

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 12 月 17 日 (17.12.2020)



(10) 国际公布号
WO 2020/248366 A1

(51) 国际专利分类号:
G06F 16/33 (2019.01) **G06F 16/35** (2019.01)

(21) 国际申请号: PCT/CN2019/102207

(22) 国际申请日: 2019 年 8 月 23 日 (23.08.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910525743.3 2019年6月14日 (14.06.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国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, Guangdong 518000 (CN)。

(72) 发明人: 王健宗(WANG, JianZong); 中国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, Guangdong 518000 (CN)。
程宁(CHENG, Ning); 中国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, Guangdong 518000 (CN)。

(74) 代理人: 深圳市沃德知识产权代理事务所(普通合伙)(SHENZHEN WORLD INTELLECTUAL PROPERTY AGENCY (GENERAL PARTNERSHIP)); 中国广东省深圳市福田区园岭街道八卦四路 10 号中浩大厦 1528-1530 室于志光, Guangdong 518000 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,

(54) **Title:** TEXT INTENTION INTELLIGENT CLASSIFICATION METHOD AND DEVICE, AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 文本意图智能分类方法、装置及计算机可读存储介质

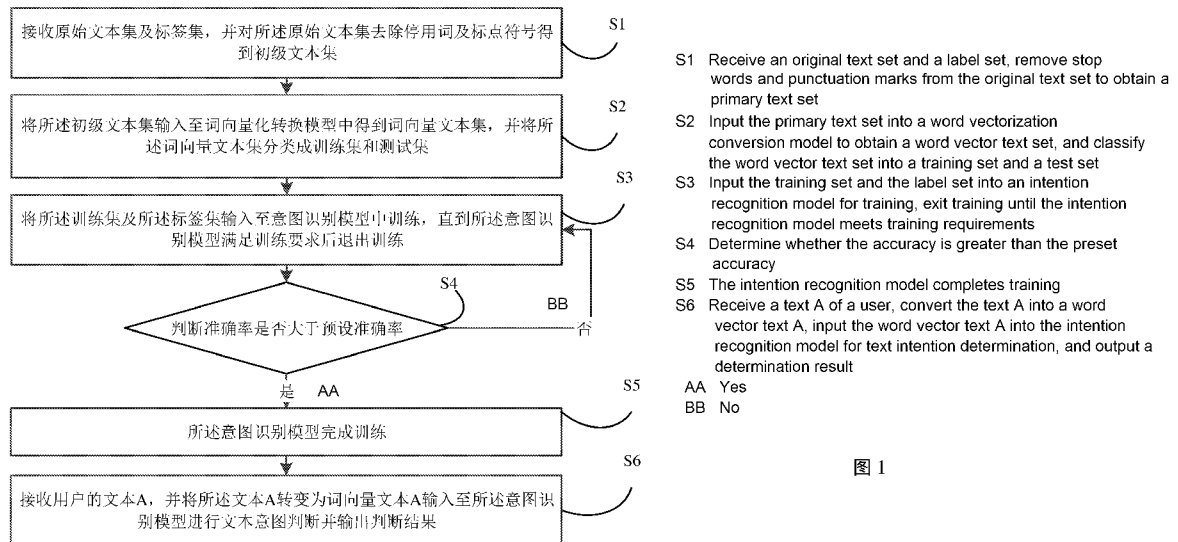


图 1

(57) **Abstract:** A text intention intelligent classification method, a text intention intelligent classification device, and a computer-readable storage medium. The method comprises: receiving an original text set and a label set, removing stop words and punctuation marks from the original text set to obtain a primary text set (S1); inputting the primary text set into a word vectorization conversion model to obtain a word vector text set, and classifying the word vector text set into a training set and a test set (S2); inputting the training set and the label set into an intention recognition model for training, exiting training until the intention recognition model meets training requirements (S3); and receiving a text A of a user, converting the text A into a word vector text A, inputting the word vector text A into the intention recognition model for text intention determination, and outputting a determination result (S6). An accurate text intention intelligent classification function can be implemented.



GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 一种文本意图智能分类方法, 文本意图智能分类装置, 以及一种计算机可读存储介质, 其中方法包括: 接收原始文本集及标签集, 并对原始文本集去除停用词及标点符号得到初级文本集(S1), 将初级文本集输入至词向量化转换模型中得到词向量文本集, 并将词向量文本集分类成训练集和测试集(S2), 将训练集及标签集输入至意图识别模型中训练, 直到意图识别模型满足训练要求后退出训练(S3), 接收用户的文本A, 并将文本A转变为词向量文本A输入至意图识别模型进行文本意图判断并输出判断结果(S6)。可以实现精准的文本意图智能分类功能。

文本意图智能分类方法、装置及计算机可读存储介质

本申请基于巴黎公约申明享有2019年6月14日递交的申请号为CN 201910525743.3、名称为“文本意图智能分类方法、装置及计算机可读存储介
5 质”的中国专利申请的优先权，该中国专利申请的整体内容以参考的方式结合在本申请中。

技术领域

本申请涉及人工智能技术领域，尤其涉及一种可用于智能化的文本意图
10 分类方法、装置及计算机可读存储介质。

背景技术

目前，各个行业，特别是电商行业，智能回答系统都有了一定程度的普及。但目前的智能回答系统一般使用关键字规则或简单的检索来匹配答案后
15 回答客户的问题，这样的回答方式往往答非所问，智能程度不够高，因此不能很好的解决客户所提出的问题，既浪费客户的时间，也浪费服务器的计算资源，并没有达到减轻人工客服压力的初衷。

发明内容

20 鉴于以上内容，本申请提供一种文本意图智能分类方法、装置及计算机可读存储介质。其目的在于当用户输入文本数据时，对所述用户输入的文本进行意图判断并输出判断结果。

为实现上述目的，本申请提供一种文本意图智能分类方法，该方法包括：

步骤 A：接收原始文本集及标签集，并对所述原始文本集去除停用词及
25 标点符号得到初级文本集；

步骤 B：将所述初级文本集转换为词向量文本集，并将所述词向量文本集分类成训练集和测试集；

步骤 C：将所述训练集及所述标签集输入至预先构建的意图识别模型中训练，直到所述意图识别模型满足预设训练要求后退出训练；

30 步骤 D：将所述测试集输入至所述意图识别模型中进行文本意图判断，

计算对所述文本意图的判断结果与所述标签集中内容的匹配准确率，若所述匹配准确率小于预设准确率，返回步骤 C，若所述匹配准确率大于所述预设准确率，所述意图识别模型完成训练；

5 步骤 E：接收用户的文本，并将所述文本转变为词向量文本输入至所述意图识别模型进行文本意图判断，并输出判断结果。

此外，为实现上述目的，本申请还提供一种文本意图智能分类装置，所述装置包括存储器和处理器，所述存储器上存储有可在所述处理器上运行的文本意图智能分类程序，所述文本意图智能分类程序被所述处理器执行，可实现如下步骤：

