日志文件中录入点击事件时触发与点击事件对应的通知消息,通过订阅通知消息获取用户终端针对推送的关键词触发的反馈结果。订阅通知消息的方式能够在第一时间接收到反馈结果,实时性较高,而且,该种方式尤其适用于关键词的浏览频率不固定的场景中,从而在浏览频率较低的时间段内避免查询开销。

[0063] 在获取反馈结果之后,根据反馈结果对预设的神经网络模型进行增强学习。其中,增强学习的算法包括:反向传播算法。具体地,在本实施例中,能够通过反向传播算法来监督神经网络模型的学习过程,该算法能够将训练输入送入网络以获得激励响应,将激励响应同训练输入对应的目标输出求差,从而获得隐藏层和输出层的响应误差,然后通过调节每个词向量的权重、参数等属性来对应调整每个词向量,使神经网络模型得以调整和修正。并且,除此之外,增强学习的方式还可以有多种,例如,可以设置一个回报函数,每当针对某关键词触发一次点击事件时,针对该关键词的回报结果+1或者回报结果为正等。例如,当反馈结果为第一类反馈结果时,相应的关键词的回报结果为负;当反馈结果为第二类反馈结果时,相应的关键词的回报结果为正,且第二类反馈结果对应的结果分值越高,相应的关键

词的回报结果也越高。在这里,增强学习的具体实现方式可以由本领域技术人员根据实际情况进行设置,本发明对此不作限制。

[0064] 步骤S260:根据增强学习的学习结果对预设的神经网络模型进行调整。

[0065] 具体地,在本步骤中,根据步骤S250中增强学习的学习结果对预设的神经网络模型调整,使预设的神经网络模型能够根据动态的反馈结果来实现对应的动态调整,即:使神经网络模型能够针对用户的需求或者针对用户的兴趣来动态调整输出倾向,以此来进一步提高神经网络模型输出结果的准确性。其中,调整的参数可以包括关键词的权重、逻辑关系以及与其它相关词汇的填补关系等与关键词语义相关的关联关系。

[0066] 由此可见,在上述方式中,能够利用神经网络模型自动学习,并且,该神经网络模型能够根据用户反馈而不断完善,从而使输出结果愈加精准。在此过程中,无需人工干预,减少了人工提取特征的时间成本和人力成本,通过机器学习的方式便捷准确地实现了关键词的推送。另外,由于神经网络模型能够持续地根据新增样本和用户反馈进行完善和更新,因而能够对新兴词汇进行学习,且能够感知到词汇之间的关联关系的动态变化情况,因而具备较高的时效性,尤其适用于新闻领域的关键词推送。

[0067] 另外,本领域技术人员还可以对上述实施例中的技术细节进行各种改动和变形。例如,对关键词进行推送时,可以为各个关键词设置多个不同的点击操作入口,以便于用户实现不同类型的操作。比如:用于实现用户评分的评分操作入口、用于打开并浏览与关键词对应的文档的浏览操作入口、用于根据该关键词进行相关搜索的搜索操作入口等。相应地,在步骤S250中根据用户终端针对推送的关键词触发的反馈结果对预设的神经网络模型进行增强学习时,可以进一步将查询到的用户日志文件中包含的与关键词对应的点击事件细分为多种类型的点击操作事件,比如细分为评分点击操作事件、浏览点击操作事件和搜索点击操作事件。并且,预先为每种类型的点击操作事件设置不同的分析优先级,并根据不同类型的点击操作事件的分析优先级设置对应的反馈结果的结果分值,该结果分值用于确定对应的样本的样本权重。由此可见,通过对反馈结果进行细化,能够更加准确地反映用户的满意度,进一步提升神经网络模型的准确率。

[0068] 综上所述,在本发明提供的基于文档的关键词推送方法中,首先对获取到的文档进行预处理,将预处理结果作为文档对应的文档内容,然后将获取到的文档所对应的文档内容输入预设的神经网络模型,并根据预设的筛选策略对输出结果进行筛选,将筛选结果确定为文档的关键词,并对上述关键词进行推送。并且,当监测到用户终端针对关键词触发的点击事件时,将点击事件记录到用户日志文件中,获取用户终端针对推送的关键词触发的反馈结果,根据反馈结果对预设的神经网络模型进行增强学习。由此可见,本发明中的方案解决了在针对文档进行关键词推送时,由于无法根据文本内容中的语义来挖掘关键词进而导致推送的关键词过于粗糙,不够准确并且时效性较低的问题,实现了针对文档中文本的语义关系、概念关系的来生成或者抽取文档的关键词的效果,并且能够针对用户的需求以及针对用户的兴趣来动态调整输出倾向,有效提高了文档关键词的推送质量。

