



(12) 发明专利

(10) 授权公告号 CN 112380861 B

(45) 授权公告日 2024. 06. 21

(21) 申请号 202011268695.3

(22) 申请日 2020.11.13

(65) 同一申请的已公布的文献号

申请公布号 CN 112380861 A

(43) 申请公布日 2021.02.19

(73) 专利权人 北京京东尚科信息技术有限公司

地址 100083 北京市海淀区知春路76号8层

专利权人 北京京东世纪贸易有限公司

(72) 发明人 乐雨泉 赵宇明 陈蒙 何从庆

(74) 专利代理机构 北京德琦知识产权代理有限公司

公司 11018

专利代理师 孙清然 王琦

(51) Int. Cl.

G06N 3/084 (2023.01)

G06F 40/284 (2020.01)

G06F 40/30 (2020.01)

G06F 16/35 (2019.01)

G06N 3/045 (2023.01)

G06N 3/0442 (2023.01)

G06N 3/048 (2023.01)

(56) 对比文件

Congying Xia.Zero-shot User Intent Detection via Capsule Neural Networks.arXiv:1809.00385v1 [cs.CL] 2 Sep 2018.2018,1-10.

刘娇;李艳玲;林民.胶囊网络用于短文本多意图识别的研究.计算机科学与探索.2020,14 (第10期),1735-1743.

审查员 张文

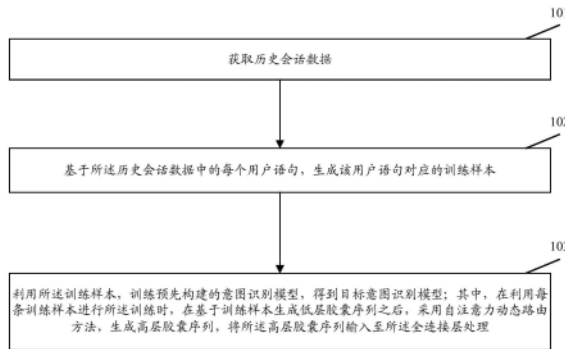
权利要求书3页 说明书11页 附图2页

(54) 发明名称

模型训练方法、装置及意图识别方法、装置

(57) 摘要

本申请公开了一种模型训练方法、装置及意图识别方法、装置,其中,模型训练方法,包括:获取历史会话数据;基于所述历史会话数据中的每个用户语句,生成该用户语句对应的训练样本;利用所述训练样本,训练预先构建的意图识别模型,得到目标意图识别模型;其中,在利用每条训练样本进行所述训练时,在基于训练样本生成低层胶囊序列之后,采用自注意力动态路由方法,生成高层胶囊序列,将所述高层胶囊序列输入至全连接层处理。采用本申请,可以提高意图识别准确度。



1. 一种模型训练方法,其特征在于,包括:
 - 获取历史会话数据;
 - 基于所述历史会话数据中的每个用户语句,生成该用户语句对应的训练样本;
 - 利用所述训练样本,训练预先构建的意图识别模型,得到目标意图识别模型;其中,在利用每条训练样本进行所述训练时,在基于训练样本生成低层胶囊序列之后,采用自注意力动态路由方法,生成高层胶囊序列,将所述高层胶囊序列输入至全连接层处理;
 - 其中,所述采用自注意力动态路由方法,生成高层胶囊序列包括:
 - 利用所述低层胶囊序列,计算注意力权重;
 - 利用所述注意力权重计算所述高层胶囊序列中的每个元素。
2. 根据权利要求1所述的方法,其特征在于,所述生成该用户语句对应的训练样本包括:
 - 为所述用户语句,生成训练文本;
 - 基于所述训练文本,生成词序列;
 - 根据预设的意图类别集合,确定与所述用户语句相匹配的意图类别;
 - 基于所述词序列和所述意图类别,生成所述用户语句对应的训练样本。
3. 根据权利要求2所述的方法,其特征在于,所述生成训练文本包括:
 - 如果所述用户语句为所述历史会话数据中的第一个用户语句,则将所述用户语句作为训练文本;否则,将所述用户语句与在所述用户语句之前的预设数量用户语句组合,得到所述训练文本。
4. 根据权利要求1所述的方法,其特征在于,利用每条所述训练样本进行所述训练包括:
 - 将所述训练样本的词序列,转化为词向量序列;
 - 将所述词向量序列,输入双向长短时记忆网络,并根据输出结果,生成低层胶囊序列;
 - 基于所述低层胶囊序列,采用自注意力动态路由方法,生成高层胶囊序列;
 - 将所述高层胶囊序列执行拉平操作后输入至全连接层,并生成输出结果的概率分布值;
 - 基于所述概率分布值,生成模型损失函数值;
 - 根据所述模型损失函数值和相应训练样本的意图类别标签,调整所述意图识别模型的网络参数。
5. 根据权利要求4所述的方法,其特征在于,所述根据输出结果,生成低层胶囊序列包括:
 - 按照 $u_i = [\bar{u}_i, \tilde{u}_i]$,对双向长短时记忆网络输出的每组前向隐藏状态值和后向隐藏状态值进行拼接,得到所述低层胶囊序列中的每个元素 u_i ;
 - 其中, $\bar{u}_i$ 为所述词向量序列中的第i个词向量对应的前向隐藏状态值, $\tilde{u}_i$ 为所述词向量序列中的第i个词向量对应的后向隐藏状态值; $u = \{u_1, \dots, u_i \dots u_n\}$; u 为所述低层胶囊序列; n 为所述词向量序列中的向量数量; $1 \leq i \leq n$ 。
6. 根据权利要求5所述的方法,其特征在于,所述利用所述低层胶囊序列,计算注意力权重;利用所述注意力权重计算所述高层胶囊序列中的每个元素,包括:

利用所述低层胶囊序列,按照
$$a_{ij} = \frac{e^{\frac{(u_i W^Q)(u_j W^K)^T}{\sqrt{d_z}}}}{\sum_{k=1}^n e^{\frac{(u_i W^Q)(u_k W^K)^T}{\sqrt{d_z}}}}$$
, 计算注意力权重 $a_{i,j}$; 其中, $i \in \{1, 2, \dots, n\}$, $j \in \{1, 2, \dots, m\}$; m 为预设的高层胶囊序列的元素数量;