10 步骤 A：接收原始文本集及标签集，并对所述原始文本集去除停用词及标点符号得到初级文本集；

步骤 B：将所述初级文本集输入至词向量化转换模型中得到词向量文本集，并将所述词向量文本集分类成训练集和测试集；

15 步骤 C：将所述训练集及所述标签集输入至意图识别模型中训练，直到所述意图识别模型满足训练要求后退出训练；

步骤 D：将所述测试集输入至所述意图识别模型中进行文本意图判断，计算所述文本意图判断与所述标签集中内容是否相同并得到判断准确率，若所述判断准确率小于预设准确率，返回步骤 C，若判断准确率大于预设准确率，所述意图识别模型完成训练；

20 步骤 E：接收用户的文本 A，并将所述文本 A 转变为词向量文本 A 输入至所述意图识别模型进行文本意图判断，并输出判断结果。

此外，为实现上述目的，本申请还提供一种计算机可读存储介质，所述计算机可读存储介质中包括文本意图智能分类程序，所述文本意图智能分类程序被处理器执行时，可实现如上所述文本意图智能分类方法中的任意步骤。

25 本申请通过将所述初级文本集转换为词向量文本集可以有效的提高对文本特征的采集，增加文本分类的准确率，同时意图识别模型是基于深度学习，可有效的根据所述文本特征识别出文本关键字，并基于关键字进行意图分类，因此本申请提出的文本意图智能分类方法、装置及计算机可读存储介质可以实现精准的文本意图智能分类功能。

30

附图说明

图 1 为本申请一实施例提供的文本意图智能分类方法的流程示意图；

图 2 为本申请一实施例提供的文本意图智能分类装置的内部结构示意图；

图 3 为本申请一实施例提供的文本意图智能分类装置中文本意图智能分类程序的模块示意图。

本申请目的的实现、功能特点及优点将结合实施例，参照附图做进一步说明。

具体实施方式

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。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

本申请提供一种文本意图智能分类方法。参照图 1 所示，为本申请一实施例提供的文本意图智能分类方法的流程示意图。该方法可以由一个装置执行，该装置可以由软件和/或硬件实现。

在本实施例中，文本意图智能分类方法包括：

S1、接收原始文本集及标签集，并对所述原始文本集去除停用词及标点符号得到初级文本集。

较佳地，所述原始文本集包括各种通过文字记录下的问题，如问题 A：“什么时候发货”、问题 B：“为什么还没有物流信息”、问题 C：“双十一是否会降价”、问题 D：“近期价格是否会发生变化？”等之类的问题。

进一步地，所述标签集记录所述原始文本集内各种问题的分类结果，如问题 A：“什么时候发货”和问题 B：“为什么还没有物流信息”在所述标签集中被记录为“发货咨询”；问题 C：“双十一是否会降价”和问题 D：“近期价格是否会发生变化？”在所述标签集中被记录为“价格变动咨询”。

本申请较佳实施例，所述停用词包括那个、这个、哎呀、比如、吧、但是等，所述标点符号包括句号、逗号、问好等。进一步地，本申请通过制作中文停用词表和中文标点符号表，通过比对所述原始文本集与所述中文停用

词表和中文标点符号表是否有相同部分，达到去除所述停用词及所述标点符号的目的。如问题 D：“近期价格是否会发生变化？”经过去除所述停用词及所述标点符号的操作变为：“近期价格发生变化”。

S2、将所述初级文本集输入至词向量化转换模型中得到词向量文本集，
5 并将所述词向量文本集分类成训练集和测试集。

进一步地，利用词向量化转换模型将所述初级文本集转换为词向量文本集，所述词向量化转换模型包括输入层、投影层和输出层。

较佳地，本申请较佳实施例将所述初级文本集输入至所述输入层，所述输入层确定所述初级文本集内每个词语 ω 在所述初级文本集的出现位置
10 $Context(\omega)$ ，并将所述出现位置 $Context(\omega)$ 输入至所述投射层。如上述“近期价格发生变化”经过所述输入层处理后得到 $Context(\omega_1)$ 、 $Context(\omega_2)$ 、 $Context(\omega_3)$ 、 $Context(\omega_4)$ ，其中 ω_1 为近期， ω_2 为价格， ω_3 为发生， ω_4 为变化， $Context$ 为所述“近期价格发生变化”，则 $Context(\omega_1) = p(1|4)$ 、 $Context(\omega_2) = p(2|4)$ ，依次类推。

15 进一步地，所述投射层基于所述出现位置 $Context(\omega)$ 做累加求和操作得到累加求和矩阵 X_ω ，并根据所述 X_ω 建立概率模型。所述累加求和操作得到累加求和矩阵 X_ω 为：

$$X_\omega = \sum_{i=1}^{2^c} V(Context(\omega_i))$$

其中， $V(Context(\omega_i))$ 是所述出现位置 $Context(\omega)$ 的矩阵表示形式， c 表示所述初级文本集每个文本的词语数量，如所述“近期价格发生变化”的词语数量为 4。所述概率模型为：

$$p(\omega|Context(\omega)) = \prod_{j=2}^{\omega} p(d_j^\omega|X_\omega, \theta_{j-1}^\omega)$$

其中， $p(\omega|Context(\omega))$ 为所述概率模型， $Context$ 为所述初级文本集， ω 为所述初级文本集内每个词语， d_j^ω 表示在路径 p^ω 内，第 j 个结点对应的
25 Huffman 编码 θ_{j-1}^ω 表示路径 p^ω 内第 j 个非叶子结点对应的向量。所述 Huffman 编码是根据数据通信知识用 0、1 码的不同排列来表示词语，词语称为叶子结点，各叶子结点的权值通过 Huffman 编码表现。如所述问题“近期价格发生变化”经过所述输入层处理后得到 $Context(\omega_1)$ ，其中 ω_1 为“近期”，则所述“近期”为所述叶子结点，通过所述 Huffman 编码操作后得到编码形式，所述
30 编码形式可用 00010 表示，当所述“近期”为所述叶子结点时，所述“价

格”、“发生”、“变化”为非叶子结点，所述叶子结点与所述非叶子结点统称为结点。

进一步地，所述输出层根据所述概率模型建立对数似然函数，并最大化所述对数似然函数得到词向量文本集，所述对数似然函数 ζ 为：

$$\zeta = \sum_{\omega \in \varsigma} p(\omega | \text{Context}(\omega))$$

其中， ζ 为所述对数似然函数， ς 是包含了所述初级文本集所有内容的集合，称为语料，进一步地，所述对数似然函数 ζ 基于所述概率模型可进一步扩展为：

$$\zeta = \sum_{\omega \in \varsigma} \log \prod_{j=2}^{l^\omega} \{ [\sigma(X_\omega^T \theta_{j-1}^\omega)]^{1-d_j^\omega} [1 - \sigma(X_\omega^T \theta_{j-1}^\omega)]^{d_j^\omega} \}$$

