



(12) 发明专利申请

(10) 申请公布号 CN 114722774 A

(43) 申请公布日 2022.07.08

(21) 申请号 202210361729.6

(22) 申请日 2022.04.07

(71) 申请人 平安科技(深圳)有限公司

地址 518000 广东省深圳市福田区福田街
道福安社区益田路5033号平安金融中
心23楼

(72) 发明人 李平 马骏 王少军

(74) 专利代理机构 广州嘉权专利商标事务所有
限公司 44205

专利代理师 陈春芹

(51) Int. Cl.

G06F 40/126 (2020.01)

G06F 40/30 (2020.01)

G06N 20/00 (2019.01)

H03M 7/30 (2006.01)

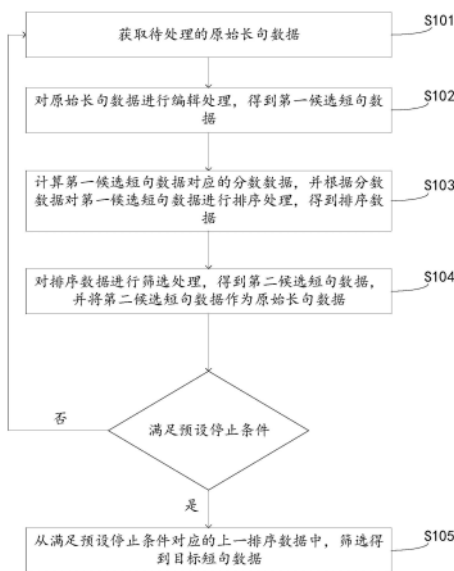
权利要求书2页 说明书14页 附图5页

(54) 发明名称

数据压缩方法、装置、电子设备及存储介质

(57) 摘要

本申请实施例提供了一种数据压缩方法、装置、电子设备及存储介质,属于人工智能技术领域。该方法包括:获取待处理的原始长句数据;对所述原始长句数据进行编辑处理,得到第一候选短句数据;计算所述第一候选短句数据对应的分数数据,并根据所述分数数据对所述第一候选短句数据进行排序处理,得到排序数据;对所述排序数据进行筛选处理,得到第二候选短句数据,并将所述第二候选短句数据作为所述原始长句数据,重新获取待处理的所述原始长句数据直至满足预设停止条件;从满足所述预设停止条件对应的上一所述排序数据中,筛选得到目标短句数据,本申请实施例能够有效提高数据的压缩效率。



1. 一种数据压缩方法,其特征在于,所述方法包括:
 - 获取待处理的原始长句数据;
 - 对所述原始长句数据进行编辑处理,得到第一候选短句数据;
 - 计算所述第一候选短句数据对应的分数数据,并根据所述分数数据对所述第一候选短句数据进行排序处理,得到排序数据;
 - 对所述排序数据进行筛选处理,得到第二候选短句数据,并将所述第二候选短句数据作为所述原始长句数据,重新获取待处理的所述原始长句数据直至满足预设停止条件;
 - 从满足所述预设停止条件对应的上一所述排序数据中,筛选得到目标短句数据。
2. 根据权利要求1所述的数据压缩方法,其特征在于,在所述对所述原始长句数据进行编辑处理,得到第一候选短句数据的步骤之前,包括:
 - 对所述原始长句数据进行短语结构分析操作,得到短语结构树,其中,所述短语结构树包括从句数据和/或短语数据。
3. 根据权利要求2所述的数据压缩方法,其特征在于,所述对所述原始长句数据进行编辑处理,得到第一候选短句数据的步骤,包括以下至少之一:
 - 对所述原始长句数据中与所述短语结构树对应相同的数据进行删除操作,得到第一候选短句数据;或者,
 - 在所述短语结构树包括从句数据的情况下,从所述短语结构树中提取所述从句数据,作为所述第一候选短句数据;或者,
 - 在所述短语结构树包括短语数据的情况下,从所述短语结构树中提取所述短语数据,作为所述第一候选短句数据。
4. 根据权利要求1至3任一项所述的数据压缩方法,其特征在于,所述计算所述第一候选短句数据对应的分数数据的步骤,包括:
 - 对所述第一候选短句数据与所述原始长句数据进行语义相似度计算,得到语义相似度数据;
 - 通过统计语言模型对所述第一候选短句数据进行语句通顺度计算,得到语句通顺度数据;
 - 对所述第一候选短句数据与所述原始长句数据进行距离计算和归一化处理,得到候选短句长度数据;
 - 对所述语义相似度数据、所述语句通顺度数据和所述候选短句长度数据进行乘积计算,得到所述第一候选短句数据对应的分数数据。
5. 根据权利要求4所述的数据压缩方法,其特征在于,所述对所述第一候选短句数据与所述原始长句数据进行语义相似度计算,得到语义相似度数据的步骤,包括:
 - 计算所述第一候选短句数据对应的第一句向量和所述原始长句数据对应的第二句向量;
 - 对所述第一句向量与所述第二句向量进行余弦相似度计算,得到余弦相似度数据。
6. 根据权利要求4所述的数据压缩方法,其特征在于,所述对所述第一候选短句数据与所述原始长句数据进行距离计算和归一化处理,得到候选短句长度数据的步骤,包括:
 - 计算所述第一候选短句数据对应的第一句子长度和所述原始长句数据对应的第二句子长度;

对所述第二句子长度与所述第一句子长度进行减法计算,得到距离数据;

对所述距离数据进行归一化处理,得到候选短句长度数据。

7. 根据权利要求1至3任一项所述的数据压缩方法,其特征在于,所述预设停止条件为:

当前排序数据中的第二候选短句数据对应的最大分数数据,小于前一排序数据中的第二候选短句数据对应的最大分数数据。

8. 一种数据压缩装置,其特征在于,所述装置包括:

数据获取模块,用于获取待处理的原始长句数据;

编辑处理模块,用于对所述原始长句数据进行编辑处理,得到第一候选短句数据;

排序处理模块,用于计算所述第一候选短句数据对应的分数数据,并根据所述分数数据对所述第一候选短句数据进行排序处理,得到排序数据;

循环处理模块,用于对所述排序数据进行筛选处理,得到第二候选短句数据,并将所述第二候选短句数据作为所述原始长句数据,重新获取待处理的所述原始长句数据直至满足预设停止条件;

目标筛选模块,用于从满足所述预设停止条件对应的上一所述排序数据中,筛选得到目标短句数据。

9. 一种电子设备,其特征在于,所述电子设备包括存储器、处理器、存储在所述存储器上并可在所述处理器上运行的程序以及用于实现所述处理器和所述存储器之间的连接通信的数据总线,所述程序被所述处理器执行时实现如权利要求1至7任一项所述的数据压缩方法的步骤。

10. 一种存储介质,所述存储介质为计算机可读存储介质,用于计算机可读存储,其特征在于,所述存储介质存储有一个或者多个程序,所述一个或者多个程序可被一个或者多个处理器执行,以实现权利要求1至7中任一项所述的数据压缩方法的步骤。

数据压缩方法、装置、电子设备及存储介质

技术领域

[0001] 本申请涉及人工智能技术领域,尤其涉及一种数据压缩方法、装置、电子设备及存储介质。

背景技术

[0002] 相关技术中,针对长句数据的压缩方法通常是基于监督学习算法和Sequence to Sequence模型的(Sequence to Sequence模型是一种从序列到序列的转换模型框架,可应用在机器翻译、自动应答等场景中)。由于监督学习算法主要依赖于带标签的样本集,故在将带标签的样本集输入监督学习算法之前,还需要人工标注并行语料以获得该样本集。但当样本集数据量大、人工标注时间长时,容易影响数据的压缩效率。

发明内容

[0003] 本申请实施例的主要目的在于提出一种数据压缩方法、装置、电子设备及存储介质,能够有效提高数据的压缩效率。

[0004] 为实现上述目的,本申请实施例的第一方面提出了一种数据压缩方法,所述方法包括:

[0005] 获取待处理的原始长句数据;

[0006] 对所述原始长句数据进行编辑处理,得到第一候选短句数据;

[0007] 计算所述第一候选短句数据对应的分数数据,并根据所述分数数据对所述第一候选短句数据进行排序处理,得到排序数据;