W^Q 为当前模型中的第一权重矩阵; W^K 为当前模型中的第二权重矩阵; d_z 是预设的缩放因子;

利用所述 $a_{i,j}$, 按照
$$v_j = \frac{\left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|^2}{1 + \left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|^2} \cdot \frac{\sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right)}{\left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|}$$
, 计算所述高

层胶囊序列中的每个元素 v_j ;

其中, $v = \{v_1, \dots, v_j, \dots, v_m\}$; v 为所述高层胶囊序列; $j \in \{1, 2, \dots, m\}$; W^V 为当前模型中的第三权重矩阵, $w_{i,j}$ 为当前模型中的耦合系数。

7. 一种意图识别方法, 其特征在于, 包括:

获取目标文本;

基于所述目标文本, 生成词序列;

基于所述词序列, 利用预先训练的意图识别模型, 确定所述目标文本的意图类别;

其中, 所述意图识别模型采用权利要求1至6所述的任一方法训练得到。

8. 一种模型训练装置, 其特征在于, 包括:

数据获取单元, 用于获取历史会话数据;

训练样本生成单元, 用于基于所述历史会话数据中的每个用户语句, 生成该用户语句对应的训练样本;

训练单元, 用于利用所述训练样本, 训练预先构建的意图识别模型, 得到目标意图识别模型; 其中, 在利用每条训练样本进行所述训练时, 在基于训练样本生成低层胶囊序列之后, 采用自注意力动态路由方法, 生成高层胶囊序列, 将所述高层胶囊序列输入至全连接层处理;

其中, 所述训练单元, 具体用于采用自注意力动态路由方法, 生成高层胶囊序列时, 利用所述低层胶囊序列, 计算注意力权重; 利用所述注意力权重计算所述高层胶囊序列中的每个元素。

9. 一种意图识别装置, 其特征在于, 包括:

文本获取单元, 用于获取目标文本;

文本处理单元, 用于基于所述目标文本, 生成词序列;

意图识别单元, 用于基于所述词序列, 利用预先训练的意图识别模型, 确定所述目标文本的意图类别;

其中, 所述意图识别模型采用权利要求1至6所述的任一方法训练得到。

10. 一种模型训练设备, 其特征在于, 包括处理器和存储器;

所述存储器中存储有可被所述处理器执行的应用程序,用于使得所述处理器执行如权利要求1至6中任一项所述的模型训练方法。

11.一种计算机可读存储介质,其特征在于,其中存储有计算机可读指令,该计算机可读指令用于执行如权利要求1至6中任一项所述的模型训练方法。

12.一种意图识别设备,其特征在于,包括处理器和存储器;

所述存储器中存储有可被所述处理器执行的应用程序,用于使得所述处理器执行如权利要求7所述的意图识别方法。

13.一种计算机可读存储介质,其特征在于,其中存储有计算机可读指令,该计算机可读指令用于执行如权利要求7所述的意图识别方法。

模型训练方法、装置及意图识别方法、装置

技术领域

[0001] 本发明涉及计算机技术领域,特别是涉及一种模型训练方法、装置及意图识别方法、装置。

背景技术

[0002] 随着深度学习技术的快速发展,越来越多的传统行业被人工智能改变。人工智能会话场景是一个非常受欢迎的领域,因此吸引了众多研究者。对用户意图进行识别是很多智能会话场景的重要任务,它是信息检索、问答系统等很多应用的核心,准确识别用户意图,能够增强用户的体验感。

[0003] 意图识别是指通过分析用户当前的对话语句,利用机器学习、深度学习、自然语言处理等技术自动判断用户意图的过程。目前意图识别的做法主要包含基于特征工程、基于深度学习的方法。

[0004] 发明人在实现本发明的过程中发现有意图识别方案存在识别准确度不够高的问题。具体分析如下:

[0005] 基于特征工程的方法需要大量的、繁琐的人工操作去提取语句特征。但是,采用人工提取的方式,很难提取到全面的、有效地特征,从而影响该类方法的识别准确性。并且,由于提取的特征不容易扩展,难以具备良好的泛化性。

[0006] 基于深度学习的方法可以自动地提取有效特征,然而,在很多实际的会话场景(如客服场景)下,用户意图类型数量规模比较大,并且意图类别的数量分布极其不平衡。此种场景下,用户意图往往频繁出现在部分头部意图类别中,例如,客服场景下的“配送周期”、“什么时间出库”、“售后商品使用问题”等意图类别。而对于一些长尾意图类别(即出现频次很低的意图类别),比如“分包少件发货形式”、“什么是京贴”等,会由于出现频次很低,而导致这些类别意图的样本数据量比较少,这样,就会导致所训练得到的意图识别模型对低频次意图识别的准确性较低,进而降低基于深度学习意图识别方法的整体识别准确度。

发明内容

[0007] 有鉴于此,本发明的主要目的在于提供一种模型训练方法、装置及意图识别方法、装置,可以提高意图识别准确度。

[0008] 为了达到上述目的,本发明实施例提出的技术方案为:

[0009] 一种模型训练方法,包括:

[0010] 获取历史会话数据;

[0011] 基于所述历史会话数据中的每个用户语句,生成该用户语句对应的训练样本;

[0012] 利用所述训练样本,训练预先构建的意图识别模型,得到目标意图识别模型;其中,在利用每条训练样本进行所述训练时,在基于训练样本生成低层胶囊序列之后,采用自注意力动态路由方法,生成高层胶囊序列,将所述高层胶囊序列输入至全连接层处理。

[0013] 在一个实施方式中,所述生成该用户语句对应的训练样本包括:

[0014] 为所述用户语句,生成训练文本;

[0015] 基于所述训练文本,生成词序列;

[0016] 根据预设的意图类别集合,确定与所述用户语句相匹配的意图类别;

[0017] 基于所述词序列和所述意图类别,生成所述用户语句对应的训练样本。

[0018] 在一个实施方式中,所述生成训练文本包括:

[0019] 如果所述用户语句为所述历史会话数据中的第一个用户语句,则将所述用户语句作为训练文本;否则,将所述用户语句与在所述用户语句之前的预设数量用户语句组合,得到所述训练文本。

[0020] 在一个实施方式中,利用每条所述训练样本,进行所述训练包括:

[0021] 将所述训练样本的词序列,转化为词向量序列;

[0022] 将所述词向量序列,输入双向长短时记忆网络,并根据输出结果,生成低层胶囊序列;

[0023] 基于所述低层胶囊序列,采用自注意力动态路由方法,生成高层胶囊序列;

[0024] 将所述高层胶囊序列执行拉平操作后输入至全连接层,并生成输出结果的概率分布值;

[0025] 基于所述概率分布值,生成模型损失函数值;

[0026] 根据所述模型损失函数值和相应训练样本的意图类别标签,调整所述意图识别模型的网络参数。

[0027] 在一个实施方式中,所述根据输出结果,生成低层胶囊序列包括:

[0028] 按照 $u_i = [\vec{u}_i, \bar{u}_i]$, 对双向长短时记忆网络输出的每组前向隐藏状态值和后向隐藏状态值进行拼接,得到所述低层胶囊序列中的每个元素 u_i ;

[0029] 其中, $\vec{u}_i$ 为所述词向量序列中的第 i 个词向量对应的前向隐藏状态值, $\bar{u}_i$ 为所述词向量序列中的第 i 个词向量对应的后向隐藏状态值; $u = \{u_1, \dots, u_i \dots u_n\}$; u 为所述低层胶囊序列; n 为所述词向量序列中的向量数量; $1 \leq i \leq n$ 。

[0030] 在一个实施方式中,所述采用自注意力动态路由方法,生成高层胶囊序列包括:

[0031] 利用所述低层胶囊序列,按照 $a_{ij} = \frac{e^{\frac{(u_i W^Q)(u_j W^K)^T}{\sqrt{d_z}}}}{\sum_{k=1}^n e^{\frac{(u_i W^Q)(u_k W^K)^T}{\sqrt{d_z}}}}$, 计算注意力权重 $a_{i,j}$; 其中, i

$\in \{1, 2, \dots, n\}$, $j \in \{1, 2, \dots, m\}$; m 为预设的高层胶囊序列的元素数量;

[0032] W^Q 为当前模型中的第一权重矩阵; W^K 为当前模型中的第二权重矩阵; d_z 是预设的缩放因子;

[0033] 利用所述 $a_{i,j}$, 按照 $v_j = \frac{\left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|^2}{1 + \left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|^2} \cdot \frac{\sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right)}{\left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|}$, 计算所

述高层胶囊序列中的每个元素 v_j ;

[0034] 其中, $v = \{v_1, \dots, v_j, \dots, v_m\}$; v 为所述高层胶囊序列; $j \in \{1, 2, \dots, m\}$; W^v 为当前模型中的第三权重矩阵, $w_{i,j}$ 为当前模型中的耦合系数。

[0035] 本发明实施例还公开了一种意图识别方法,包括:

[0036] 获取目标文本;

[0037] 基于所述目标文本,生成词序列;

[0038] 基于所述词序列,利用预先训练的意图识别模型,确定所述目标文本的意图类别;

[0039] 其中,所述意图识别模型采用如上所述模型训练方法训练得到。

[0040] 一种模型训练装置,包括:

[0041] 数据获取单元,用于获取历史会话数据;

[0042] 训练样本生成单元,用于基于所述历史会话数据中的每个用户语句,生成该用户语句对应的训练样本;

[0043] 训练单元,用于利用所述训练样本,训练预先构建的意图识别模型,得到目标意图识别模型;其中,在利用每条训练样本进行所述训练时,在基于训练样本生成低层胶囊序列之后,采用自注意力动态路由方法,生成高层胶囊序列,将所述高层胶囊序列输入至全连接层处理。

[0044] 在一个实施方式中,所述训练单元,具体用于利用每条所述训练样本,进行所述训练,包括:

[0045] 将所述训练样本的词序列,转化为词向量序列;

[0046] 将所述词向量序列,输入双向长短时记忆网络,并根据输出结果,生成低层胶囊序列;

[0047] 基于所述低层胶囊序列,采用自注意力动态路由方法,生成高层胶囊序列;

[0048] 将所述高层胶囊序列执行拉平操作后输入至全连接层,并生成输出结果的概率分布值;

[0049] 基于所述概率分布值,生成模型损失函数值;

[0050] 根据所述模型损失函数值和相应训练样本的意图类别标签,调整所述意图识别模型的网络参数。

[0051] 在一个实施方式中,所述训练单元,具体用于根据输出结果,生成低层胶囊序列,包括:

[0052] 按照 $u_i = [\vec{u}_i, \bar{u}_i]$,对双向长短时记忆网络输出的每组前向隐藏状态值和后向隐藏状态值进行拼接,得到所述低层胶囊序列中的每个元素 u_i ;

[0053] 其中, $\vec{u}_i$ 为所述词向量序列中的第 i 个词向量对应的前向隐藏状态值, $\bar{u}_i$ 为所述词向量序列中的第 i 个词向量对应的后向隐藏状态值; $u = \{u_1, \dots, u_i, \dots, u_n\}$; u 为所述低层胶囊序列; n 为所述词向量序列中的向量数量; $1 \leq i \leq n$ 。

[0054] 在一个实施方式中,所述训练单元,具体用于采用自注意力动态路由方法,生成高层胶囊序列,包括:

[0055] 利用所述低层胶囊序列,按照 $a_{ij} = \frac{e^{\frac{(u_i W^Q)(u_j W^K)^T}{\sqrt{d_z}}}}{\sum_{k=1}^n e^{\frac{(u_i W^Q)(u_k W^K)^T}{\sqrt{d_z}}}}$, 计算注意力权重 $a_{i,j}$; 其中, i

$\in \{1, 2, \dots, n\}$, $j \in \{1, 2, \dots, m\}$; m 为预设的高层胶囊序列的元素数量;