其中， l^ω 表示所述路径 p^ω 中包括结点的数量， σ 为阈值函数，所述阈值函数可用 sigmoid 函数。

较佳地，所述最大化所述对数似然函数的方法为：

$$\begin{aligned} \frac{\partial \zeta(\omega, j)}{\partial \theta_{j-1}^\omega} &= \frac{\partial}{\partial \theta_{j-1}^\omega} \{ (1 - d_j^\omega) \log [\sigma(X_\omega^T \theta_{j-1}^\omega)] + d_j^\omega \log [1 - \sigma(X_\omega^T \theta_{j-1}^\omega)] \} \\ &= (1 - d_j^\omega) [1 - \sigma(X_\omega^T \theta_{j-1}^\omega)] X_\omega - d_j^\omega \sigma(X_\omega^T \theta_{j-1}^\omega) X_\omega \\ &= [1 - d_j^\omega - \sigma(X_\omega^T \theta_{j-1}^\omega)] X_\omega \end{aligned}$$

基于所述方法最大化 $\frac{\partial \zeta(\omega, j)}{\partial \theta_{j-1}^\omega}$ 后得到最优化所述概率模型，基于所述最优化

概率模型得到最优化累加求和矩阵 X_ω ，进一步地，基于所述最优化累加求和

矩阵 X_ω 得到词向量，因此所述初级文本集可转变为词向量文本集。

较佳地，将所述词向量文本集按照 8:2 的数量随机划分为训练集和测试集。

S3、将所述训练集及所述标签集输入至意图识别模型中训练，直到所述意图识别模型满足训练要求后退出训练。

本申请较佳实施所述意图识别模型包括卷积神经网络，所述卷积神经网络共有十六层卷积层和十六池化层、一层全连接层。当所述卷积神经网络接收所述训练集后，将所述训练集输入至第一层卷积层，所述第一层卷积层进行卷积操作后得到第一卷积数据集输入至第一层池化层；所述第一层池化层对所述第一卷积数据集进行最大化池化操作后得到第一降维数据集输入至第

二层卷积层；所述第二层卷积层再次进行所述卷积操作后得到第二卷积数据集输入至第二层池化层进行所述最大化池化操作得到第二降维数据集，以此类推，直至最终得到第十六降维数据集，将所述第十六降维数据集输入至全连接层。

5 较佳地，所述全连接层接收所述第十六降维数据集，并结合激活函数计算得到所述训练值集合，并将所述训练值集合和所述标签集输入至所述意图识别模型的损失函数中，所述损失函数计算出损失值，判断所述损失值与预设训练阈值的大小关系，直至所述损失值小于所述预设训练阈值时，所述意图识别模型退出训练，所述预设训练阈值一般设置为 0.1。

10 本申请较佳实施例所述卷积层的卷积操作为：

$$\omega' = \frac{\omega + 2p - k}{s} + 1$$

其中 ω' 为输出数据， ω 为输入数据， k 为所述卷积层的卷积核大小， s 为所述卷积操作的步幅， p 为数据补零矩阵；

所述激活函数为：

$$y = \frac{1}{1 + e^{-\omega}}$$

其中 y 为所述训练值集合， e 为无限不循环小数。

15 本申请较佳实施例所述损失值 T 为：

$$T = \sum_{t=1}^n |y_t - \mu_t|^2$$

其中， n 为所述标签集的大小， y_t 为所述训练值集合， μ_t 为所述标签集。

S4、将所述测试集输入至所述意图识别模型中进行意图判断，判断准确率和预设准确率的大小关系。

若判断准确率小于预设准确率，返回 S3。

20 进一步地，所述步骤 S4 包括：

所述意图识别模型基于所述测试集进行上述卷积操作、池化操作、激活操作后得到所述测试集的意图分类集合，依次比对所述意图分类集合与所述标签集是否相同，并计算相同的数量，将所述相同的数量除以所述测试集总数得到准确率。

25 例如，所述意图识别模型将测试集中“近期价格发生变化”判断为“价格变动咨询”，与实际被分类为“价格变动咨询”一致，因此所述意图识别模

型判断正确。进一步地,所述准确率一般设置为 95%,当所述准确率小于 95% 时,重新返回 S3 继续训练。

若判断准确率大于预设准确率,则 S5、所述意图识别模型完成训练。

5 S6、接收用户的文本 A,并将所述文本 A 转变为词向量文本 A 输入至所述意图识别模型进行意图判断,并输出判断结果。

本申请提供一种文本意图智能分类装置。参照图 2 所示,为本申请一实施例提供的文本意图智能分类装置的内部结构示意图。

10 在本实施例中,所述文本意图智能分类装置 1 可以是 PC (Personal Computer,个人电脑),或者是智能手机、平板电脑、便携计算机等终端设备,也可以是一种服务器等。该文本意图智能分类装置 1 至少包括存储器 11、处理器 12,通信总线 13,以及网络接口 14。

15 其中,存储器 11 至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器(例如,SD 或 DX 存储器等)、磁性存储器、磁盘、光盘等。存储器 11 在一些实施例中可以是文本意图智能分类装置 1 的内部存储单元,例如该文本意图智能分类装置 1 的硬盘。存储器 11 在另一些实施例中也可以是文本意图智能分类装置 1 的外部存储设备,例如文本意图智能分类装置 1 上配备的插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital, SD)卡,闪存卡(Flash Card)等。进一步地,存储器 11 还可以既包括文本意图智能分类装置 1 的内部存储单元也包括外部存储设备。存储器 11 不仅可以用于存储安装于文本意图智能分类装置 1 的应用软件及各类数据,例如文本意图智能分类程序 01 的代码等,还可以用于暂时地存储已经输出或者将要输出的数据。

25 处理器 12 在一些实施例中可以是一中央处理器(Central Processing Unit, CPU)、控制器、微控制器、微处理器或其他数据处理芯片,用于运行存储器 11 中存储的程序代码或处理数据,例如执行文本意图智能分类程序 01 等。

通信总线 13 用于实现这些组件之间的连接通信。

网络接口 14 可选的可以包括标准的有线接口、无线接口(如 WI-FI 接口),通常用于在该装置 1 与其他电子设备之间建立通信连接。

30 可选地,该装置 1 还可以包括用户接口,用户接口可以包括显示器(Display)、输入单元比如键盘(Keyboard),可选的用户接口还可以包括标

准的有线接口、无线接口。可选地，在一些实施例中，显示器可以是 LED 显示器、液晶显示器、触控式液晶显示器以及 OLED（Organic Light-Emitting Diode，有机发光二极管）触摸器等。其中，显示器也可以适当的称为显示屏或显示单元，用于显示在文本意图智能分类装置 1 中处理的信息以及用于显示可视化的用户界面。

图 2 仅示出了具有组件 11-14 以及文本意图智能分类程序 01 的文本意图智能分类装置 1，本领域技术人员可以理解的是，图 1 示出的结构并不构成对文本意图智能分类装置 1 的限定，可以包括比图示更少或者更多的部件，或者组合某些部件，或者不同的部件布置。

在图 2 所示的装置 1 实施例中，存储器 11 中存储有文本意图智能分类程序 01；处理器 12 执行存储器 11 中存储的文本意图智能分类程序 01 时实现如下步骤：

步骤一、接收原始文本集及标签集，并对所述原始文本集去除停用词及标点符号得到初级文本集。

较佳地，所述原始文本集包括各种通过文字记录下的问题，如问题 A：“什么时候发货”、问题 B：“为什么还没有物流信息”、问题 C：“双十一是否会降价”、问题 D：“近期价格是否会发生变化？”等之类的问题。