[0008] 对所述排序数据进行筛选处理,得到第二候选短句数据,并将所述第二候选短句数据作为所述原始长句数据,重新获取待处理的所述原始长句数据直至满足预设停止条件;

[0009] 从满足所述预设停止条件对应的上一所述排序数据中,筛选得到目标短句数据。

[0010] 在一些实施例,在所述对所述原始长句数据进行编辑处理,得到第一候选短句数据的步骤之前,包括:

[0011] 对所述原始长句数据进行短语结构分析操作,得到短语结构树,其中,所述短语结构树包括从句数据和/或短语数据。

[0012] 在一些实施例,所述对所述原始长句数据进行编辑处理,得到第一候选短句数据的步骤,包括以下至少之一:

[0013] 对所述原始长句数据中与所述短语结构树对应相同的数据进行删除操作,得到第一候选短句数据;或者,

[0014] 在所述短语结构树包括从句数据的情况下,从所述短语结构树中提取所述从句数据,作为所述第一候选短句数据;或者,

[0015] 在所述短语结构树包括短语数据的情况下,从所述短语结构树中提取所述短语数据,作为所述第一候选短句数据。

- [0016] 在一些实施例,所述计算所述第一候选短句数据对应的分数数据的步骤,包括:
- [0017] 对所述第一候选短句数据与所述原始长句数据进行语义相似度计算,得到语义相似度数据;
- [0018] 通过统计语言模型对所述第一候选短句数据进行语句通顺度计算,得到语句通顺度数据;
- [0019] 对所述第一候选短句数据与所述原始长句数据进行距离计算和归一化处理,得到候选短句长度数据;
- [0020] 对所述语义相似度数据、所述语句通顺度数据和所述候选短句长度数据进行乘积计算,得到所述第一候选短句数据对应的分数数据。
- [0021] 在一些实施例,所述对所述第一候选短句数据与所述原始长句数据进行语义相似度计算,得到语义相似度数据的步骤,包括:
- [0022] 计算所述第一候选短句数据对应的第一句向量和所述原始长句数据对应的第二句向量;
- [0023] 对所述第一句向量与所述第二句向量进行余弦相似度计算,得到余弦相似度数据。
- [0024] 在一些实施例,所述对所述第一候选短句数据与所述原始长句数据进行距离计算和归一化处理,得到候选短句长度数据的步骤,包括:
- [0025] 计算所述第一候选短句数据对应的第一句子长度和所述原始长句数据对应的第二句子长度;
- [0026] 对所述第二句子长度与所述第一句子长度进行减法计算,得到距离数据;
- [0027] 对所述距离数据进行归一化处理,得到候选短句长度数据。
- [0028] 在一些实施例,所述预设停止条件为:
- [0029] 当前排序数据中的第二候选短句数据对应的最大分数数据,小于前一排序数据中的第二候选短句数据对应的最大分数数据。
- [0030] 为实现上述目的,本申请实施例的第二方面提出了一种数据压缩装置,所述装置包括:
- [0031] 数据获取模块,用于获取待处理的原始长句数据;
- [0032] 编辑处理模块,用于对所述原始长句数据进行编辑处理,得到第一候选短句数据;
- [0033] 排序处理模块,用于计算所述第一候选短句数据对应的分数数据,并根据所述分数数据对所述第一候选短句数据进行排序处理,得到排序数据;
- [0034] 循环处理模块,用于对所述排序数据进行筛选处理,得到第二候选短句数据,并将所述第二候选短句数据作为所述原始长句数据,重新获取待处理的所述原始长句数据直至满足预设停止条件;
- [0035] 目标筛选模块,用于从满足所述预设停止条件对应的上一所述排序数据中,筛选得到目标短句数据。
- [0036] 为实现上述目的,本申请实施例的第三方面提出了一种电子设备,所述电子设备包括存储器、处理器、存储在所述存储器上并可在所述处理器上运行的程序以及用于实现所述处理器和所述存储器之间的连接通信的数据总线,所述程序被所述处理器执行时实现上述第一方面所述的数据压缩方法。

[0037] 为实现上述目的,本申请实施例的第四方面提出了一种存储介质,所述存储介质为计算机可读存储介质,用于计算机可读存储,所述存储介质存储有一个或者多个程序,所述一个或者多个程序可被一个或者多个处理器执行,以实现上述第一方面所述的数据压缩方法。

[0038] 本申请提出的数据压缩方法、装置、电子设备及存储介质,其通过获取待处理的原始长句数据,之后对原始长句数据进行编辑处理,得到第一候选短句数据,再计算第一候选短句数据对应的分数数据,并根据分数数据对第一候选短句数据进行排序处理,得到排序数据,之后对排序数据进行筛选处理,得到第二候选短句数据,并将第二候选短句数据作为原始长句数据,重新获取待处理的原始长句数据直至满足预设停止条件,从满足预设停止条件对应的上一排序数据中,筛选得到目标短句数据。本申请实施例在不需要监督数据的情况下,通过编辑处理和排序处理等无监督学习的方式逐步对原始长句数据进行压缩简化,相对于需要人工标注并行语料以获得带标签的样本集的相关技术,本申请实施例无需对数据进行标注和监督,能够有效提高数据的压缩效率。

附图说明

[0039] 图1是本申请实施例提供的数据压缩方法的流程示意图;

[0040] 图2是图1中的步骤S102之前的子流程示意图;

[0041] 图3是图1中的步骤S102的子流程示意图;

[0042] 图4是图1中的步骤S103的子流程示意图;

[0043] 图5是图4中的步骤S401的子流程示意图;

[0044] 图6是图4中的步骤S403的子流程示意图;

[0045] 图7是本申请实施例提供的候选短句长度数据的曲线示意图;

[0046] 图8是本申请实施例提供的数据压缩装置的结构示意图;

[0047] 图9是本申请实施例提供的电子设备的硬件结构示意图。

具体实施方式

[0048] 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,此处所描述的具体实施例仅用以解释本申请,并不用于限定本申请。

[0049] 需要说明的是,虽然在装置示意图中进行了功能模块划分,在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于装置中的模块划分,或流程图中的顺序执行所示出或描述的步骤。说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。

[0050] 除非另有定义,本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同。本文中所使用的术语只是为了描述本申请实施例的目的,不是旨在限制本申请。

[0051] 首先,对本申请中涉及的若干名词进行解析:

[0052] 人工智能(Artificial Intelligence, AI):是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学;人工智能是计算机科学的一个

分支,人工智能企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器,该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。人工智能可以对人的意识、思维的信息过程的模拟。人工智能还是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

[0053] 自然语言处理(Natural Language Processing,NLP):NLP用计算机来处理、理解以及运用人类语言(如中文、英文等),NLP属于人工智能的一个分支,是计算机科学与语言学的交叉学科,又常被称为计算语言学。自然语言处理包括语法分析、语义分析、篇章理解等。自然语言处理常用于机器翻译、手写体和印刷体字符识别、语音识别及文语转换、信息意图识别、信息抽取与过滤、文本分类与聚类、舆情分析和观点挖掘等技术领域,它涉及与语言处理相关的数据挖掘、机器学习、知识获取、知识工程、人工智能研究和与语言计算相关的语言学研究等。

[0054] 机器学习(Machine Learning,ML):是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科,它专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域,机器学习(深度学习)通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、式教学习等技术。

[0055] BERT(Bidirectional Encoder Representations from Transformer)模型:BERT模型进一步增加词向量模型泛化能力,充分描述字符级、词级、句子级甚至句间关系特征,其基于Transformer构建而成。BERT中有三种embedding,即token embedding,segment embedding,position embedding;其中token embedding是词向量,第一个单词是CLS标志,可以用于之后的分类任务;segment embedding用来区别两种句子,因为预训练不光做LM还要做以两个句子为输入的分类任务;position embedding,这里的位置词向量不是Transfor中的三角函数,而是BERT经过训练学到的。但BERT直接训练一个position embedding来保留位置信息,每个位置随机初始化一个向量,加入模型训练,最后就得到一个包含位置信息的embedding,最后这个position embedding和word embedding的结合方式上,BERT选择直接拼接。