[0056] W^Q 为当前模型中的第一权重矩阵; W^K 为当前模型中的第二权重矩阵; d_z 是预设的缩放因子;

[0057] 利用所述 $a_{i,j}$, 按照 $v_j = \frac{\left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|^2 \cdot \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right)}{1 + \left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|^2 \cdot \left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|}$, 计算所

述高层胶囊序列中的每个元素 v_j ;

[0058] 其中, $v = \{v_1, \dots, v_j, \dots, v_m\}$; v 为所述高层胶囊序列; $j \in \{1, 2, \dots, m\}$; W^V 为当前模型中的第三权重矩阵, $w_{i,j}$ 为当前模型中的耦合系数。

[0059] 本发明实施例公开了一种意图识别装置, 包括:

[0060] 文本获取单元, 用于获取目标文本;

[0061] 文本处理单元, 用于基于所述目标文本, 生成词序列;

[0062] 意图识别单元, 用于基于所述词序列, 利用预先训练的意图识别模型, 确定所述目标文本的意图类别;

[0063] 其中, 所述意图识别模型采用如上所述模型训练方法训练得到。

[0064] 本发明实施例公开了一种模型训练设备, 包括处理器和存储器;

[0065] 所述存储器中存储有可被所述处理器执行的应用程序, 用于使得所述处理器执行如上所述的模型训练方法。

[0066] 本发明实施例公开了一种计算机可读存储介质, 其中存储有计算机可读指令, 该计算机可读指令用于执行如上所述的模型训练方法。

[0067] 本发明实施例公开了一种意图识别设备, 包括处理器和存储器;

[0068] 所述存储器中存储有可被所述处理器执行的应用程序, 用于使得所述处理器执行如上所述的意图识别方法。

[0069] 本发明实施例公开了一种计算机可读存储介质, 其中存储有计算机可读指令, 该计算机可读指令用于执行如上所述的意图识别方法。

[0070] 由上述技术方案可见, 本发明提出的模型训练方法、装置及意图识别方法、装置, 在进行模型训练时, 在基于训练样本生成低层胶囊序列之后, 采用自注意力动态路由方法, 生成高层胶囊序列, 将所述高层胶囊序列输入至全连接层处理。如此, 低层胶囊序列通过动态路由方法得到高层胶囊序列, 高层胶囊序列为模型从训练文本里学习到的高层泛化特征, 该特征对低频意图识别非常有效, 因此, 采用本发明实施例, 可以在一定程度上缓解低频意图识别精确度低的问题, 从而可以有效地提升意图识别准确度。

附图说明

- [0071] 图1为本发明实施例一的方法流程示意图；
- [0072] 图2为本发明实施例二的方法流程示意图；
- [0073] 图3为本发明实施例三的装置结构示意图；
- [0074] 图4为本发明实施例四的装置结构示意图。

具体实施方式

[0075] 为使本发明的目的、技术方案和优点更加清楚,下面将结合附图及具体实施例对本发明作进一步地详细描述。

[0076] 图1为本发明实施例的流程示意图,如图1所示,该实施例实现的模型训练方法主要包括:

[0077] 步骤101、获取历史会话数据。

[0078] 本步骤用于获取历史会话数据,以便在后续步骤中生成用于模型训练的训练样本。

[0079] 步骤102、基于所述历史会话数据中的每个用户语句,生成该用户语句对应的训练样本。

[0080] 本步骤中,基于历史会话数据中的每个用户语句,生成每个用户语句对应的训练样本。具体地,每条训练样本将包括一词序列和相应用户语句的意图类别标签。

[0081] 在一个实施方式中,为了提高模型对意图识别的准确性,可以采用下述步骤a1~a4生成该用户语句对应的训练样本:

[0082] 步骤a1、为所述用户语句,生成训练文本。

[0083] 在一个实施方式中,可以采用下述方法,为每个用户语句生成相应的训练文本:

[0084] 如果所述用户语句为所述历史会话数据中的第一个用户语句,则将所述用户语句作为训练文本;否则,将所述用户语句与在所述用户语句之前的预设数量用户语句组合,得到所述训练文本。

[0085] 上述方法中,对于历史会话数据中第二个用户语句之后各用户语句,将利用该用户语句之前的历史用户语句,与该用户语句相组合,得到所述训练文本,即结合用户语句的上文,生成相应的训练文本,如此,可以使得训练文本包含更多的上下文信息,有利用模型从相应的训练样本中提取出更泛化的特征,从而可以提高所训练的模型的意图识别准确性。

[0086] 具体地,本领域技术人员可根据实际需要,选择进行组合的历史用户语句数量,例如,可以将当前用户语句的上一相邻用户语句与当前用户语句相组合,得到一训练文本。

[0087] 步骤a2、基于所述训练文本,生成词序列。

[0088] 在一个实施方式中,具体可以采用下述方法生成词序列:

[0089] 步骤a21、去除所述训练文本中与语义无关的标点符号。

[0090] 步骤a22、对所述去除的结果分词,得到所述训练文本的词序列。

[0091] 本步骤可采用现有的分词方法实现,在此不再赘述。

[0092] 步骤a3、根据预设的意图类别集合,确定与所述用户语句相匹配的意图类别。

[0093] 在一个实施方式中,具体可以将用户语句与预设的意图类别集合中的意图类别进

行匹配,得到与该用户语句相匹配的意图类别。

[0094] 步骤a4、基于所述词序列和所述意图类别,生成所述用户语句对应的训练样本。

[0095] 本步骤中,将利用用户语句的词序列和意图类别,构建一条训练样本。

[0096] 步骤103、利用所述训练样本,训练预先构建的意图识别模型,得到目标意图识别模型。其中,在利用每条训练样本进行所述训练时,在基于训练样本生成低层胶囊序列之后,采用自注意力动态路由方法,生成高层胶囊序列,将所述高层胶囊序列输入至全连接层处理。

[0097] 本步骤中,通过在利用每条训练样本进行所述训练时,在基于训练样本生成低层胶囊序列之后,采用自注意力动态路由方法,生成高层胶囊序列,可以从训练样本里学习到高层泛化特征,该特征对低频意图识别非常有效,从而可以在一定程度上缓解低频意图识别精确度低的问题,进而可以有效地提升意图识别准确度。