[0069] 实施例三

[0070] 图3示出了根据本发明实施例三提供的一种基于文档的关键词推送装置的结构框图。如图3所示,该装置包括:输入模块31、第一获取模块32、确定模块33、第二获取模块34、增强学习模块35以及调整模块36。

[0071] 输入模块31适于每当获取到文档时,将获取到的文档所对应的文档内容输入预设的神经网络模型。

[0072] 具体地,文档包括新闻、网页、邮件等各类文字信息,文档内容具体为针对上述文档进行一定的处理之后的处理结果。其中,文档内容可以为针对上述文档进行分词处理或者词向量转换处理等相关处理之后的处理结果,具体的处理方式可取决于神经网络模型的具体类型。

[0073] 预设的神经网络模型用于根据输入的文档内容来确定对应的关键词。本发明对神经网络模型所使用的具体算法以及训练方式不做限定。例如,在生成预设的神经网络模型时,可以将大量的原始语料作为训练样本集,通过对该训练样本集进行训练来生成上述神经网络模型。其中,上述预设的神经网络模型可以包括卷积神经网络以及循环神经网络等。

[0074] 具体地,在获取到文档之后,将文档经过一定处理获取与文档所对应的文档内容,然后通过输入模块31将上述文档内容输入预设的神经网络模型。

[0075] 第一获取模块32适于获取神经网络模型的输出结果。

[0076] 具体地,输出结果具体包括对输入的文档内容进行语义分析、抽象以及概括处理等相关处理之后,针对该文档内容生成的一个或多个关键词。在输入模块31将文档内容输入预设的神经网络模型之后,第一获取模块32从上述神经网络模型中获取对应的输出结果。

[0077] 确定模块33适于根据输出结果确定文档的关键词,并对关键词进行推送。

[0078] 具体地,在第一获取模块32获取输出结果之后,确定模块33根据输出结果确定文档的关键词。其中,确定模块33可以直接将输出结果作为对应文档的关键词,也可以先对输出结果进行进一步的筛选或排序等处理后,再根据处理结果确定对应文档的关键词,然后将上述确定的关键词进行推送。

[0079] 第二获取模块34适于获取用户终端针对推送的关键词触发的反馈结果。

[0080] 具体地,用户终端在获取并显示推送的关键词后,第二获取模块34若接收到用户针对推送的关键词执行预设操作的操作请求,则根据操作请求的次数和类型生成相应的反馈结果,并将该反馈结果传送至增强学习模块35。其中,预设操作包括点击操作和/或搜索操作;反馈结果可以通过用户日志的方式进行记录。

[0081] 增强学习模块35适于根据反馈结果对预设的神经网络模型进行增强学习。

[0082] 具体地,增强学习模块35用于接收第二获取模块34发送的反馈结果,并根据上述反馈结果对预设的神经网络模型进行增强学习,以实现能够针对用户的反馈结果来实时调整神经网络模型的目的。

[0083] 调整模块36适于根据增强学习的学习结果对预设的神经网络模型进行调整。

[0084] 具体地,调整模块36根据增强学习模块35中增强学习的学习结果对预设的神经网络模型调整,使预设的神经网络模型能够根据动态的反馈结果来实现对应的动态调整,即:使神经网络模型能够针对用户的需求或者针对用户的兴趣来动态调整输出倾向,以此来进一步提高神经网络模型输出结果的准确性。其中,调整的参数可以包括关键词的权重、逻辑关系以及与其它相关词汇的填补关系等与关键词语义相关的关联关系。

[0085] 由此可见,在本发明提供的一种基于文档的关键词推送装置中,一方面,通过输入模块31将获取到的文档所对应的文档内容输入预设的神经网络模型,通过第一获取模块32

获取神经网络模型的输出结果,并通过确定模块 33根据神经网络模型的输出结果确定并推送文档的关键词,能够利用神经网络模型的机器学习功能实现对文本语义关系的挖掘和概括,从而提取出更加准确的关键词。另一方面,通过第二获取模块34获取用户终端针对推送的关键词触发的反馈结果,通过增强学习模块35根据反馈结果对预设的神经网络模型进行增强学习,并通过调整模块36根据反馈结果对预设的神经网络模型进行调整,能够根据用户反馈对神经网络模型进行更新,从而使神经网络模型不断优化,进而进一步提升关键词推送的准确性。