进一步地，所述标签集记录所述原始文本集内各种问题的分类结果，如问题 A：“什么时候发货”和问题 B：“为什么还没有物流信息”在所述标签集中被记录为“发货咨询”；问题 C：“双十一是否会降价”和问题 D：“近期价格是否会发生变化？”在所述标签集中被记录为“价格变动咨询”。

本申请较佳实施例，所述停用词包括那个、这个、哎呀、比如、吧、但是等，所述标点符号包括句号、逗号、问好等。进一步地，本申请通过制作中文停用词表和中文标点符号表，通过比对所述原始文本集与所述中文停用词表和中文标点符号表是否有相同部分，达到去除所述停用词及所述标点符号的目的。如问题 D：“近期价格是否会发生变化？”经过去除所述停用词及所述标点符号的操作变为：“近期价格发生变化”。

步骤二、将所述初级文本集输入至词向量化转换模型中得到词向量文本集，并将所述词向量文本集分类成训练集和测试集。

进一步地，利用词向量化转换模型将所述初级文本集转换为词向量文本

集，所述词向量化转换模型包括输入层、投影层和输出层。

较佳地，本申请较佳实施例将所述初级文本集输入至所述输入层，所述输入层确定所述初级文本集内每个词语 ω 在所述初级文本集的出现位置 $Context(\omega)$ ，并将所述出现位置 $Context(\omega)$ 输入至所述投射层。如上述“近期价格发生变化”经过所述输入层处理后得到 $Context(\omega_1)$ 、 $Context(\omega_2)$ 、 $Context(\omega_3)$ 、 $Context(\omega_4)$ ，其中 ω_1 为近期， ω_2 为价格， ω_3 为发生， ω_4 为变化， $Context$ 为所述“近期价格发生变化”，则 $Context(\omega_1) = p(1|4)$ 、 $Context(\omega_2) = p(2|4)$ ，依次类推。

进一步地，所述投射层基于所述出现位置 $Context(\omega)$ 做累加求和操作得到累加求和矩阵 X_ω ，并根据所述 X_ω 建立概率模型。所述累加求和操作得到累加求和矩阵 X_ω 为：

$$X_\omega = \sum_{i=1}^{2^c} V(Context(\omega_i))$$

其中， $V(Context(\omega_i))$ 是所述出现位置 $Context(\omega)$ 的矩阵表示形式， c 表示所述初级文本集每个文本的词语数量，如所述“近期价格发生变化”的词语数量为4。所述概率模型为：

$$p(\omega|Context(\omega)) = \prod_{j=2}^{l^\omega} p(d_j^\omega|X_\omega, \theta_{j-1}^\omega)$$

其中， $p(\omega|Context(\omega))$ 为所述概率模型， $Context$ 为所述初级文本集， ω 为所述初级文本集内每个词语， d_j^ω 表示在路径 p^ω 内，第 j 个结点对应的 Huffman 编码 θ_{j-1}^ω 表示路径 p^ω 内第 j 个非叶子结点对应的向量。所述 Huffman 编码是根据数据通信知识用 0、1 码的不同排列来表示词语，词语称为叶子结点，各叶子结点的权值通过 Huffman 编码表现。如所述问题“近期价格发生变化”经过所述输入层处理后得到 $Context(\omega_1)$ ，其中 ω_1 为“近期”，则所述“近期”为所述叶子结点，通过所述 Huffman 编码操作后得到编码形式，所述编码形式可用 00010 表示，当所述“近期”为所述叶子结点时，所述“价格”、“发生”、“变化”为非叶子结点，所述叶子结点与所述非叶子结点统称为结点。

进一步地，所述输出层根据所述概率模型建立对数似然函数，并最大化所述对数似然函数得到词向量文本集，所述对数似然函数 ζ 为：

$$\zeta = \sum_{\omega \in \varsigma} p(\omega | \text{Context}(\omega))$$

其中， ζ 为所述对数似然函数， ς 是包含了所述初级文本集所有内容的集合，称为语料，进一步地，所述对数似然函数 ζ 基于所述概率模型可进一步扩展为：

$$\zeta = \sum_{\omega \in \varsigma} \log \prod_{j=2}^{l^\omega} \{ [\sigma(X_\omega^T \theta_{j-1}^\omega)]^{1-d_j^\omega} [1 - \sigma(X_\omega^T \theta_{j-1}^\omega)]^{d_j^\omega} \}$$

其中， l^ω 表示所述路径 p^ω 中包括结点的数量， σ 为阈值函数，所述阈值函数可用 sigmoid 函数。

较佳地，所述最大化所述对数似然函数的方法为：

$$\begin{aligned} \frac{\partial \zeta(\omega, j)}{\partial \theta_{j-1}^\omega} &= \frac{\partial}{\partial \theta_{j-1}^\omega} \{ (1 - d_j^\omega) \log [\sigma(X_\omega^T \theta_{j-1}^\omega)] + d_j^\omega \log [1 - \sigma(X_\omega^T \theta_{j-1}^\omega)] \} \\ &= (1 - d_j^\omega) [1 - \sigma(X_\omega^T \theta_{j-1}^\omega)] X_\omega - d_j^\omega \sigma(X_\omega^T \theta_{j-1}^\omega) X_\omega \\ &= [1 - d_j^\omega - \sigma(X_\omega^T \theta_{j-1}^\omega)] X_\omega \end{aligned}$$

基于所述方法最大化 $\frac{\partial \zeta(\omega, j)}{\partial \theta_{j-1}^\omega}$ 后得到最优化所述概率模型，基于所述最优化

10 概率模型得到最优化累加求和矩阵 X_ω ，进一步地，基于所述最优化累加求和矩阵 X_ω 得到词向量，因此所述初级文本集可转变为词向量文本集。

较佳地，将所述词向量文本集按照 8:2 的数量随机划分为训练集和测试集。

步骤三、将所述训练集及所述标签集输入至意图识别模型中训练，直到所述意图识别模型满足训练要求后退出训练。

15 本申请较佳实施所述意图识别模型包括卷积神经网络，所述卷积神经网络共有十六层卷积层和十六池化层、一层全连接层。当所述卷积神经网络接收所述训练集后，将所述训练集输入至第一层卷积层，所述第一层卷积层进行卷积操作后得到第一卷积数据集输入至第一层池化层；所述第一层池化层对所述第一卷积数据集进行最大化池化操作后得到第一降维数据集输入至第二层卷积层；所述第二层卷积层再次进行所述卷积操作后得到第二卷积数据集输入至第二层池化层进行所述最大化池化操作得到第二降维数据集，以此类推，直至最终得到第十六降维数据集，将所述第十六降维数据集输入至全连接层。

20

较佳地，所述全连接层接收所述第十六降维数据集，并结合激活函数计算得到所述训练值集合，并将所述训练值集合和所述标签集输入至所述意图识别模型的损失函数中，所述损失函数计算出损失值，判断所述损失值与预设训练阈值的大小关系，直至所述损失值小于所述预设训练阈值时，所述意图识别模型退出训练，所述预设训练阈值一般设置为 0.1。