[0056] 统计语言模型(Statistical Language Model):是自然语言处理的基础模型,是从概率统计角度出发,解决自然语言上下文相关的特性的数学模型。统计语言模型的核心就是判断一个句子在文本中出现的概率。

[0057] Transformer模型:是一种基于多头注意力机制的模型。多头注意力机制为:使用多个注意力机制进行单独计算,以获取更多层面的语义信息,然后将各个注意力机制获取的结果进行拼接组合,以得到最终的结果。

[0058] 嵌入(embedding):embedding是一种向量表征,是指用一个低维的向量表示一个物体,该物体可以是一个词,或是一个商品等等;这个embedding向量的性质是能使距离相近的向量对应的物体有相近的含义,比如embedding(复仇者联盟)和embedding(钢铁侠)之间的距离就会很接近,但embedding(复仇者联盟)和embedding(乱世佳人)的距离就会远一些。embedding实质是一种映射,从语义空间到向量空间的映射,同时尽可能在向量空间保

持原样本在语义空间的关系,如语义接近的两个词汇在向量空间中的位置也比较接近。embedding能够用低维向量对物体进行编码还能保留其含义,常应用于机器学习,在机器学习模型构建过程中,通过把物体编码为一个低维稠密向量再传给DNN,以提高效率。

[0059] 在数据例如文本分类等场景中,通常句子的长度越长,通过分类模型输出的置信度分数会越低,导致分类出错的概率也会越大。而现如今大多数效果突出的自然语言处理模型都是基于Transformer模型的,该模型在实际应用场景中会通过预先设定一个最大文本长度N,以对输入到模型的文本的长度进行严格限制。例如,常用的最大文本长度N有128个字符或256个字符等,当文本超过设定的最大文本长度N都会被截断,即模型只会读取前N个字符,后续的字符会被模型直接丢弃,然而那些被丢弃的字符可能正是表达语义所需要的最重要的部分。所以,在实际应用场景中,当遇到文本长度太长时,通常会通过某种方法将这些文本进行压缩简化,使其变得更加简短,然后用压缩简化过后的文本来替代原始的文本,以便于进行后续的分类任务等。可理解的是,上述文本可以为一个句子,例如定义超过设定的最大文本长度N为长句,将原始的长句进行压缩简化,可得到更加简短的短句,并将得到的短句替代原始的长句以进行后续的任务。

[0060] 受到机器翻译领域的启发和影响,相关技术中,针对长句数据的压缩方法通常是基于监督学习算法和Sequence to Sequence模型的(Sequence to Sequence模型是一种从序列到序列的转换模型框架,可应用在机器翻译、自动应答等场景中)。由于监督学习算法主要依赖于带标签的样本集,故在将带标签的样本集输入监督学习算法之前,还需要人工标注并行语料以获得该样本集。但当样本集数据量大、人工标注时间长时,容易影响数据的压缩效率。此外,通过人工标注出来的并行语料通常是针对特定领域的,由此训练得到的模型的领域适应性较差,如果将通过这些语料训练的模型应用在其他领域时,往往效果欠佳。

[0061] 基于此,本申请实施例提供了一种数据压缩方法、装置、电子设备及存储介质,能够有效提高数据的压缩效率,且通过本申请实施例的数据压缩方法得到的目标短句数据并不局限于特定领域,由此训练得到的模型将具备良好的领域适应性。例如,本申请实施例提供的数据压缩方法可应用于文本处理、文本分类等。

[0062] 本申请实施例提供的数据压缩方法、装置、电子设备及存储介质,具体通过如下实施例进行说明,首先描述本申请实施例中的数据压缩方法。

[0063] 本申请实施例可以基于人工智能技术对相关的数据进行获取和处理。其中,人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

[0064] 人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、机器人技术、生物识别技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0065] 本申请实施例提供的数据压缩方法,涉及人工智能技术领域。本申请实施例提供的数据压缩方法可应用于终端中,也可应用于服务器端中,还可以是运行于终端或服务器端中的软件。在一些实施例中,终端可以是智能手机、平板电脑、笔记本电脑、台式计算机等;服务器端可以配置成独立的物理服务器,也可以配置成多个物理服务器构成的服务器

集群或者分布式系统,还可以配置成提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN以及大数据和人工智能平台等基础云计算服务的云服务器;软件可以是实现数据压缩方法的应用等,但并不局限于以上形式。

[0066] 本申请可用于众多通用或专用的计算机系统环境或配置中。例如:个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器系统、基于微处理器的系统、置顶盒、可编程的消费电子设备、网络PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述,例如程序模块。一般地,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本申请,在这些分布式计算环境中,由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中,程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0067] 图1是本申请实施例提供的的数据压缩方法的一个可选的流程示意图,图1中的方法可以包括但不限于包括步骤S101至步骤S105。

[0068] 步骤S101,获取待处理的原始长句数据;

[0069] 步骤S102,对原始长句数据进行编辑处理,得到第一候选短句数据;

[0070] 步骤S103,计算第一候选短句数据对应的分数数据,并根据分数数据对第一候选短句数据进行排序处理,得到排序数据;

[0071] 步骤S104,对排序数据进行筛选处理,得到第二候选短句数据,并将第二候选短句数据作为原始长句数据;

[0072] 执行循环过程:重新获取待处理的原始长句数据直至满足预设停止条件;

[0073] 步骤S105,从满足预设停止条件对应的上一排序数据中,筛选得到目标短句数据。

[0074] 本申请实施例的步骤S101至步骤S105,通过获取待处理的原始长句数据,之后对原始长句数据进行编辑处理,得到第一候选短句数据,再计算第一候选短句数据对应的分数数据,并根据分数数据对第一候选短句数据进行排序处理,得到排序数据,之后对排序数据进行筛选处理,得到第二候选短句数据,并将第二候选短句数据作为原始长句数据,重新获取待处理的原始长句数据直至满足预设停止条件,从满足预设停止条件对应的上一排序数据中,筛选得到目标短句数据。本申请实施例在不需要监督数据的情况下,通过编辑处理和排序处理等无监督学习的方式逐步对原始长句数据进行压缩简化,相对于需要人工标注并行语料以获得带标签的样本集的相关技术,本申请实施例无需对数据进行标注和监督,能够有效提高数据的压缩效率,且能够有效减少人工成本。

[0075] 可以理解的是,原始长句数据可以为原始的待简化的长句,而目标短句数据即为本申请实施例通过对待简化的长句进行压缩简化后最终得到的短句。

[0076] 本申请实施例通过步骤S101至步骤S104,以对原始长句数据进行编辑处理得到压缩简化的第一候选短句数据,之后对第一候选短句数据进行排序处理,以便于从排序数据中筛选得到第二候选短句数据。例如,本申请实施例的排序处理可设置为分数数据由小到大排序,或者分数数据由大到小排序等。可以理解的是,第二候选短句数据是从进行排序处理后的第一候选短句数据即排序数据中筛选出来的,故可定义将排序数据中的beam_size个第一候选短句数据作为第二候选短句数据,在一些实施例中,beam_size的数值可设置为大于或等于1的整数。

[0077] 为了进一步获取更加简短且能够与原始长句数据所表达的语义基本匹配的目标短句数据,本申请实施例将筛选得到的第二候选短句数据作为原始长句数据,以重新执行步骤S101至步骤S104,即执行循环过程:重新获取待处理的原始长句数据直至满足预设停止条件。通过对原始长句数据不断进行迭代优化,直到满足预设停止条件时,停止执行循环过程,此时,将满足预设停止条件对应的排序数据定义为当前排序数据时,则对当前排序数据对应的上一排序数据进行筛选得到目标短句数据,即执行步骤S105。可以理解的是,由于在一些实施例的步骤S104中,通过对满足预设停止条件对应的上一排序数据进行筛选处理,得到第二候选短句数据,故还可从上一排序数据对应的第二候选短句数据中,筛选得到目标短句数据。

[0078] 可以理解的是,本申请实施例的第一候选短句数据可以为一个或多个,第二候选短句数据可以为一个或多个,而再将第二候选短句数据重新作为原始长句数据执行循环过程时,原始长句数据也可以为一个或多个,本申请实施例对此不作具体限定。