[0098] 在一个实施方式中,可以采用下述步骤x1~x6,利用每条训练样本对训练模型进行训练:

[0099] 步骤x1、将所述训练样本的词序列,转化为词向量序列。

[0100] 在一个实施方式中,可以采用利用词序列word2vec工具在公开语料训练得到的词向量,将事实描述文本词序列中的每个词映射为其对应的词向量,从而得到当前训练样本的词序列所对应的词向量序列。

[0101] 步骤x2、将所述词向量序列,输入双向长短时记忆网络,并根据输出结果,生成低层胶囊序列。

[0102] 在一个实施方式中,具体可以采用下述方法根据输出结果,生成低层胶囊序列:

[0103] 按照 $u_i = [\vec{u}_i, \bar{u}_i]$,对双向长短时记忆网络输出的每组前向隐藏状态值和后向隐藏状态值进行拼接,得到所述低层胶囊序列中的每个元素 u_i ;

[0104] 其中, $\vec{u}_i$ 为所述词向量序列中的第i个词向量对应的前向隐藏状态值, $\bar{u}_i$ 为所述词向量序列中的第i个词向量对应的后向隐藏状态值; $u = \{u_1, \dots, u_i \dots u_n\}$;u为所述低层胶囊序列;n为所述词向量序列中的向量数量; $1 \leq i \leq n$ 。

[0105] 具体地,词向量序列中的每个词向量,将依次输入至双向长短时记忆网络(Bi-LSTM)处理。

[0106] 双向长短时记忆网络具体可以采用下述公式对词向量序列 $u = \{u_1, \dots, u_i \dots u_n\}$ 进行处理输出每个时刻t的前向隐藏状态值 $\vec{u}_t$:

$$[0107] \quad f_t = \sigma(W_f \cdot [\vec{u}_{t-1}, \text{vec}_t] + b_f)$$

$$[0108] \quad i_t = \sigma(W_i \cdot [\vec{u}_{t-1}, \text{vec}_t] + b_i)$$

$$[0109] \quad o_t = \sigma(W_o \cdot [\vec{u}_{t-1}, \text{vec}_t] + b_o)$$

$$[0110] \quad \tilde{C}_t = \tanh(W_c \cdot [\vec{u}_{t-1}, \text{vec}_t] + b_c)$$

$$[0111] \quad C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$$[0112] \quad \vec{u}_t = o_t * \tanh(C_t)$$

[0113] 其中 f_t, i_t, o_t 分别是LSTM的遗忘门、输入门、输出门; $\tilde{C}_t$ 表示当前时刻的候选值, C_t 表示当前时刻 t 的状态; C_{t-1} 表示上一时刻 $t-1$ 的状态; W_f, W_t, W_o, W_c 为权重矩阵; b_f, b_i, b_o, b_c 为偏置向量; $\vec{u}_{t-1}$ 表示上一时刻 $t-1$ 输出的前向隐藏状态值; $\sigma()$ 表示sigmoid函数; $\tanh()$ 为激活函数; vec_t 表示词向量序列中的第 t 个词向量。

[0114] 对于后向隐藏状态值 $\vec{u}_t$,可以采用与前向隐藏状态值 $\vec{u}_t$ 同样的计算方式得到,在此不再赘述。

[0115] 步骤x3、基于所述低层胶囊序列,采用自注意力动态路由方法,生成高层胶囊序列。

[0116] 在一个实施方式中,可以采用下述方法利用自注意力动态路由方法,生成高层胶囊序列:

[0117] 步骤x31、利用所述低层胶囊序列,按照
$$a_{ij} = \frac{e^{\frac{(u_i W^Q)(u_j W^K)^T}{\sqrt{d_z}}}}{\sum_{k=1}^n e^{\frac{(u_i W^Q)(u_k W^K)^T}{\sqrt{d_z}}}}$$
,计算注意力权重

$a_{i,j}$;其中, $i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, m\}$; m 为预设的高层胶囊序列的元素数量;

[0118] W^Q 为当前模型中的第一权重矩阵; W^K 为当前模型中的第二权重矩阵; d_z 是预设的缩放因子。

[0119] 步骤x32、利用所述 $a_{i,j}$,按照
$$v_j = \frac{\left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|^2 \cdot \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right)}{1 + \left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|^2 \cdot \left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|^2},$$

计算所述高层胶囊序列中的每个元素 v_j ;

[0120] 其中, $v = \{v_1, \dots, v_j, \dots, v_m\}$; v 为所述高层胶囊序列; $j \in \{1, 2, \dots, m\}$; W^V 为当前模型中的第三权重矩阵, $w_{i,j}$ 为当前模型中的耦合系数。

[0121] 上述方法利用自注意力动态路由方法,来生成高层胶囊序列,如此,可以从训练样本中提取出更泛化的特征,从而可以有效克服低频次意图类别的样本数量少而导致的低频意图识别精确度低的问题。

[0122] 步骤x4、将所述高层胶囊序列执行拉平操作后输入至全连接层,并生成输出结果的概率分布值。

[0123] 本步骤中,对所述高层胶囊序列执行拉平操作,是为了获得 $1 \times R$ 维的序列,以便可以输入至全连接层,其中 $R = n \times m$ 。

[0124] 在一个实施方式中,可以利用softmax函数,生成所述全连接层的输出结果的概率分布值。

[0125] 步骤x5、基于所述概率分布值,生成模型损失函数值。

[0126] 在一个实施方式中,可以利用focal loss损失函数,即按照 $L = -\alpha(1 - \hat{y})^r \log(\hat{y})$,基于所述概率分布值,生成模型损失函数值。