本申请较佳实施例所述卷积层的卷积操作为：

$$\omega' = \frac{\omega + 2p - k}{s} + 1$$

其中 ω' 为输出数据， ω 为输入数据， k 为所述卷积层的卷积核大小， s 为所述卷积操作的步幅， p 为数据补零矩阵；

所述激活函数为：

$$y = \frac{1}{1 + e^{-\omega}}$$

其中 y 为所述训练值集合， e 为无限不循环小数。

本申请较佳实施例所述损失值 T 为：

$$T = \sum_{t=1}^n |y_t - \mu_t|^2$$

其中， n 为所述标签集的大小， y_t 为所述训练值集合， μ_t 为所述标签集。

步骤四、将所述测试集输入至所述意图识别模型中进行意图判断，判断准确率和预设准确率的大小关系。

若判断准确率小于预设准确率，返回步骤三。

较佳地，所述意图识别模型基于所述测试集进行上述卷积操作、池化操作、激活操作后得到所述测试集的意图分类集合，依次比对所述意图分类集合与所述标签集是否相同，并计算相同的数量，将所述相同的数量除以所述测试集总数得到准确率。如所述意图识别模型将测试集中“近期价格发生变化”判断为“价格变动咨询”，与实际被分类为“价格变动咨询”一致，因此所述意图识别模型判断正确。进一步地，所述准确率一般设置为 95%，当所述准确率小于 95%时，重新返回 S3 继续训练。

步骤五、若判断准确率大于预设准确率，则所述意图识别模型完成训练。

步骤六、接收用户的文本 A，并将所述文本 A 转变为词向量文本 A 输入至所述意图识别模型进行意图判断，并输出判断结果。

可选地，在其他实施例中，文本意图智能分类程序还可以被分割为一个

或者多个模块，一个或者多个模块被存储于存储器 11 中，并由一个或多个处理器（本实施例为处理器 12）所执行以完成本申请，本申请所称的模块是指能够完成特定功能的一系列计算机程序指令段，用于描述文本意图智能分类程序在文本意图智能分类装置中的执行过程。

例如，参照图 3 所示，为本申请文本意图智能分类装置一实施例中的文本意图智能分类程序的程序模块示意图，该实施例中，所述文本意图智能分类程序可以被分割为数据接收模块 10、数据处理模块 20、模型训练模块 30、文本意图智能分类输出模块 40 示例性地：

所述数据接收模块 10 用于：接收原始文本集及标签集，并对所述原始文本集去除停用词及标点符号得到初级文本集。

所述数据处理模块 20 用于：将所述初级文本集输入至词向量化转换模型中得到词向量文本集，并将所述词向量文本集分类成训练集和测试集。

所述模型训练模块 30 用于：将所述训练集及所述标签集输入至意图识别模型中训练，直到所述意图识别模型满足训练要求后退出训练，将所述测试集输入至所述意图识别模型中进行文本意图判断，计算所述文本意图判断与
所述标签集中内容是否相同并得到判断准确率，若所述判断准确率小于预设准确率则继续训练，若判断准确率大于预设准确率，所述意图识别模型完成训练。

所述文本意图智能分类输出模块 40 用于：接收用户的文本 A，并将所述文本 A 转变为词向量文本 A 输入至所述意图识别模型进行文本意图判断，并输出判断结果。

上述数据接收模块 10、数据处理模块 20、模型训练模块 30、文本意图智能分类输出模块 40 等程序模块被执行时所实现的功能或操作步骤与上述实施例大体相同，在此不再赘述。

此外，本申请实施例还提出一种计算机可读存储介质，该计算机可读存储介质可以是硬盘、多媒体卡、SD 卡、闪存卡、SMC、只读存储器（ROM）、可擦除可编程只读存储器（EPROM）、便携式紧致盘只读存储器（CD-ROM）、USB 存储器等等中的任意一种或者几种的任意组合。所述计算机可读存储介质中包括文本意图智能分类程序 10，所述文本意图智能分类程序 10 被处理器
执行时实现如下操作：

接收原始文本集及标签集，并对所述原始文本集去除停用词及标点符号得到初级文本集。

将所述初级文本集输入至词向量化转换模型中得到词向量文本集，并将所述词向量文本集分类成训练集和测试集。

5 将所述训练集及所述标签集输入至意图识别模型中训练，直到所述意图识别模型满足训练要求后退出训练，将所述测试集输入至所述意图识别模型中进行文本意图判断，计算所述文本意图判断与所述标签集中内容是否相同并得到判断准确率，若所述判断准确率小于预设准确率则继续训练，若判断准确率大于预设准确率，所述意图识别模型完成训练。

10 接收用户的文本 A，并将所述文本 A 转变为词向量文本 A 输入至所述意图识别模型进行文本意图判断，并输出判断结果。

本申请之计算机可读存储介质的具体实施方式与上述文本意图智能分类方法的具体实施方式大致相同，在此不再赘述。

上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。

15 需要说明的是，上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。并且本文中的术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、装置、物品或者方法不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、装置、物品或者方法所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、装置、物品或者方法中还存在另外的相同要素。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。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如上所述的一个存储介质(如 ROM/RAM、磁碟、光盘)中，包括若干指令用以使得一台终端设备(可以是手机，计算机，服务器，或者网络设备等)执行本申请各个实施例所述的方法。

25 以上仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接

接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

5

10

15

20

25

30

权利要求书

1、一种文本意图智能分类方法，其特征在于，所述方法包括：

步骤 A：接收原始文本集及标签集，并对所述原始文本集去除停用词及标点符号得到初级文本集；

5 步骤 B：将所述初级文本集转换为词向量文本集，并将所述词向量文本集分类成训练集和测试集；

步骤 C：将所述训练集及所述标签集输入至预先构建的意图识别模型中训练，直到所述意图识别模型满足预设训练要求后退出训练；

10 步骤 D：将所述测试集输入至所述意图识别模型中进行文本意图判断，计算对所述文本意图的判断结果与所述标签集中内容的匹配准确率，若所述匹配准确率小于预设准确率，返回步骤 C，若所述匹配准确率大于所述预设准确率，所述意图识别模型完成训练；

步骤 E：接收用户的文本，并将所述文本转变为词向量文本输入至所述意图识别模型进行文本意图判断，并输出判断结果。

15 2、如权利要求 1 所述的文本意图智能分类方法，其特征在于，所述步骤 B 包括：

利用词向量化转换模型将所述初级文本集转换为词向量文本集，所述词向量化转换模型包括输入层、投影层和输出层。

20 3、如权利要求 1 所述的文本意图智能分类方法，其特征在于，所述将所述初级文本集转换为词向量文本集包括：

确定所述初级文本集内每个词语 ω 在所述初级文本集的出现位置 $\text{Context}(\omega)$ ；

基于所述出现位置 $\text{Context}(\omega)$ 做累加求和操作得到累加求和矩阵 X_ω ，并根据所述 X_ω 建立概率模型；

25 根据所述概率模型建立对数似然函数，并最大化所述对数似然函数得到所述词向量文本集。

4、如权利要求 3 所述的文本意图智能分类方法，其特征在于，所述概率模型为：

$$p(\omega|\text{Context}(\omega)) = \prod_{j=2}^{\omega} p(d_j^\omega|X_\omega, \theta_{j-1}^\omega)$$