[0086] 实施例四

[0087] 图4示出了根据本发明实施例四提供的一种基于文档的关键词推送装置的结构框图。如图4所示,该装置包括:输入模块41、第一获取模块42、确定模块43、第二获取模块44、增强学习模块45、调整模块46、预处理模块 47、监测模块48以及第三获取模块49。其中,第二获取模块44进一步包括:查询单元441、第一确定单元442以及第二确定单元443。

[0088] 预处理模块47适于对获取到的文档进行预处理,将预处理结果作为文档对应的文档内容;其中,预处理包括以下中的至少一个:分词处理、以及词向量转换处理。

[0089] 具体地,预处理模块47用于对获取到的文档进行预处理。其中,预处理包括以下中的至少一个:分词处理、以及词向量转换处理。分词处理具体为:在获取到文档之后,将文档语句中包含的词进行词性划分、词义划分等拆分处理,抽取并保留其中的关键词。例如,若文档中包含语句为“原子能应用与发展的重要意义”,则对上述语句进行分词处理之后的结果为:“原子能”、“应用”、“发展”、“重要意义”。词向量转换处理具体为:通过向量形式表示各个词汇之间的关系,从而把文本处理简化为向量运算,通过计算向量空间上的相似度来表示文本语义上的相似度。因此,词向量能够在一定程度上刻画出词与词之间的语义距离。发明人在实现本发明的过程中发现,通过将文档转换为词向量的方式能够显著提升神经网络模型的预测效率。其中,词向量可以通过多种方式得到,例如,可以将大量文章作为训练词向量的训练数据集,通过训练文章中的词汇得到词向量;也可以根据各个词汇的出现频率来确定词向量。具体实施中,预处理模块47在针对获取到的文档进行预处理时,可以根据实际情况来选择上述处理方式中的任意一种方式,或者,也可以将上述两种方式进行结合和优化,从而实现更优的预处理效果。

[0090] 输入模块41适于每当获取到文档时,将获取到的文档所对应的文档内容输入预设的神经网络模型。

[0091] 具体地,预设的神经网络模型一般包含三层结构:输入层、输出层以及位于输入层和输出层之间的隐藏层。其中,上述输入层用于接收输入信息,为神经网络模型中的输入端口;上述输出层用于输出处理结果,为神经网络模型中的输出端口;隐藏层位于输入层和输出层之间,具体用于对上述输入信息进行特征提取。具体实施中,在建立预设的神经网络模型时,可以通过对大量的原始语料进行训练来生成上述神经网络模型。

[0092] 在本实施例中,预设的神经网络模型为一种基于深度学习的神经网络模型,其包括以下中的至少一个:卷积神经网络(CNN)、以及循环神经网络(RNN)。其中,卷积神经网络包括:至少一个卷积核,该卷积核用于对文档内容进行特征提取。循环神经网络包括:长短期记忆网络,其中,长短期记忆网络中的当前时间点的输入数据为上一个时间点的输出数据。

[0093] 具体地,卷积神经网络是一种基于深度学习建立的架构。卷积神经网络中可以包含多个不同的卷积核,卷积核可以通过与指定对象进行卷积运算来提取指定对象的不同特征,其具体过程可以为:将卷积核与文档内容做卷积,其卷积结果即为对文档内容进行特征提取的结果,即:卷积结果为针对文档内容中对应关键字的抽取和概括的结果。具体实施中,本领域技术人员可以根据需要提取的文档内容的特征来设置对应的卷积核。其中,在上述卷积过程中,文档内容中的信息由于与卷积核进行卷积运算而被赋予不同的权重,进而能够针对上述赋予的权重信息来对文档中的内容进行特征提取。并且,由于卷积神经网络中的卷积核具有卷积共享性(即针对于上述权值来讲,具有权重共享性),因此在本实施例中,应用卷积神经网络进行特征提取能够有效降低神经网络模型的复杂度,减少权值的数量,有效避免传统识别算法中复杂的特征提取和数据重建过程。具体地,输入模块41在将获取到的文档所对应的文档内容输入上述卷积神经网络时,具体为将文档内容以每篇文章为单位输入上述卷积神经网络。