[0079] 请参阅图2,在一些实施例中,在步骤S102之前,包括:

[0080] 步骤S201,对原始长句数据进行短语结构分析操作,得到短语结构树,其中,短语结构树包括从句数据和/或短语数据。

[0081] 可以理解的是,本申请实施例为基于待简化的原始长句数据的短语结构树。在一些实施例的步骤S201中,可通过利用第三方的解析工具,例如Stanford Parser (Stanford NLP) 工具,对原始长句数据进行短语结构分析操作,得到短语结构树,继而可以得到从句数据和/或短语数据,该从句数据和/或短语数据为原始长句数据中的一部分。可以理解的是,从句数据可以为:主语从句、宾语从句、同位语从句等;短语数据可以为:主谓短语、动宾短语、偏正短语等,还可以为名词短语、形容词短语等,本申请实施例对此不作具体限定。

[0082] 需说明的是,短语结构树是用来表达句子的句法结构,其只有叶子结点与输入句子中的词语相关联,其他结点都是标记短语成分。即短语结构树是基于短语结构文法得到的自然语言或形式语言句子结构的树状表示,其中的叶子结点与终极符号对应,其他结点与非终极符号对应。

[0083] 通过短语结构树不断地将原始长句数据中的成分(包括短语和句子)按照规则组成新的短语,从而得到句子的结构。本申请实施例通过短语结构分析操作得到短语结构树,便于进一步对原始长句数据的压缩简化,从而便于进一步保留表达原始长句数据的语义的重要部分。

[0084] 请参阅图3,在一些实施例中,步骤S102可以包括但不限于包括步骤S301至步骤S303:

[0085] 步骤S301,对原始长句数据中与短语结构树对应相同的数据进行删除操作,得到第一候选短句数据;或者,

[0086] 步骤S302,在短语结构树包括从句数据的情况下,从短语结构树中提取从句数据,作为第一候选短句数据;或者,

[0087] 步骤S303,在短语结构树包括短语数据的情况下,从短语结构树中提取短语数据,作为第一候选短句数据。

[0088] 本申请实施例可以采用不同的编辑方法对原始长句数据进行编辑处理,例如删除原始长句数据中的某些词,或者交换原始长句数据中的两个短语数据等。采用不同的编辑

方法作用于原始长句数据中的不同“部位”后,可以获得若干个长度更短的第一候选短句数据。

[0089] 具体地,在一些实施例的步骤S301中,在短语结构树包括从句数据,且原始长句数据中具有与从句数据对应相同的数据的情况下,将该相同的数据(即从句数据)从原始长句数据中进行删除,并将原始长句数据中除该相同的数据(即从句数据)的其余部分作为第一候选短句数据,即将进行删除操作后的原始长句数据作为第一候选短句数据。又或者,在短语结构树包括短语数据,且原始长句数据中具有与短语数据对应相同的数据的情况下,将该相同的数据(即短语数据)从原始长句数据中进行删除,并将原始长句数据中除该相同的数据(即短语数据)的其余部分作为第一候选短句数据。

[0090] 在一些实施例的步骤S302中,当短语结构树包括从句数据,从短语结构树中提取从句数据,并将该从句数据直接作为第一候选短句数据。

[0091] 在一些实施例的步骤S303中,当短语结构树包括短语数据,从短语结构树中提取短语数据,并将该短语数据直接作为第一候选短句数据。

[0092] 可以理解的是,通过对原始长句数据进行编辑处理,例如对原始长句数据中对应于短语结构树中的每一个从句数据或短语数据,均进行若干次上述的删除操作(即步骤S301)或提取操作(即步骤S302或步骤S303),从而得到若干个比原始长句数据更加简短的句子,即得到第一候选短句数据。

[0093] 需要说明的是,由于本申请实施例可以为循环处理的过程,故上述的原始长句数据不一定是原始的待简化的长句,也可以是上一排序数据中,进行筛选处理后得到的第二候选短句数据,本申请对此不作具体限定。

[0094] 请参阅图4,在一些实施例中,步骤S103中的计算第一候选短句数据对应的分数数据的步骤,可以包括但不限于包括步骤S401至步骤S404:

[0095] 步骤S401,对第一候选短句数据与原始长句数据进行语义相似度计算,得到语义相似度数据;

[0096] 步骤S402,通过统计语言模型对第一候选短句数据进行语句通顺度计算,得到语句通顺度数据;

[0097] 步骤S403,对第一候选短句数据与原始长句数据进行距离计算和归一化处理,得到候选短句长度数据;

[0098] 步骤S404,对语义相似度数据、语句通顺度数据和候选短句长度数据进行乘积计算,得到第一候选短句数据对应的分数数据。

[0099] 为了便于获取与原始长句数据的语义相近且简短的数据,本申请实施例需要计算第一候选短句数据对应的分数数据。通过对每一第一候选短句数据进行逐个打分,以计算得到对应的分数数据。

[0100] 定义第一候选短句数据对应的分数数据的函数表示为打分函数 $f(C)$,其中, C 表示第一候选短句数据。可理解的是,打分函数 $f(C)$ 包括有语义相似度数据、语句通顺度数据和候选短句长度数据,具体表示为语义相似度数据、语句通顺度数据和候选短句长度数据之间的乘积。

[0101] 可以理解的是,语义相似度数据、语句通顺度数据和候选短句长度数据能够分别对第一候选短句数据的语义相似度、语句通顺度和短句长度等方面的特性进行打分。

[0102] 定义语义相似度数据的函数表示为 $f_{sim}(X, C)$, 其中, X 表示原始长句数据。对于每一第一候选短句数据, 均需要计算第一候选短句数据与原始长句数据之间的语义相似度数据。可以理解的是, 语义相似度数据可以从不同的层次进行度量, 例如从词、短语到段落、文档等, 通过对第一候选短句数据与原始长句数据进行语义相似度计算, 使得语义相似度数据能够捕捉到原始长句数据的本质结构, 以便于进一步促进后续的学习任务, 如分类任务等。

[0103] 定义语句通顺度数据的函数表示为 $f_{smooth}(C)$, 可以理解的是, 该函数表示主要用于评估经过编辑处理得到的第一候选短句数据对应的语句流畅性和通顺度。在一些实施例的步骤S402中, 可利用统计语言模型来对第一候选短句数据的语句通顺度进行计算, 以得到语句通顺度数据。在一些实施例中, 将第一候选短句数据 C 输入至统计语言模型, 得到语句通顺度数据, 该语句通顺度数据表征第一候选短句数据 C 在原始语料中出现的概率 $P(C)$ 。

[0104] 定义候选短句长度数据的函数表示为归一化编辑距离函数 $f_{edit}(X, C)$, 本申请实施例具体是对第一候选短句数据 C 的短句长度进行打分。可以理解的是, 第一候选短句数据 C 是根据对原始长句数据 X 进行编辑处理得到的, 即第一候选短句数据 C 可通过对原始长句数据 X 进行迭代的删除操作或提取操作得到, 故第一候选短句数据 C 可理解为由原始长句数据 X 中的若干从句数据或短语数据拼接而成的, 故通过对第一候选短句数据与原始长句数据进行距离计算和归一化处理, 可得到候选短句长度数据。

[0105] 因此, 在一些实施例的步骤S404中, 打分函数 $f(C)$ 可具体表示为 $f(C) = f_{sim}(X, C) * f_{smooth}(C) * f_{edit}(X, C)$, 即通过对语义相似度数据、语句通顺度数据和候选短句长度数据进行乘积计算, 得到第一候选短句数据对应的分数数据, 之后, 根据分数数据对第一候选短句数据进行排序处理, 得到排序数据。

[0106] 请参阅图5, 在一些实施例中, 步骤S401可以包括但不限于包括步骤S501至步骤S502:

[0107] 步骤S501, 计算第一候选短句数据对应的第一句向量和原始长句数据对应的第二句向量;