30 其中， $p(\omega|\text{Context}(\omega))$ 为所述概率模型， Context 为所述初级文本集， ω 为

所述初级文本集内每个词语， d_j^ω 表示在路径 p^ω 内,第j个结点对应的 Huffman 编码， θ_{j-1}^ω 表示路径 p^ω 内,第j个非叶子结点对应的向量。

5、如权利要求 4 中的文本意图智能分类方法，其特征在于，所述对数似然函数为：

$$\zeta = \sum_{\omega \in \varsigma} p(\omega | \text{Context}(\omega))$$

其中， ζ 为所述对数似然函数， ς 是包含了所述初级文本集所有内容的集合，其中，所述对数似然函数 ζ 基于所述概率模型可进一步扩展为：

$$\zeta = \sum_{\omega \in \varsigma} \log \prod_{j=2}^{l^\omega} \{ [\sigma(X_\omega^T \theta_{j-1}^\omega)]^{1-d_j^\omega} [1 - \sigma(X_\omega^T \theta_{j-1}^\omega)]^{d_j^\omega} \}$$

其中， l^ω 表示所述路径 p^ω 中包括结点的数量， σ 为阈值函数。

10 6、如权利要求 1 至 5 任意一项所述的文本意图智能分类方法，其特征在于，所述意图识别模型包括卷积神经网络、激活函数、损失函数，其中，所述卷积神经网络包括十六层卷积层和十六层池化层、一层全连接层；

所述步骤 C 包括：

15 所述卷积神经网络接收所述训练集后，将所述训练集输入至所述十六层卷积层和十六层池化层进行卷积操作和最大池化操作得到降维数据集，并将所述降维数据集输入至全连接层；

20 所述全连接层接收所述降维数据集，并结合所述激活函数计算得到训练值集合，并将所述训练值集合和所述标签集输入至所述损失函数中，所述损失函数计算出损失值，判断所述损失值与预设训练阈值的大小关系，直至所述损失值小于所述预设训练阈值时，所述意图识别模型满足所述预设训练要求并退出训练。

7、如权利要求 6 所述的文本意图智能分类方法，其特征在于，所述步骤 D 包括：

25 所述意图识别模型基于所述测试集进行卷积操作、池化操作、激活操作，得到所述测试集的意图分类集合，依次比对所述意图分类集合与所述标签集是否相同，并计算相同的数量，将所述相同的数量除以所述测试集总数得到准确率。

8、一种文本意图智能分类装置，其特征在于，所述装置包括存储器和处理器，所述存储器上存储有可在所述处理器上运行的文本意图智能分类程序，30 所述文本意图智能分类程序被所述处理器执行时实现如下步骤：

步骤 A :接收原始文本集及标签集 ,并对所述原始文本集去除停用词及标点符号得到初级文本集 ;

步骤 B :将所述初级文本集转换为词向量文本集 ,并将所述词向量文本集分类成训练集和测试集 ;

5 步骤 C :将所述训练集及所述标签集输入至预先构建的意图识别模型中训练 ,直到所述意图识别模型满足预设训练要求后退出训练 ;

步骤 D : 将所述测试集输入至所述意图识别模型中进行文本意图判断 ,计算对所述文本意图的判断结果与所述标签集中内容的匹配准确率 ,若所述匹配准确率小于预设准确率 ,返回步骤 C ,若所述匹配准确率大于所述预设准确率 ,所述意图识别模型完成训练 ;

10 步骤 E :接收用户的文本 ,并将所述文本转变为词向量文本输入至所述意图识别模型进行文本意图判断 ,并输出判断结果。

9、如权利要求 8 所述的文本意图智能分类装置 ,其特征在于 ,所述步骤 B 包括 :

15 利用词向量化转换模型将所述初级文本集转换为词向量文本集 ,所述词向量化转换模型包括输入层、投影层和输出层。

10、如权利要求 8 所述的文本意图智能分类装置 ,其特征在于 ,所述将所述初级文本集转换为词向量文本集包括 :

20 确定所述初级文本集内每个词语 ω 在所述初级文本集的出现位置 $\text{Context}(\omega)$;

基于所述出现位置 $\text{Context}(\omega)$ 做累加求和操作得到累加求和矩阵 X_ω ,并根据所述 X_ω 建立概率模型 ;

根据所述概率模型建立对数似然函数 ,并最大化所述对数似然函数得到所述词向量文本集。

25 11、如权利要求 10 所述的文本意图智能分类装置 ,其特征在于 ,所述概率模型为 :

$$p(\omega|\text{Context}(\omega)) = \prod_{j=2}^{l_\omega} p(d_j^\omega|X_\omega, \theta_{j-1}^\omega)$$

30 其中 , $p(\omega|\text{Context}(\omega))$ 为所述概率模型 , Context 为所述初级文本集 , ω 为所述初级文本集内每个词语 , d_j^ω 表示在路径 p^ω 内,第 j 个结点对应的 Huffman 编码 , θ_{j-1}^ω 表示路径 p^ω 内,第 j 个非叶子结点对应的向量。

12、如权利要求 11 中的文本意图智能分类装置，其特征在于，所述对数似然函数为：

$$\zeta = \sum_{\omega \in \zeta} p(\omega | \text{Context}(\omega))$$

其中， ζ 为所述对数似然函数， ζ 是包含了所述初级文本集所有内容的集合，
5 其中，所述对数似然函数 ζ 基于所述概率模型可进一步扩展为：

$$\zeta = \sum_{\omega \in \zeta} \log \prod_{j=2}^{l^\omega} \{ [\sigma(X_\omega^T \theta_{j-1}^\omega)]^{1-d_j^\omega} [1 - \sigma(X_\omega^T \theta_{j-1}^\omega)]^{d_j^\omega} \}$$

其中， l^ω 表示所述路径 p^ω 中包括结点的数量， σ 为阈值函数。

13、如权利要求 8 至 12 任意一项所述的文本意图智能分类装置，其特征在于，所述意图识别模型包括卷积神经网络、激活函数、损失函数，其中，
10 所述卷积神经网络包括十六层卷积层和十六层池化层、一层全连接层；

所述步骤 C 包括：

所述卷积神经网络接收所述训练集后，将所述训练集输入至所述十六层卷积层和十六层池化层进行卷积操作和最大池化操作得到降维数据集，并将所述降维数据集输入至全连接层；

15 所述全连接层接收所述降维数据集，并结合所述激活函数计算得到训练值集合，并将所述训练值集合和所述标签集输入至所述损失函数中，所述损失函数计算出损失值，判断所述损失值与预设训练阈值的大小关系，直至所述损失值小于所述预设训练阈值时，所述意图识别模型满足所述预设训练要求并退出训练。

20 14、如权利要求 13 所述的文本意图智能分类装置，其特征在于，所述步骤 D 包括：

所述意图识别模型基于所述测试集进行卷积操作、池化操作、激活操作，得到所述测试集的意图分类集合，依次比对所述意图分类集合与所述标签集是否相同，并计算相同的数量，将所述相同的数量除以所述测试集总数得到
25 准确率。