[0094] 循环神经网络是一种用来处理序列数据的网络,其针对的问题为:当两个关联数据之间的距离较远时,如何使这两个关联数据之间产生关联关系。具体地,循环神经网络的具体表现形式为:对前面的信息进行记忆并将上述记忆应用于当前输出的计算中。即:在循环神经网络中,隐藏层之间的节点不再是无连接的节点,而是有连接的节点;并且,隐藏层的输入不仅包括输入层的输出,还进一步包括上一时刻隐藏层的输出。在本实施例中,循环神经网络包括:长短期记忆网络(LSTM)。长短期记忆网络能够针对输入的文档内容进行转码操作,并对文档内容进行拟合。在长短期记忆网络中,当前时间点的输入数据为上一个时间点的输出数据,也就是说,长短期记忆网络能够将上一个时间点输出的文档内容连接至当前时间点的文档内容学习中,增加了神经网络模型学习与预测语言的能力,使深度学习能够适应更加复杂的语言场景。具体地,输入模块41在将获取到的文档所对应的文档内容输入长短期记忆网络时,具体为将文档内容以句子为单位输入长短期记忆网络。

[0095] 第一获取模块42适于获取神经网络模型的输出结果。

[0096] 具体地,输出结果具体包括对输入的文档内容进行语义分析、抽象以及概括处理等相关处理之后,针对该文档内容生成的一个或多个关键词。在输入模块41将文档内容输入预设的神经网络模型之后,第一获取模块42从上述神经网络模型获取对应的输出结果。

[0097] 确定模块43适于根据输出结果确定文档的关键词,并对关键词进行推送。

[0098] 具体地,在第一获取模块42获取输出结果之后,确定模块43根据输出结果确定文档的关键词。其中,确定模块43可以直接将输出结果作为对应文档的关键词,也可以根据预设的筛选策略对输出结果进行筛选,将筛选结果确定为文档的关键词,并对上述关键词进行推送。其中,针对上述第二种情况,预设的筛选策略包括以下中的至少一个:根据词性进行筛选的策略;以及,根据词频进行筛选的策略。具体地,根据词性的筛选策略具体为:根据每个词在语法上的词性,比如动词、名词以及形容词等词性属性来筛选输出结果,确定输出结果中的关键词。例如,当关键词的词性为副词、连词、助词等词性时,由于这些词性的词一般都不属于实词,因此,可以通过词性筛选策略将属于副词、连词、助词等词性的关键词筛除。根据词频进行筛选的策略为:根据关键词出现的频率来筛除关键词。具体实施中,可以通过TF-IDF (term frequency-inverse document frequency,词频--反转文件频率)算法针对输出结果中的关键词的词频进行相关的权重计算和赋值,抽取出权重较大的关键词作

为文档的关键词,并对确定的关键词进行推送。在这里,确定模块 42 针对输出结果中的关键词进行进一步的筛选处理,进一步提升了针对文档关键词推送的准确性。

[0099] 第二获取模块44适于获取用户终端针对推送的关键词触发的反馈结果。

[0100] 其中,反馈结果可以通过监测模块48中的用户日志文件进行获取。具体地,反馈结果可以通过用户针对关键词的点击操作次数和/或点击操作频率得到。第二获取模块44进一步包括查询单元441、第一确定单元442以及第二确定单元443。在获取反馈结果时,首先通过查询单元441在对应的时间范围内查询用户日志文件是否包含对应的点击事件,当查询结果为否时,通过第一确定单元442确定用户终端针对推送的关键词触发的反馈结果为第一类反馈结果,将第一类反馈结果对应的关键词作为负样本提供给神经网络模型进行增强学习,以供神经网络模型通过第一类反馈结果能够确定推送效果不理想的关键词,并在后续的推送过程中改进相关的推送策略;当查询结果为是时,通过第二确定单元443确定用户终端针对推送的关键词触发的反馈结果为第二类反馈结果;并且,当查询结果为是时,第二确定单元443还可以进一步确定用户日志文件中包含的与关键词对应的点击事件的次数,根据次数确定第二类反馈结果对应的结果分值。具体地,当反馈结果为第二类反馈结果时,第二确定单元443对用户日志文件中包含的与关键词对应的点击事件的记录进行+1操作,将第二类反馈结果对应的关键词作为正样本提供给神经网络模型进行增强学习,以供神经网络模型通过第二类反馈结果能够确定推送效果理想的关键词,并在后续的推送过程中强化相关的推送策略。