[0108] 步骤S502, 对第一句向量与第二句向量进行余弦相似度计算, 得到余弦相似度数据。

[0109] 可以理解的是, 从语义角度而言, 若第一候选短句数据的语义与原始长句数据的语义越接近, 则余弦相似度数据也将越高。故本申请实施例通过对第一候选短句数据与原始长句数据进行余弦相似度计算, 以得到余弦相似度数据。具体地, 分别计算第一候选短句数据对应的第一句向量和原始长句数据对应的第二句向量。其中, 句向量(sentence embedding)可以通过多种方式计算得到, 例如, 可以通过第三方发布的、经过大规模语料预训练得到的BERT模型来获取句向量。在一些实施例中, 通过将第一候选短句数据和原始长句数据分别输入到上述BERT模型中, 可得到第一候选短句数据对应的第一句向量和原始长句数据对应的第二句向量。可以理解的是, 为了得到较精确的句向量, 还可以通过实际应用场景中的训练数据来对预训练的BERT模型再次进行微调, 以得到性能更好的BERT模型。

[0110] 之后, 通过对第一句向量与第二句向量进行余弦相似度计算, 以得到余弦相似度数据。可以理解的是, 余弦相似度数据具体可表示为: 通过获取第一句向量与第二句向量之间的夹角的大小, 来判断第一句向量与第二句向量之间的相似程度。其中, 夹角越小, 则代

表两个句向量之间越相似,即对应的第一候选短句数据的语义与原始长句数据的语义也越接近。例如,在一些实施例中,余弦相似度数据越接近1,则表示夹角越接近 0° ,即代表第一句向量与第二句向量之间越相似。

[0111] 请参阅图6,在一些实施例中,步骤S403可以包括但不限于包括步骤S601至步骤S603:

[0112] 步骤S601,计算第一候选短句数据对应的第一句子长度和原始长句数据对应的第二句子长度;

[0113] 步骤S602,对第二句子长度与第一句子长度进行减法计算,得到距离数据;

[0114] 步骤S603,对距离数据进行归一化处理,得到候选短句长度数据。

[0115] 可以理解的是,第一候选短句数据C可理解为由原始长句数据X中的若干从句数据或短语数据拼接而成的,故可分别计算第一候选短句数据对应的第一句子长度和原始长句数据对应的第二句子长度。

[0116] 定义第一句子长度表示为 $|C|$,第二句子长度表示为 $|X|$,距离数据表示为 $w(X,C)$,候选短句长度数据的函数表示为归一化编辑距离函数 $f_{edit}(X,C)$ 。其中, $|\cdot|$ 表示计算句子长度,距离数据为大于或者等于0的整数。

[0117] 在一些实施例的步骤S602中,通过对第二句子长度 $|X|$ 与第一句子长度 $|C|$ 进行减法计算,得到距离数据 $w(X,C)$,故 $w(X,C) = |X| - |C|$ 。

[0118] 可以理解的是,为了和语义相似度数据、语句通顺度数据处于同一个数值范围,本申请实施例需要对距离数据进行归一化处理,以得到候选短句长度数据。

[0119] 具体地,定义本申请实施例的候选短句长度数据为:距离数据除以第一句子长度和第二句子长度之间的和,故在一些实施例的步骤S603中,候选短句长度数据可表示为:

$$f_{edit}(X,C) = \frac{|X| - |C|}{|X| + |C|} = \frac{2|X|}{|X| + |C|} - 1。$$

[0120] 可以理解的是,对于待简化的原始长句数据X而言,其对应的第二句子长度 $|X|$ 通常是一个确定的常数,故候选短句长度数据 $f_{edit}(X,C)$ 也只与第一候选短句数据对应的第一句子长度 $|C|$ 有关。本申请实施例可将 $\frac{2|X|}{|X| + |C|} - 1$ 看作第一句子长度 $|C|$ 的倒数,即 $\frac{1}{|C|}$

的变体。具体参照图7,图7中的曲线1表示本申请实施例的候选短句长度数据 $f_{edit}(X,C)$,曲线2表示 $\frac{1}{|C|}$ 。根据图7可知,曲线1、曲线2均随着第一句子长度 $|C|$ 的增加而递减,其中候选短句长度数据 $f_{edit}(X,C)$ 的递减过程呈现出明显的阶梯状,然而曲线2,即 $\frac{1}{|C|}$ 随着第一句子长度 $|C|$ 的变化,从一开始就呈现出断崖式递减,随后基本不再有起伏变化。故相比于曲线2的 $\frac{1}{|C|}$,本申请实施例的候选短句长度数据 $f_{edit}(X,C)$ 能够更加准确地对第一候选短句数据C的短句长度进行打分。即本申请实施例通过对距离数据进行归一化处理得到的候选短句长度数据将更加精确,能够有效反映第一候选短句数据的短句长度。

[0121] 在一些实施例中,预设停止条件为:当前排序数据中的第二候选短句数据对应的

最大分数数据,小于前一排序数据中的第二候选短句数据对应的最大分数数据。

[0122] 可以理解的是,本申请实施例在满足预设停止条件后,将退出循环过程,并从满足预设停止条件对应的上一排序数据中,筛选得到目标短句数据,否则重新执行步骤S101至步骤S104,以重复迭代编辑处理、排序处理等过程。

[0123] 例如,当完成一轮迭代/循环之后,首先判断当前得到的第二候选短句数据是否满足预设停止条件,若满足预设停止条件,即当前排序数据中的第二候选短句数据对应的最大分数数据,小于前一排序数据中的第二候选短句数据对应的最大分数数据,则对满足预设停止条件对应的上一排序数据进行筛选,得到目标短句数据。在一些实施例中,获取满足预设停止条件对应的上一排序数据中最大分数数据对应的第二候选短句数据,并将该最大分数数据对应的第二候选短句数据作为目标短句数据,则该目标短句数据即为原始长句数据经压缩简化后的最终结果。

[0124] 若不满足预设停止条件,则将第二候选短句数据作为原始长句数据,重新执行步骤S101至步骤S104。可以理解的是,在对当前排序数据进行筛选处理时,且在在不满足预设停止条件的情况下,即在当前排序数据中的第二候选短句数据对应的最大分数数据,大于或者等于前一排序数据中第二候选短句数据对应的最大分数数据的情况下,可以选择从当前排序数据中筛选出beam_size个第二候选短句数据,并将该beam_size个第二候选短句数据作为原始长句数据,重复执行循环过程。

[0125] 需说明的是,由于排序处理可设置为分数数据由小到大排序,或者分数数据由大到小排序,故在对当前排序数据进行筛选处理时,可设置条件为:将当前排序数据中,前beam_size个最大分数数据对应的第一候选短句数据作为第二候选短句数据,本申请实施例对此不作具体限定。

[0126] 在一些实施例中,为了节约成本开销,可将beam_size的数值设置为1。但当beam_size的数值设置为1,极有可能使得迭代结果陷入局部最优解,故为了避免贪婪的搜索局部最优解,本申请实施例可将beam_size的数值设置为大于1的整数。

[0127] 可以理解的是,通过每轮迭代/循环得到的第二候选短句数据不一定是单个原始长句数据经编辑处理、排序处理后的结果,也可能是多个原始长句数据经编辑处理、排序处理后的集合。

[0128] 可以理解的是,原始长句数据对应的初始的分数数据为0,通过判断预设停止条件,来选择是否执行循环过程。即当前排序数据中的第二候选短句数据对应的最大分数数据,小于前一排序数据中的第二候选短句数据对应的最大分数数据,则表示满足预设停止条件,此时停止执行循环过程,并将前一排序数据中最大分数数据对应的第二候选短句数据作为目标短句数据。即从满足预设停止条件对应的上一排序数据中,筛选出最大分数数据对应的第二候选短句数据,并将其作为目标短句数据。

[0129] 当前排序数据中的第二候选短句数据对应的最大分数数据,大于或者等于前一排序数据中的第二候选短句数据对应的最大分数数据,则表示不满足预设停止条件,需要执行循环过程直至满足预设停止条件。

[0130] 本申请实施例通过无监督的、逐步迭代的方式对文本长度过长的原始长句数据进行压缩简化,通过不断迭代简化原始长句数据,以便获取更加简短的目标短句数据,能够有效保证压缩简化过后的目标短句数据在保留了原始长句数据的重要/关键信息的同时,还