[0127] 其中, $\hat{y}$ 为步骤x4中得到的概率分布值, α 表示第一调节因子, γ 表示第二调节因子, L 表示模型损失函数值。

[0128] 在一个实施方式中, 还可以利用交叉熵, 基于所述概率分布值, 生成模型损失函数值。具体方法为本领域技术人员所掌握, 在此不再赘述。

[0129] 步骤x6、根据所述模型损失函数值和相应训练样本的意图类别标签, 调整所述意图识别模型的网络参数。

[0130] 本步骤中, 将基于所述模型损失函数值和相应训练样本的意图类别标签, 对意图识别模型的网络参数进行调整, 从而以最小化损失函数值为目标, 对模型进行优化。具体地, 可以采用Adam优化器对模型的网络参数进行优化。

[0131] 与上述实施例一的模型训练方法相对应, 本发明实施例二公开了一种意图识别方法, 如图2所示包括:

[0132] 步骤201、获取目标文本。

[0133] 本步骤用于获取当前需要进行意图识别的目标文本。

[0134] 步骤202、基于所述目标文本, 生成词序列。

[0135] 本步骤中生成词序列的方法同上述实施例一步骤1022中生成词序列的方法相同, 在此不再赘述。

[0136] 步骤203、基于所述词序列, 利用预先训练的意图识别模型, 确定所述目标文本的意图类别。

[0137] 其中, 所述意图识别模型采用上述实施例一中模型训练方法训练得到。

[0138] 本步骤中, 将使用的意图识别模型是采用上述实施例一中的方法训练得到、如上述方法实施例一中所述, 该模型在训练时采用自注意力动态路由方法, 生成高层胶囊序列, 将高层胶囊序列输入至全连接层处理, 有效提升了意图识别准确度, 因此, 本步骤采用意图识别模型对目标文本的意图类型进行识别, 可以获得较高的意图识别准确性。

[0139] 与上述实施例一的模型训练方法相对应, 本发明实施例三公开了一种模型训练装置, 如图3所示, 该装置包括:

[0140] 数据获取单元301, 用于获取历史会话数据;

[0141] 训练样本生成单元302, 用于基于所述历史会话数据中的每个用户语句, 生成该用户语句对应的训练样本;

[0142] 训练单元303, 用于利用所述训练样本, 训练预先构建的意图识别模型, 得到目标意图识别模型; 其中, 在利用每条训练样本进行所述训练时, 在基于训练样本生成低层胶囊序列之后, 采用自注意力动态路由方法, 生成高层胶囊序列, 将所述高层胶囊序列输入至全连接层处理。

[0143] 在一个实施方式中, 所述训练单元303, 具体用于利用每条所述训练样本, 进行所述训练, 包括:

[0144] 将所述训练样本的词序列, 转化为词向量序列;

[0145] 将所述词向量序列, 输入双向长短时记忆网络, 并根据输出结果, 生成低层胶囊序列;

[0146] 基于所述低层胶囊序列, 采用自注意力动态路由方法, 生成高层胶囊序列;

[0147] 将所述高层胶囊序列执行拉平操作后输入至全连接层, 并生成输出结果的概率分

布值;

[0148] 基于所述概率分布值,生成模型损失函数值;

[0149] 根据所述模型损失函数值和相应训练样本的意图类别标签,调整所述意图识别模型的网络参数。

[0150] 在一个实施方式中,所述训练单元303,具体用于根据输出结果,生成低层胶囊序列,包括:

[0151] 按照 $u_i = [\vec{u}_i, \bar{u}_i]$, 对双向长短时记忆网络输出的每组前向隐藏状态值和后向隐藏状态值进行拼接,得到所述低层胶囊序列中的每个元素 u_i ;

[0152] 其中, $\vec{u}_i$ 为所述词向量序列中的第 i 个词向量对应的前向隐藏状态值, $\bar{u}_i$ 为所述词向量序列中的第 i 个词向量对应的后向隐藏状态值; $u = \{u_1, \dots, u_i \dots u_n\}$; u 为所述低层胶囊序列; n 为所述词向量序列中的向量数量; $1 \leq i \leq n$ 。

[0153] 在一个实施方式中,所述训练单元303,具体用于采用自注意力动态路由方法,生成高层胶囊序列,包括:

[0154] 利用所述低层胶囊序列,按照 $a_{ij} = \frac{e^{\frac{(u_i W^Q)(u_j W^K)^T}{\sqrt{d_z}}}}{\sum_{k=1}^n e^{\frac{(u_i W^Q)(u_k W^K)^T}{\sqrt{d_z}}}}$, 计算注意力权重 $a_{i,j}$; 其中, $i \in$

$\{1, 2, \dots, n\}$, $j \in \{1, 2, \dots, m\}$; m 为预设的高层胶囊序列的元素数量;

[0155] W^Q 为当前模型中的第一权重矩阵; W^K 为当前模型中的第二权重矩阵; d_z 是预设的缩放因子;

[0156] 利用所述 $a_{i,j}$, 按照 $v_j = \frac{\left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|^2 \cdot \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right)}{1 + \left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|^2 \cdot \left\| \sum_{i=1}^n w_{ij} \cdot \left(\sum_{j=1}^m a_{ij} \cdot u_j W^V \right) \right\|}$, 计算所

述高层胶囊序列中的每个元素 v_j ;

[0157] 其中, $v = \{v_1, \dots, v_j, \dots, v_m\}$; v 为所述高层胶囊序列; $j \in \{1, 2, \dots, m\}$; W^V 为当前模型中的第三权重矩阵, $w_{i,j}$ 为当前模型中的耦合系数。

[0158] 与上述实施例二的意图识别方法相对应,本发明实施例四公开了本发明实施例公开了一种意图识别装置,如图4所示,该装置包括:

[0159] 文本获取单元401,用于获取目标文本;

[0160] 文本处理单元403,用于基于所述目标文本,生成词序列;

[0161] 意图识别单元403,用于基于所述词序列,利用预先训练的意图识别模型,确定所述目标文本的意图类别;

[0162] 其中,所述意图识别模型采用上述实施例二所述的模型训练方法训练得到。

[0163] 与上述实施例一的模型训练方法相对应,本发明实施例公开了一种模型训练设备,包括处理器和存储器;