[0101] 具体地,第二获取模块44在获取用户终端针对推送的关键词触发的反馈结果时,可以每隔预设的时间间隔获取一次用户终端针对推送的关键词触发的反馈结果,例如,每隔1分钟自动获取一次用户终端针对推送的关键词触发反馈结果。具体实施时,预设的时间间隔可以根据关键词的数量和/或浏览频率进行设定,而且,该种方式尤其适用于浏览频率较固定的场景中。或者,第二获取模块44在获取反馈结果时,每当用户日志文件中录入点击事件时触发与点击事件对应的通知消息,通过订阅通知消息获取用户终端针对推送的关键词触发的反馈结果。订阅通知消息的方式能够在第一时间接收到反馈结果,实时性较高,而且,该种方式尤其适用于关键词的浏览频率不固定的场景中,从而在浏览频率较低的时间段内避免查询开销。

[0102] 增强学习模块45适于根据反馈结果对预设的神经网络模型进行增强学习。

[0103] 具体地,增强学习的算法包括:反向传播算法。在本实施例中,增强学习模块45能够通过反向传播算法来监督神经网络模型的学习过程,该算法能够将训练输入送入网络以获得激励响应,将激励响应同训练输入对应的目标输出求差,从而获得隐藏层和输出层的响应误差,然后通过调节每个词向量的权重、参数等属性来对应调整每个词向量,使神经网络模型得以调整和修正。并且,除此之外,增强学习模块45增强学习的方式还可以有多种,例如,可以设置一个回报函数,每当针对某关键词触发一次点击事件时,针对该关键词的回报结果+1或者回报结果为正等。例如,当反馈结果为第一类反馈结果时,相应的关键词的回报结果为负;当反馈结果为第二类反馈结果时,相应的关键词的回报结果为正,且第二类反馈结果对应的结果分值越高,相应的关键词的回报结果也越高。在这里,增强学习的具体实现方式可以由本领域技术人员根据实际情况进行设置,本发明对此不作限制。

[0104] 调整模块46适于根据增强学习的学习结果对预设的神经网络模型进行调整。

[0105] 具体地,调整模块46根据增强学习模块45中增强学习的学习结果对预设的神经网络模型调整,使预设的神经网络模型能够根据动态的反馈结果来实现对应的动态调整,即:增强学习模块45使神经网络模型能够针对用户的需求或者针对用户的兴趣来动态调整输出倾向,以此来进一步提高神经网络模型输出结果的准确性。其中,调整的参数可以包括关键词的权重、逻辑关系以及与其它相关词汇的填补关系等与关键词语义相关的关联关系。

[0106] 监测模块48适于当监测到用户终端针对关键词触发的点击事件时,将点击事件记录到用户日志文件中。

[0107] 具体地,点击事件具体为用户在用户终端侧针对推送的关键词进行点击的操作事件。当监测到针对关键词触发的点击事件时,监测模块48将上述点击事件记录在用户日志文件中,并据此来确定用户针对关键词的反馈信息。例如,当用户对关键词1进行点击操作之后,日志中自动记录关键词1的信息并在点击次数信息上进行+1操作。

[0108] 第三获取模块49适于从预设的分布式消息队列中获取文档。

[0109] 具体地,为了便于存取,可以将文档存储在分布式消息队列中,通过第三获取模块49从分布式消息队列获取文档,并将获取到的文档传送至预处理模块47进行预处理。

[0110] 此外,上述各个模块的具体结构和工作原理可参照方法实施例相应步骤的描述,此处不再赘述。

[0111] 综上所述,本发明提供的基于文档的关键词推送装置中,一方面,首先通过预处理模块47对获取到的文档进行预处理,将预处理结果作为文档对应的文档内容,然后通过输入模块41将获取到的文档所对应的文档内容输入预设的神经网络模型,并通过第一获取模块42、确定模块43确定文档内容中的关键词,并对上述关键词进行推送。另一方面,通过监测模块48监测到用户终端针对关键词触发的点击事件,将点击事件记录到用户日志文件中作为反馈结果,并通过第二获取模块44获取用户终端针对推送的关键词触发的反馈结果,通过增强学习模块45根据反馈结果对预设的神经网络模型进行增强学习,并通过调整模块46根据增强学习的学习结果对预设的神经网络模型进行调整。由此可见,本发明中的方案解决了在针对文档进行关键词推送时,由于无法根据文本内容中的语义来挖掘关键词而导致推送的关键词过于粗糙,不够准确的问题,实现了针对文档中文本的语义关系、概念关系的来生成或者抽取文档的关键词的效果,并且能够针对用户的需求以及针对用户的兴趣来动态调整输出倾向,有效提高了文档关键词的推送质量。