使得目标短句数据具有较高的语言流畅性,进而使得后续的文本分类、语义相似度匹配等任务具有更高的准确率。此外,由于本申请实施例的数据压缩方法不需要监督数据,相对于通过人工标注出来的并行语料通常是针对特定领域的,由此训练得到的模型的领域适应性较差的相关技术,本申请实施例可以广泛应用在各个领域中,具备良好的领域适应性。

[0131] 具体地,本申请实施例通过获取待处理的原始长句数据,之后对原始长句数据进行编辑处理,得到第一候选短句数据,再计算第一候选短句数据对应的分数数据,并根据分数数据对第一候选短句数据进行排序处理,得到排序数据,之后对排序数据进行筛选处理,得到第二候选短句数据,并将第二候选短句数据作为原始长句数据,重新获取待处理的原始长句数据直至满足预设停止条件,从满足预设停止条件对应的上一排序数据中,筛选得到目标短句数据。本申请实施例能够在不需要监督数据的情况下,通过编辑处理、分数数据计算和排序处理等无监督学习的方式逐步对原始长句数据进行压缩简化,相对于需要人工标注并行语料以获得带标签的样本集的相关技术,本申请实施例能够有效提高数据的压缩效率。此外,本申请实施例中,通过压缩简化后得到的目标短句数据与原始长句数据之间仍具有较高的语义相似性,从而能够有效保留原始长句数据的重要/关键信息。同时,通过语句通顺度计算,可以进一步保证目标短句数据具有较高的语言流畅度。还可以理解的是,本申请实施例的目标短句数据可以替代原始长句数据进行后续的文本分类或语义相似度匹配等任务,进而可以提高这些任务的准确率;以及本申请实施例的目标短句数据也可以解决某些模型限制输入最大文本长度N的问题。

[0132] 请参阅图8,本申请实施例还提供一种数据压缩装置,可以实现上述数据压缩方法,该装置包括:

[0133] 数据获取模块701,用于获取待处理的原始长句数据;

[0134] 编辑处理模块702,用于对原始长句数据进行编辑处理,得到第一候选短句数据;

[0135] 排序处理模块703,用于计算第一候选短句数据对应的分数数据,并根据分数数据对第一候选短句数据进行排序处理,得到排序数据;

[0136] 循环处理模块704,用于对排序数据进行筛选处理,得到第二候选短句数据,并将第二候选短句数据作为原始长句数据,重新获取待处理的原始长句数据直至满足预设停止条件;

[0137] 目标筛选模块705,用于从满足预设停止条件对应的上一排序数据中,筛选得到目标短句数据。

[0138] 本申请实施例还提供了一种电子设备,电子设备包括:存储器、处理器、存储在存储器上并可在处理器上运行的程序以及用于实现处理器和存储器之间的连接通信的数据总线,程序被处理器执行时实现上述数据压缩方法。该电子设备可以为包括平板电脑、车载电脑等任意智能终端。

[0139] 请参阅图9,图9示意了另一实施例的电子设备的硬件结构,电子设备包括:

[0140] 处理器801,可以采用通用的CPU(Central Processing Unit,中央处理器)、微处理器、应用专用集成电路(Application Specific Integrated Circuit,ASIC)、或者一个或多个集成电路等方式实现,用于执行相关程序,以实现本申请实施例所提供的技术方案;

[0141] 存储器802,可以采用只读存储器(Read Only Memory,ROM)、静态存储设备、动态存储设备或者随机存取存储器(Random Access Memory,RAM)等形式实现。存储器802可以

存储操作系统和其他应用程序,在通过软件或者固件来实现本说明书实施例所提供的技术方案时,相关的程序代码保存在存储器802中,并由处理器801来调用执行本申请实施例的数据压缩方法;

[0142] 输入/输出接口803,用于实现信息输入及输出;

[0143] 通信接口804,用于实现本设备与其他设备的通信交互,可以通过有线方式(例如USB、网线等)实现通信,也可以通过无线方式(例如移动网络、WIFI、蓝牙等)实现通信;

[0144] 总线805,在设备的各个组件(例如处理器801、存储器802、输入/输出接口803和通信接口804)之间传输信息;

[0145] 其中处理器801、存储器802、输入/输出接口803和通信接口804通过总线805实现彼此之间在设备内部的通信连接。

[0146] 本申请实施例还提供了一种存储介质,存储介质为计算机可读存储介质,用于计算机可读存储,存储介质存储有一个或者多个程序,一个或者多个程序可被一个或者多个处理器执行,以实现上述数据压缩方法。

[0147] 存储器作为一种非暂态计算机可读存储介质,可用于存储非暂态软件程序以及非暂态性计算机可执行程序。此外,存储器可以包括高速随机存取存储器,还可以包括非暂态存储器,例如至少一个磁盘存储器件、闪存器件、或其他非暂态固态存储器件。在一些实施方式中,存储器可选包括相对于处理器远程设置的存储器,这些远程存储器可以通过网络连接至该处理器。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0148] 本申请实施例的数据压缩方法、数据压缩装置、电子设备及存储介质,其通过获取待处理的原始长句数据,之后对原始长句数据进行编辑处理,得到第一候选短句数据,再计算第一候选短句数据对应的分数数据,并根据分数数据对第一候选短句数据进行排序处理,得到排序数据,之后对排序数据进行筛选处理,得到第二候选短句数据,并将第二候选短句数据作为原始长句数据,重新获取待处理的原始长句数据直至满足预设停止条件,从满足预设停止条件对应的上一排序数据中,筛选得到目标短句数据。本申请实施例在不需要监督数据的情况下,通过编辑处理和排序处理等无监督学习的方式逐步对原始长句数据进行压缩简化,相对于需要人工标注并行语料以获得带标签的样本集的相关技术,本申请实施例无需对数据进行标注和监督,能够有效提高数据的压缩效率。

[0149] 本申请实施例描述的实施例是为了更加清楚的说明本申请实施例的技术方案,并不构成对于本申请实施例提供的技术方案的限定,本领域技术人员可知,随着技术的演变和新应用场景的出现,本申请实施例提供的技术方案对于类似的技术问题,同样适用。

[0150] 本领域技术人员可以理解的是,图1-6中示出的技术方案并不构成对本申请实施例的限定,可以包括比图示更多或更少的步骤,或者组合某些步骤,或者不同的步骤。

[0151] 以上所描述的装置实施例仅仅是示意性的,其中作为分离部件说明的单元可以是或者也可以不是物理上分开的,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0152] 本领域普通技术人员可以理解,上文中所公开方法中的全部或某些步骤、系统、设备中的功能模块/单元可以被实施为软件、固件、硬件及其适当的组合。

[0153] 本申请的说明书及上述附图中的术语“第一”、“第二”、“第三”、“第四”等(如果存

在)是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0154] 应当理解,在本申请中,“至少一个(项)”是指一个或者多个,“多个”是指两个或两个以上。“和/或”,用于描述关联对象的关联关系,表示可以存在三种关系,例如,“A和/或B”可以表示:只存在A,只存在B以及同时存在A和B三种情况,其中A,B可以是单数或者复数。字符“/”一般表示前后关联对象是一种“或”的关系。“以下至少一项(个)”或其类似表达,是指这些项中的任意组合,包括单项(个)或复数项(个)的任意组合。例如,a,b或c中的至少一项(个),可以表示:a,b,c,“a和b”,“a和c”,“b和c”,或“a和b和c”,其中a,b,c可以是单个,也可以是多个。

[0155] 在本申请所提供的几个实施例中,应该理解到,所揭露的装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,上述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0156] 上述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0157] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0158] 集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括多指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备)执行本申请各个实施例的方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read Only Memory,简称ROM)、随机存取存储器(Random Access Memory,简称RAM)、磁碟或者光盘等各种可以存储程序的介质。

[0159] 以上参照附图说明了本申请实施例的优选实施例,并非因此局限本申请实施例的权利范围。本领域技术人员不脱离本申请实施例的范围和实质内所作的任何修改、等同替换和改进,均应在本申请实施例的权利范围之内。