15、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储有文本意图智能分类程序，所述文本意图智能分类程序可被一个或者多个处理器执行，可实现如下步骤：

步骤 A：接收原始文本集及标签集，并对所述原始文本集去除停用词及标点符号得到初级文本集；
30

步骤 B :将所述初级文本集转换为词向量文本集 ,并将所述词向量文本集分类成训练集和测试集 ;

步骤 C :将所述训练集及所述标签集输入至预先构建的意图识别模型中训练 ,直到所述意图识别模型满足预设训练要求后退出训练 ;

5 步骤 D : 将所述测试集输入至所述意图识别模型中进行文本意图判断 ,计算对所述文本意图的判断结果与所述标签集中内容的匹配准确率 ,若所述匹配准确率小于预设准确率 ,返回步骤 C ,若所述匹配准确率大于所述预设准确率 ,所述意图识别模型完成训练 ;

10 步骤 E :接收用户的文本 ,并将所述文本转变为词向量文本输入至所述意图识别模型进行文本意图判断 ,并输出判断结果。

16、如权利要求 15 所述的计算机可读存储介质 ,其特征在于 ,所述步骤 B 包括 :

利用词向量化转换模型将所述初级文本集转换为词向量文本集 ,所述词向量化转换模型包括输入层、投影层和输出层。

15 17、如权利要求 15 所述的计算机可读存储介质 ,其特征在于 ,所述将所述初级文本集转换为词向量文本集包括 :

确定所述初级文本集内每个词语 ω 在所述初级文本集的出现位置 $\text{Context}(\omega)$;

20 基于所述出现位置 $\text{Context}(\omega)$ 做累加求和操作得到累加求和矩阵 X_ω ,并根据所述 X_ω 建立概率模型 ;

根据所述概率模型建立对数似然函数 ,并最大化所述对数似然函数得到所述词向量文本集。

18、如权利要求 17 所述的计算机可读存储介质 ,其特征在于 ,所述概率模型为 :

25
$$p(\omega|\text{Context}(\omega)) = \prod_{j=2}^{l^\omega} p(d_j^\omega|X_\omega, \theta_{j-1}^\omega)$$

其中 , $p(\omega|\text{Context}(\omega))$ 为所述概率模型 , Context 为所述初级文本集 , ω 为所述初级文本集内每个词语 , d_j^ω 表示在路径 p^ω 内,第 j 个结点对应的 Huffman 编码 , θ_{j-1}^ω 表示路径 p^ω 内,第 j 个非叶子结点对应的向量。

30 19、如权利要求 18 中的计算机可读存储介质 ,其特征在于 ,所述对数似然函数为 :

$$\zeta = \sum_{\omega \in \varsigma} p(\omega | \text{Context}(\omega))$$

其中, ζ 为所述对数似然函数, ς 是包含了所述初级文本集所有内容的集合, 其中, 所述对数似然函数 ζ 基于所述概率模型可进一步扩展为:

$$\zeta = \sum_{\omega \in \varsigma} \log \prod_{j=2}^{l^\omega} \{ [\sigma(x_\omega^T \theta_{j-1}^\omega)]^{1-d_j^\omega} [1 - \sigma(x_\omega^T \theta_{j-1}^\omega)]^{d_j^\omega} \}$$

5 其中, l^ω 表示所述路径 p^ω 中包括结点的数量, σ 为阈值函数。

20、如权利要求 15 至 19 任意一项所述的计算机可读存储介质, 其特征在于, 所述意图识别模型包括卷积神经网络、激活函数、损失函数, 其中, 所述卷积神经网络包括十六层卷积层和十六层池化层、一层全连接层;

所述步骤 C 包括:

10 所述卷积神经网络接收所述训练集后, 将所述训练集输入至所述十六层卷积层和十六层池化层进行卷积操作和最大池化操作得到降维数据集, 并将所述降维数据集输入至全连接层;

所述全连接层接收所述降维数据集, 并结合所述激活函数计算得到训练值集合, 并将所述训练值集合和所述标签集输入至所述损失函数中, 所述损失函数计算出损失值, 判断所述损失值与预设训练阈值的大小关系, 直至所述损失值小于所述预设训练阈值时, 所述意图识别模型满足所述预设训练要求并退出训练。