[0112] 在此提供的算法和显示不与任何特定计算机、虚拟系统或者其它设备固有相关。各种通用系统也可以与基于在此的示教一起使用。根据上面的描述,构造这类系统所要求的结构是显而易见的。此外,本发明也不针对任何特定编程语言。应当明白,可以利用各种编程语言实现在此描述的本发明的内容,并且上面对特定语言所做的描述是为了披露本发明的最佳实施方式。

[0113] 在此处所提供的说明书中,说明了大量具体细节。然而,能够理解,本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中,并未详细示出公知的方法、结构和技术,以便不模糊对本说明书的理解。

[0114] 类似地,应当理解,为了精简本公开并帮助理解各个发明方面中的一个或多个,在上面对本发明的示例性实施例的描述中,本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而,并不应将该公开的方法解释成反映如下意图:即所要求保

护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说,如下面的权利要求书所反映的那样,发明方面在于少于前面公开的单个实施例的所有特征。因此,遵循具体实施方式的权利要求书由此明确地并入该具体实施方式,其中每个权利要求本身都作为本发明的单独实施例。

[0115] 本领域那些技术人员可以理解,可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件,以及此外可以把它分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元中的至少一些是相互排斥之外,可以采用任何组合对本说明书(包括伴随的权利要求、摘要和附图)中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述,本说明书(包括伴随的权利要求、摘要和附图)中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

[0116] 此外,本领域的技术人员能够理解,尽管在此所述的一些实施例包括其它实施例中包括的某些特征而不是其它特征,但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如,在下面的权利要求书中,所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

[0117] 本发明的各个部件实施例可以以硬件实现,或者以在一个或者多个处理器上运行的软件模块实现,或者以它们的组合实现。本领域的技术人员应当理解,可以在实践中使用微处理器或者数字信号处理器(DSP)来实现根据本发明实施例的基于文档的关键词推送装置模块中的一些或者全部部件的一些或者全部功能。本发明还可以实现为用于执行这里所描述的方法的一部分或者全部的设备或者装置程序(例如,计算机程序和计算机程序产品)。这样的实现本发明的程序可以存储在计算机可读介质上,或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下下载得到,或者在载体信号上提供,或者以任何其他形式提供。

[0118] 应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制,并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中,不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中,这些装置中的若干个可以通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