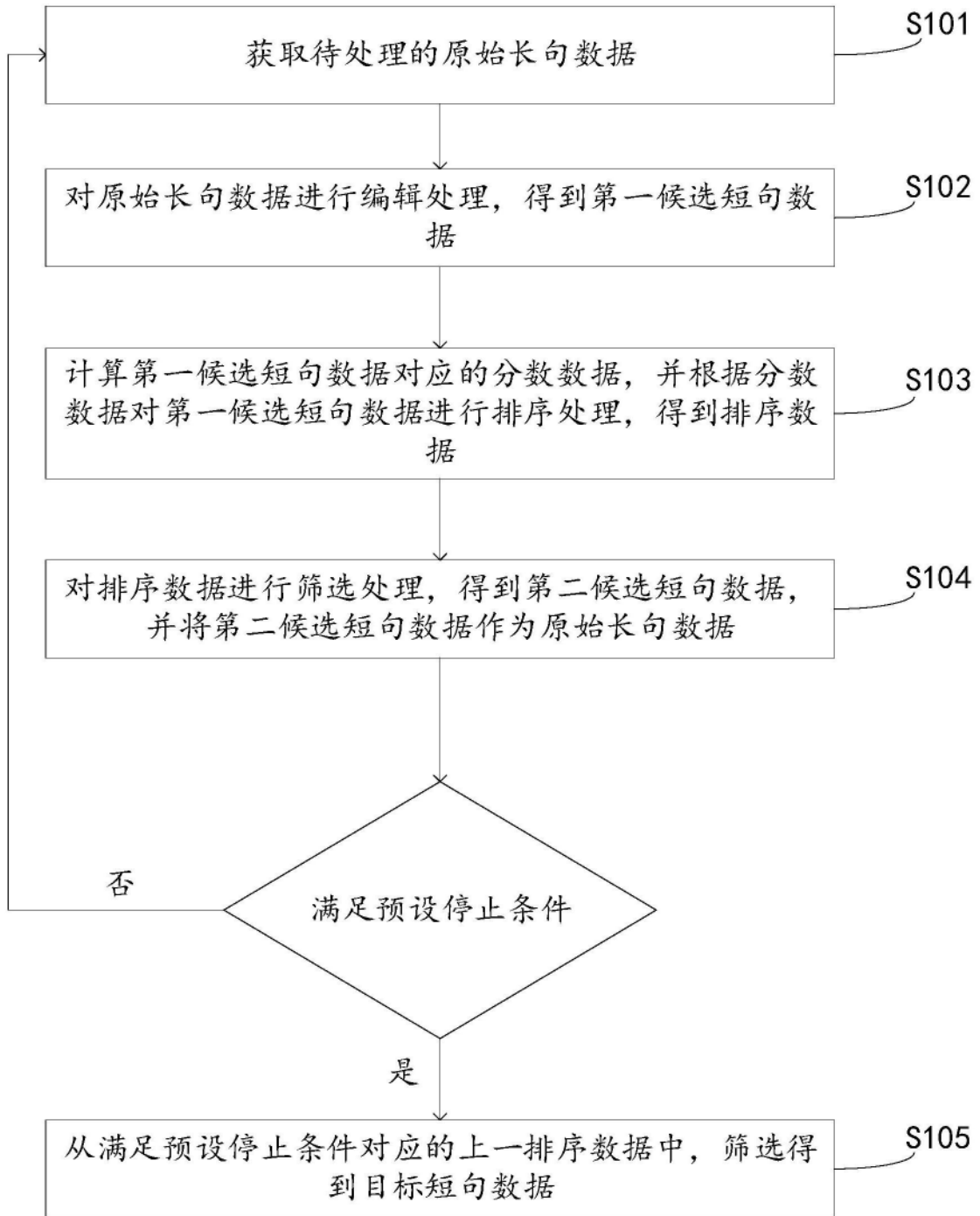


图1

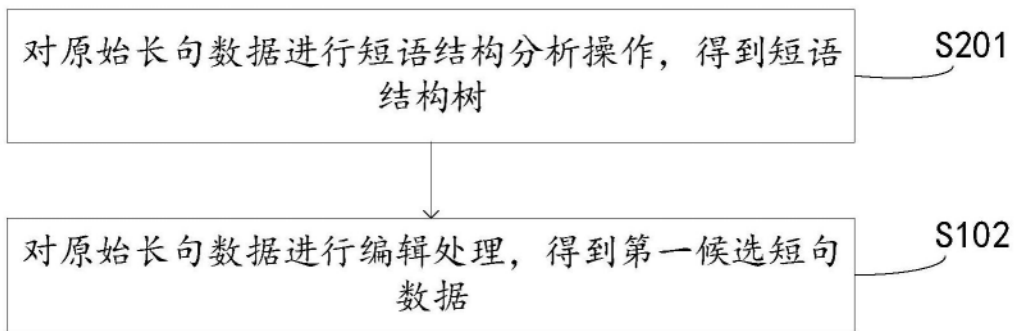


图2

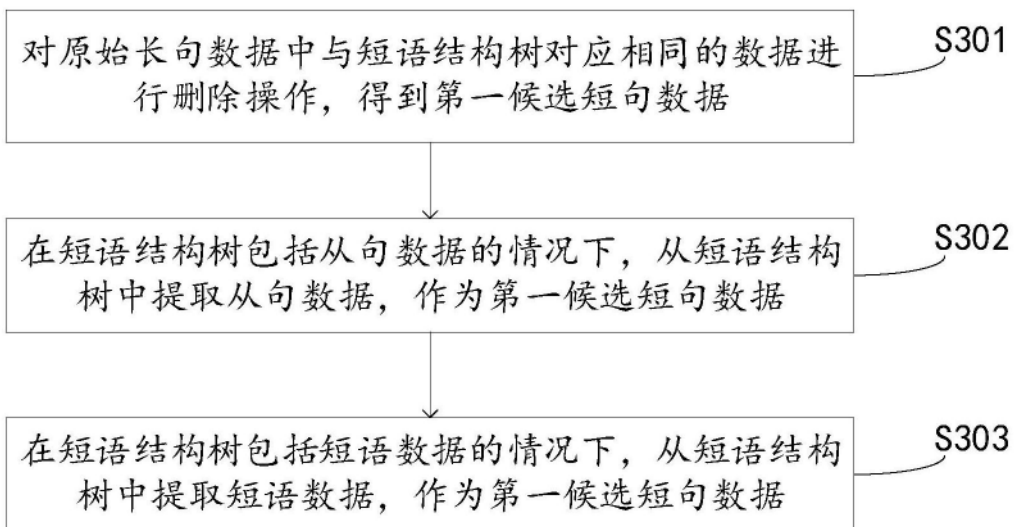


图3

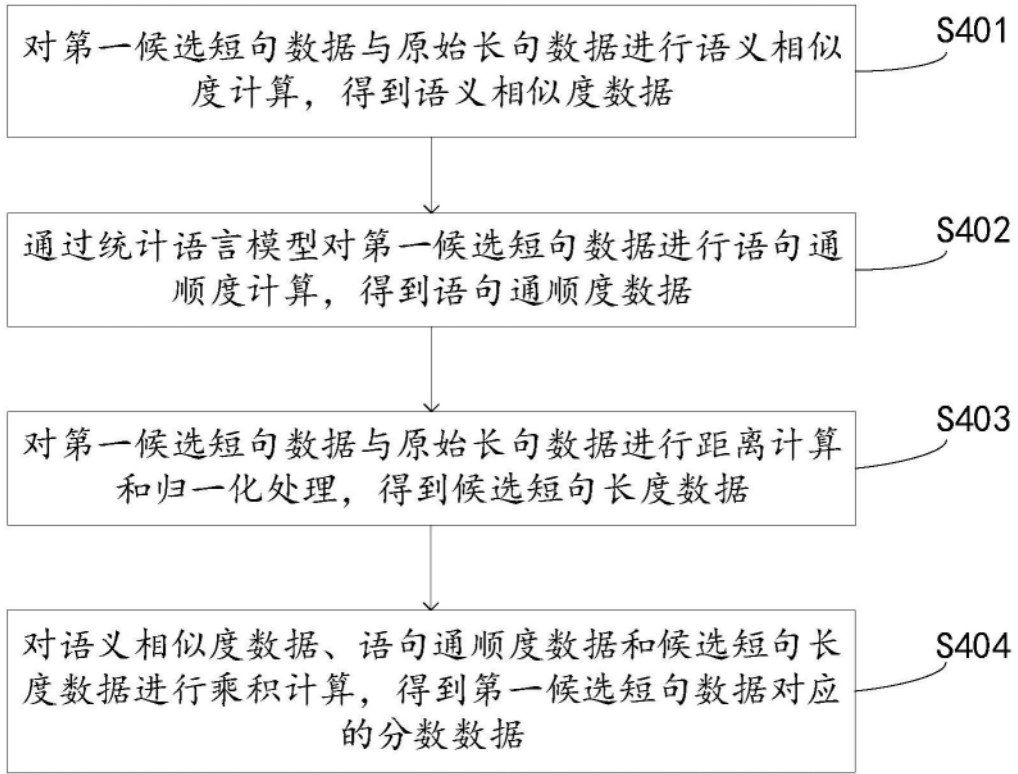


图4

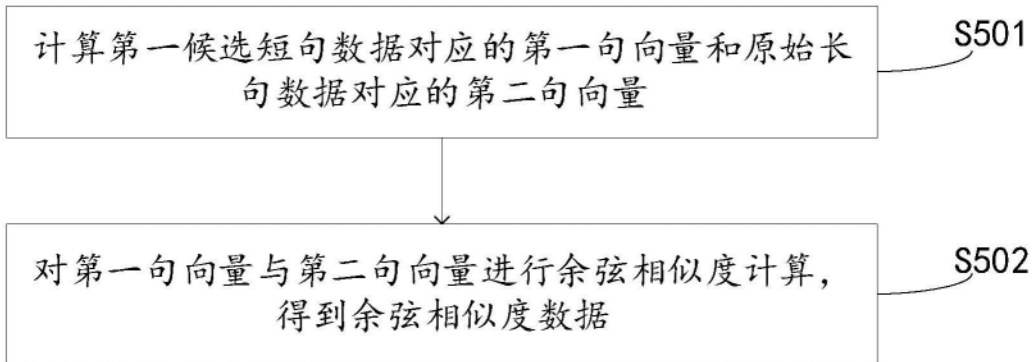