[0164] 所述存储器中存储有可被所述处理器执行的应用程序,用于使得所述处理器执行如上所述的模型训练方法。

[0165] 与上述实施例一的模型训练方法相对应,本发明实施例还公开了一种计算机可读存储介质,其中存储有计算机可读指令,该计算机可读指令用于执行如上实施例一所述的模型训练方法。

[0166] 与上述实施例二的意图识别方法相对应,本发明实施例还公开了一种意图识别设备,包括处理器和存储器;

[0167] 所述存储器中存储有可被所述处理器执行的应用程序,用于使得所述处理器执行如上实施例一所述的意图识别方法。

[0168] 与上述实施例二的意图识别方法相对应,本发明实施例公开了一种计算机可读存储介质,其中存储有计算机可读指令,该计算机可读指令用于执行如上实施例一所述的意图识别方法。

[0169] 由上述技术方案可见,本发明提出的模型训练方法、装置及意图识别方法、装置,在进行模型训练时,在基于训练样本生成低层胶囊序列之后,采用自注意力动态路由方法,生成高层胶囊序列,将所述高层胶囊序列输入至全连接层处理。如此,低层胶囊序列通过动态路由方法得到高层胶囊序列,高层胶囊序列为模型从训练文本里学习到的高层泛化特征,该特征对低频意图识别非常有效,因此,采用本发明实施例,可以在一定程度上缓解低频意图识别精确度低的问题,从而可以有效地提升意图识别准确度。在实际应用中本发明实施例可以适用于很多会话场景,例如,客服场景中的问答系统、信息检索等。

[0170] 上述实施例中,存储器具体可以实施为电可擦可编程只读存储器(EEPROM)、快闪存储器(Flash memory)、可编程程序只读存储器(PROM)等多种存储介质。处理器可以实施为包括一或多个中央处理器或一或多个现场可编程门阵列,其中现场可编程门阵列集成一或多个中央处理器核。具体地,中央处理器或中央处理器核可以实施为CPU或MCU。

[0171] 需要说明的是,上述各流程和各结构图中不是所有的步骤和模块都是必须的,可以根据实际的需要忽略某些步骤或模块。各步骤的执行顺序不是固定的,可以根据需要进行调整。各模块的划分仅仅是为了便于描述采用的功能上的划分,实际实现时,一个模块可以由多个模块实现,多个模块的功能也可以由同一个模块实现,这些模块可以位于同一个设备中,也可以位于不同的设备中。

[0172] 各实施方式中的硬件模块可以以机械方式或电子方式实现。例如,一个硬件模块可以包括专门设计的永久性电路或逻辑器件(如专用处理器,如FPGA或ASIC)用于完成特定的操作。硬件模块也可以包括由软件临时配置的可编程逻辑器件或电路(如包括通用处理器或其它可编程处理器)用于执行特定操作。至于具体采用机械方式,或是采用专用的永久性电路,或是采用临时配置的电路(如由软件进行配置)来实现硬件模块,可以根据成本和时间上的考虑来决定。

[0173] 用于提供程序代码的存储介质实施方式包括软盘、硬盘、磁光盘、光盘(如CD-ROM、CD-R、CD-RW、DVD-ROM、DVD-RAM、DVD-RW、DVD+RW)、磁带、非易失性存储卡和ROM。可选择地,可以由通信网络从服务器计算机或云上下载程序代码。

[0174] 在本文中,“示意性”表示“充当实例、例子或说明”,不应将在本文中被描述为“示意性”的任何图示、实施方式解释为一种更优选的或更具优点的技术方案。为使图面简洁,

各图中的只示意性地表示出了与本发明相关部分,而并不代表其作为产品的实际结构。另外,以使图面简洁便于理解,在有些图中具有相同结构或功能的部件,仅示意性地绘示了其中的一个,或仅标出了其中的一个。在本文中,“一个”并不表示将本发明相关部分的数量限制为“仅此一个”,并且“一个”不表示排除本发明相关部分的数量“多于一个”的情形。在本文中,“上”、“下”、“前”、“后”、“左”、“右”、“内”、“外”等仅用于表示相关部分之间的相对位置关系,而非限定这些相关部分的绝对位置。

[0175] 以上所述,仅为本发明的较佳实施例而已,并非用于限定本发明的保护范围。凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