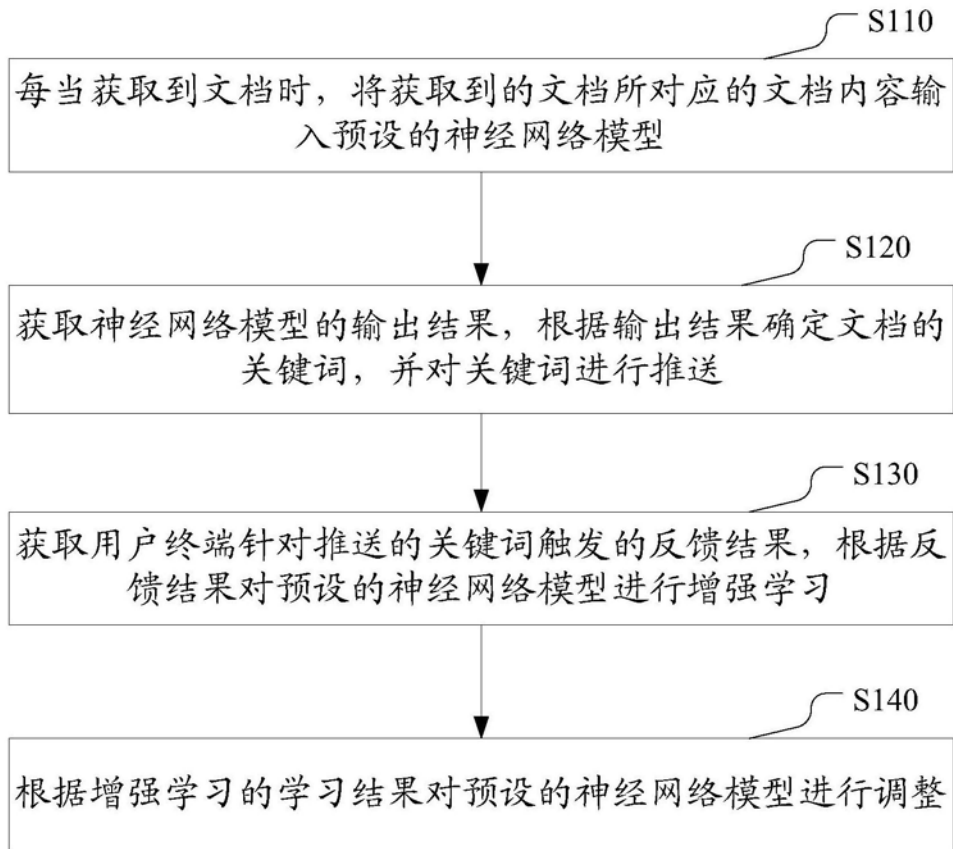


图1

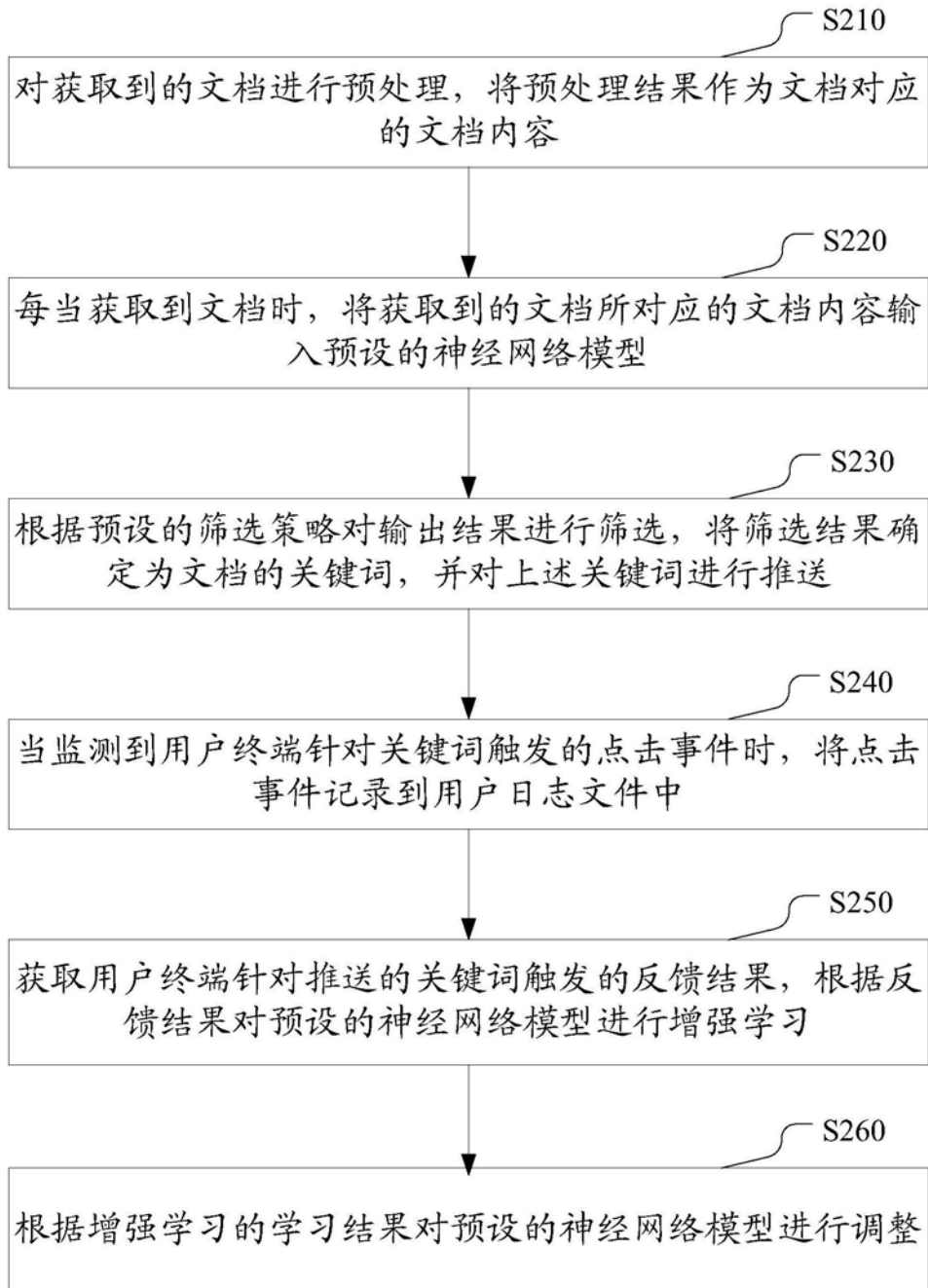


图2