图5

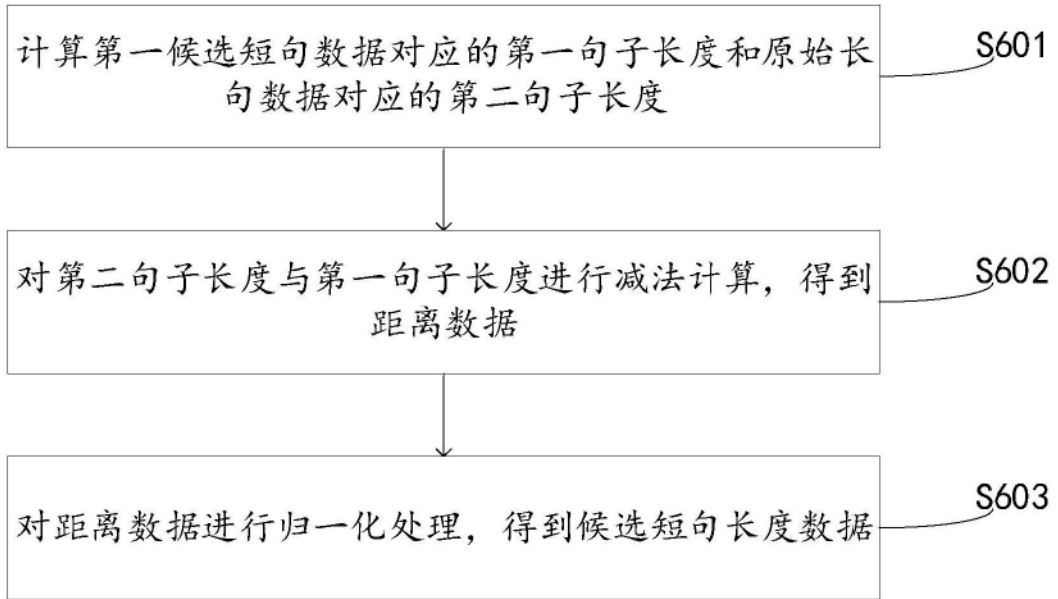


图6

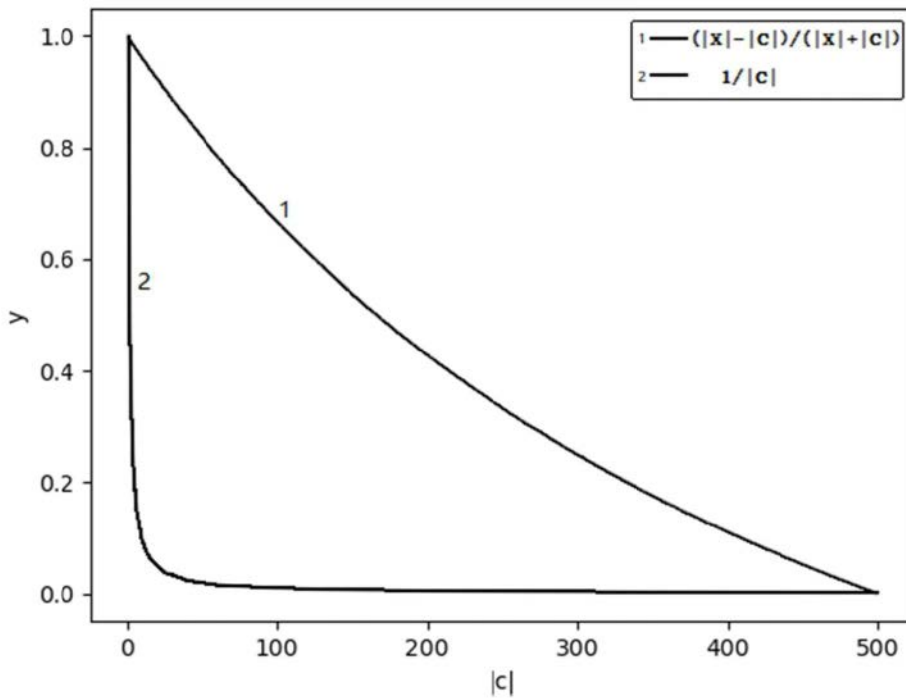


图7

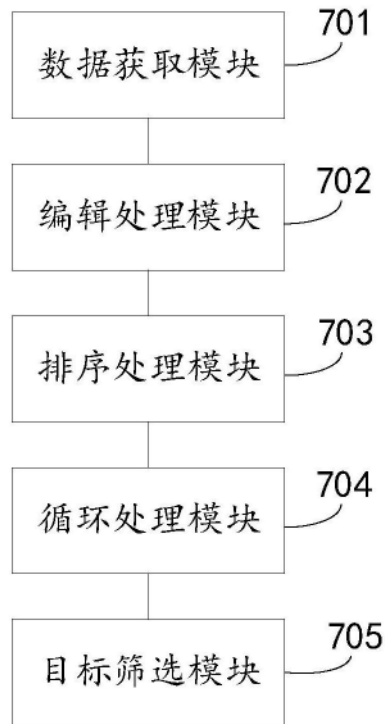


图8

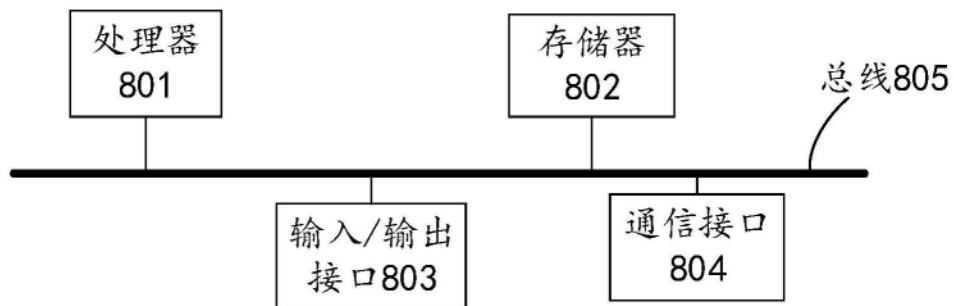


图9