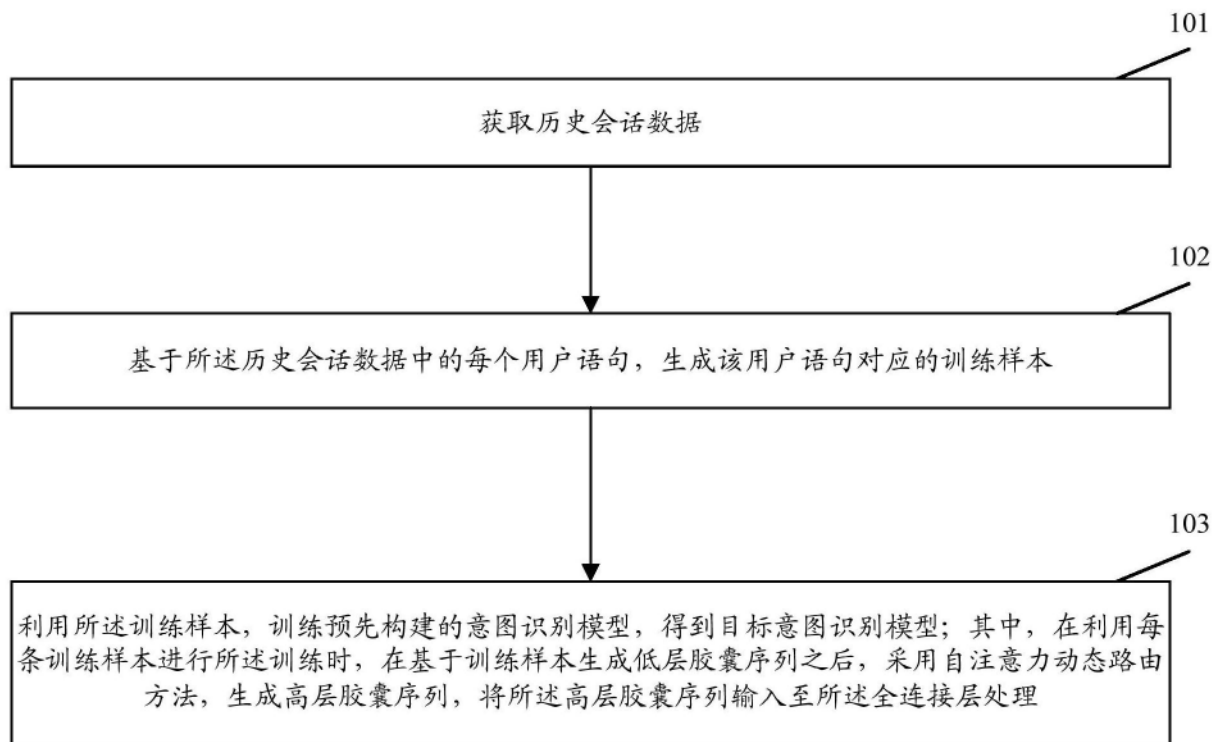


图1

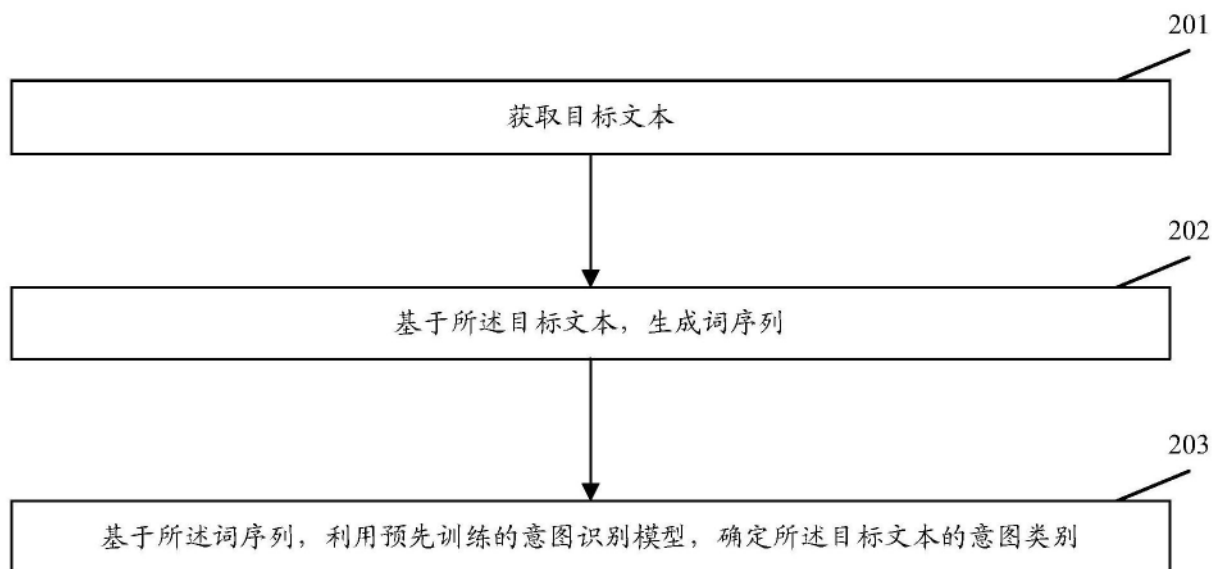


图2

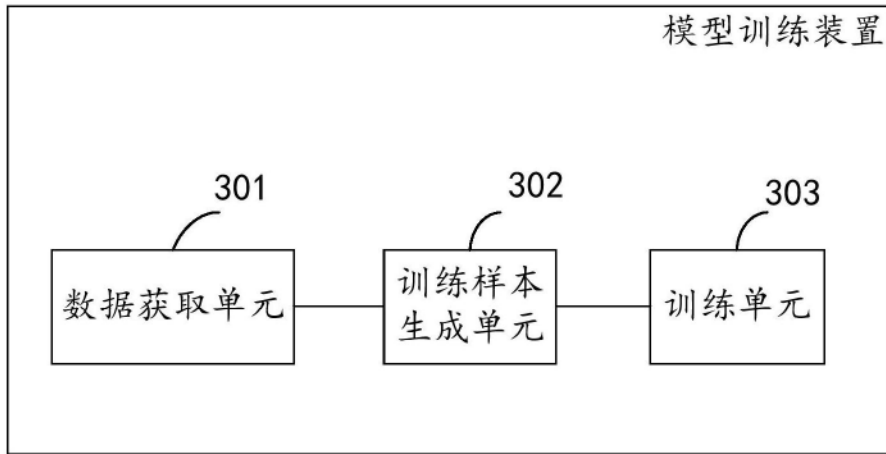


图3

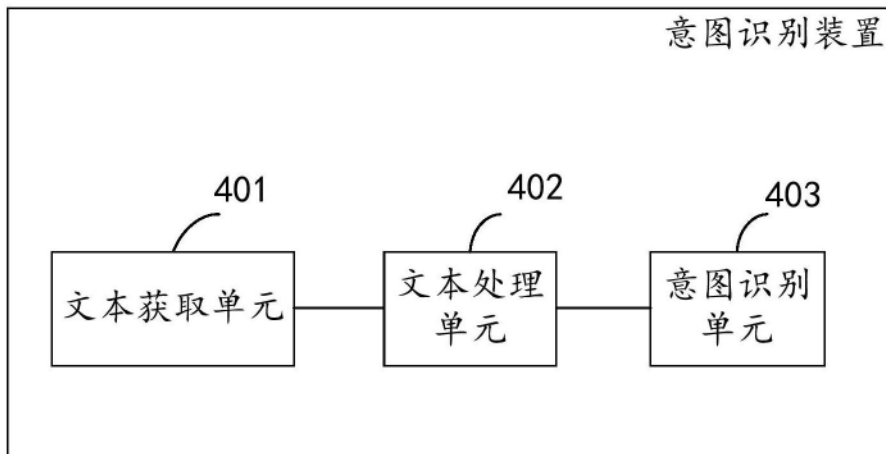


图4