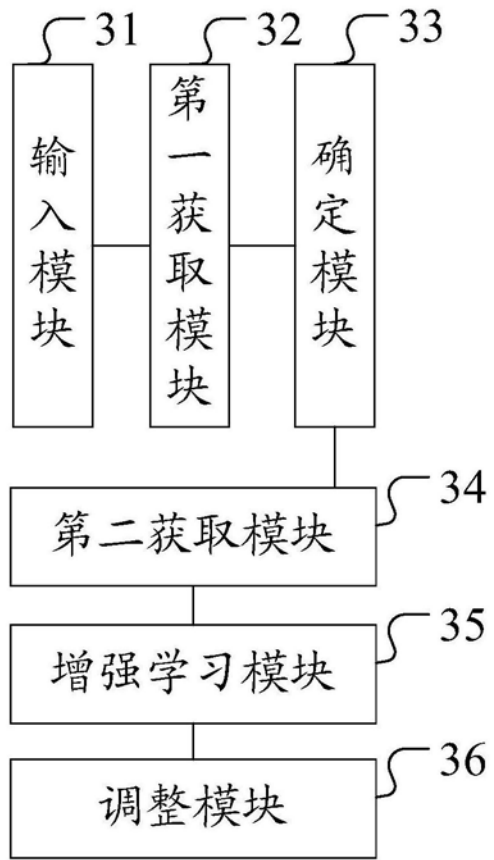


图3

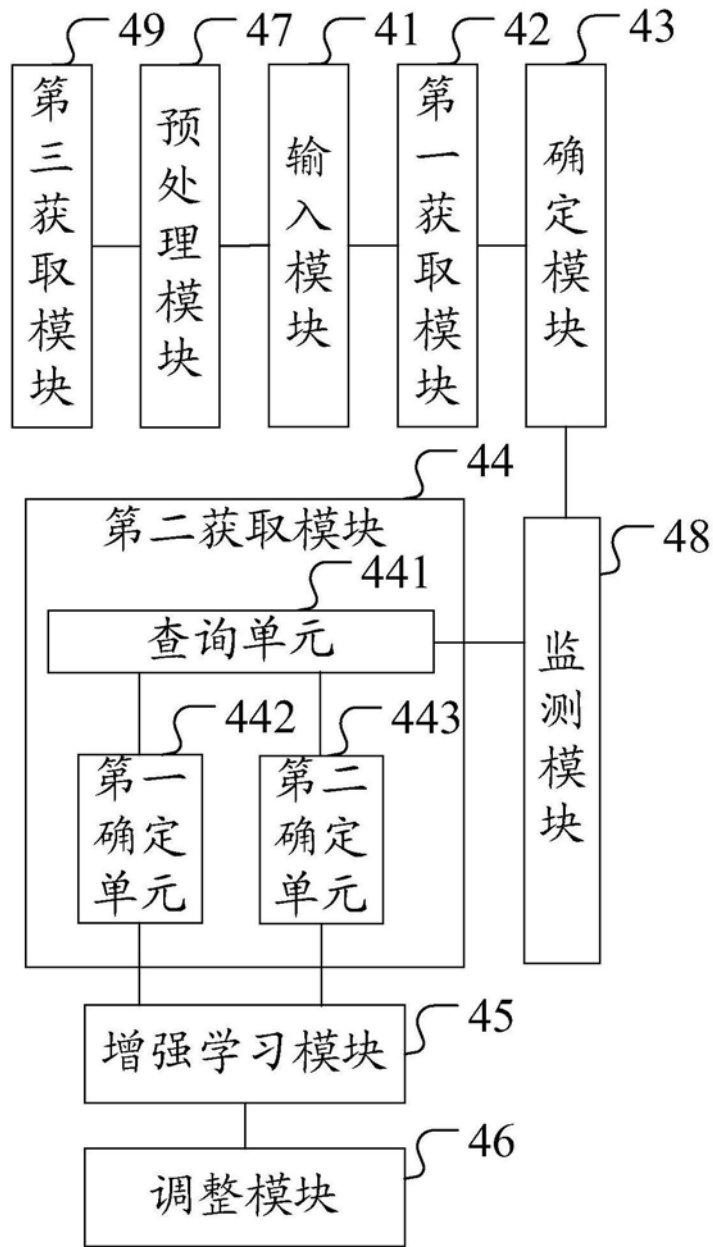


图4