20

25

30

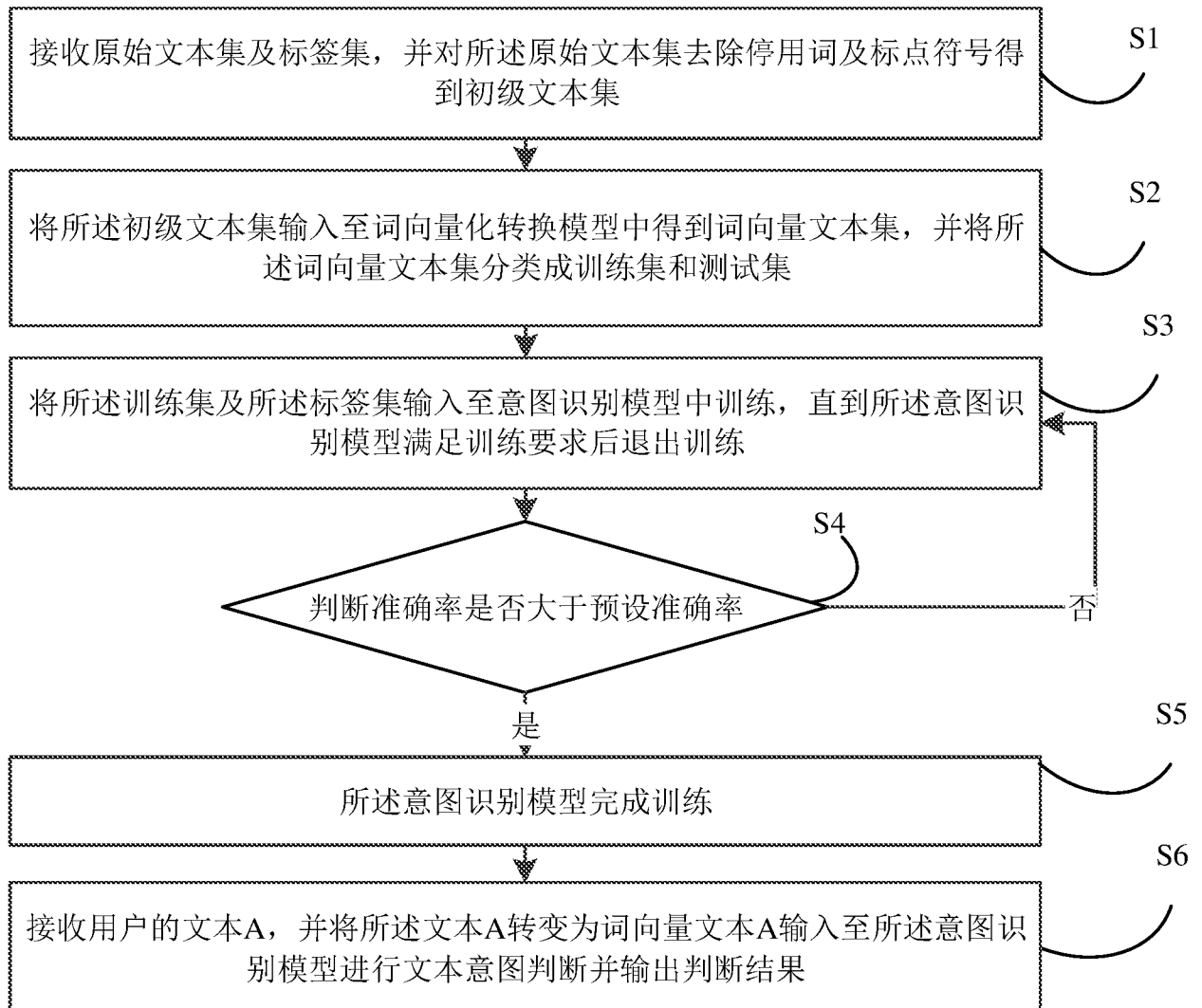


图 1

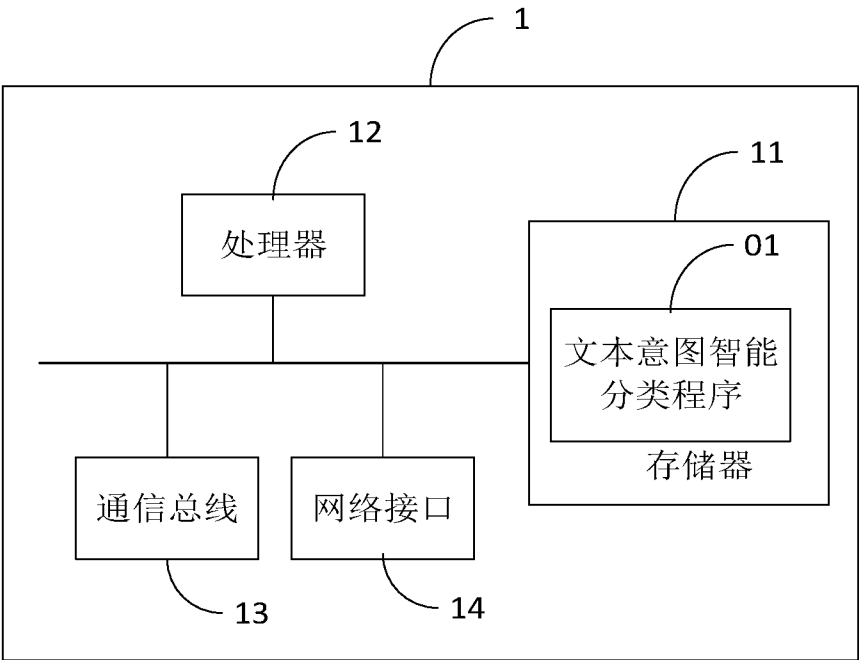


图 2

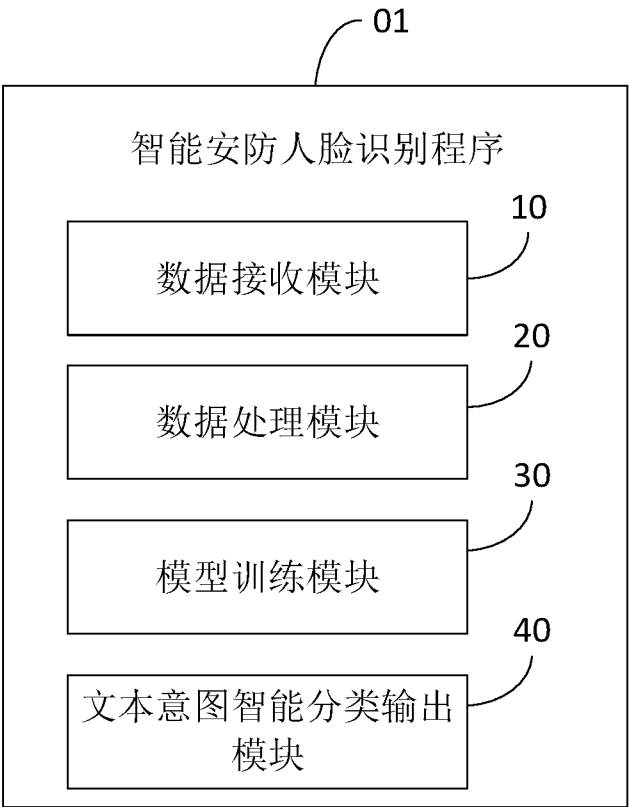


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/102207

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/33(2019.01)i; G06F 16/35(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, WPI, EPODOC: 文本, 意图, 分类, 识别, 模型, 神经网络, 训练, 测试, 词向量, 概率模型, 对数似然函数, text, intend+, classif+, model, CNN, train+, test+, label, vector

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 109325106 A (XIAMEN BIOVOICE INFORMATION TECHNOLOGY CO., LTD.) 12 February 2019 (2019-02-12) description, paragraphs [0035]-[0080], [0119]	1, 6-8, 13-15, 20
Y	CN 109325106 A (XIAMEN BIOVOICE INFORMATION TECHNOLOGY CO., LTD.) 12 February 2019 (2019-02-12) description, paragraphs [0035]-[0080], [0119]	2-5, 9-12, 16-19
X	CN 109284406 A (SICHUAN CHANGHONG ELECTRIC CO., LTD.) 29 January 2019 (2019-01-29) description, paragraphs [0036]-[0079]	1, 6-8, 13-15, 20
Y	CN 109284406 A (SICHUAN CHANGHONG ELECTRIC CO., LTD.) 29 January 2019 (2019-01-29) description, paragraphs [0036]-[0079]	2-5, 9-12, 16-19
Y	王明亚 (WANG, Mingya). "基于词向量的文本分类算法研究与改进 (Research and Improvement on Text Classification Based on Word Embedding)" 中国优秀硕士学位论文全文数据库信息科技辑 (China Master's Theses Full-Text Database, Information Science and Technology), No. 10, 15 October 2016 (2016-10-15), ISSN: 1674-0246, pp. 23-26	2-5, 9-12, 16-19



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

06 March 2020

Date of mailing of the international search report

16 March 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/102207**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 109635117 A (MOXI (BEIJING) TECHNOLOGY CO., LTD.) 16 April 2019 (2019-04-16) entire document	1-20
A	US 2008104037 A1 (INQUIRA, INC.) 01 May 2008 (2008-05-01) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/102207

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109325106	A	12 February 2019	None			
CN	109284406	A	29 January 2019	None			
CN	109635117	A	16 April 2019	None			
US	2008104037	A1	01 May 2008	US	2012131033	A1	24 May 2012

国际检索报告

国际申请号

PCT/CN2019/102207

A. 主题的分类

G06F 16/33 (2019.01) i; G06F 16/35 (2019.01) i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNKI, CNPAT, WPI, EP0D0C: 文本, 意图, 分类, 识别, 模型, 神经网络, 训练, 测试, 词向量, 概率模型, 对数似然函数, text, intend+, classif+, model, CNN, train+, test+, label, vector

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 109325106 A (厦门快商通信息技术有限公司) 2019年 2月 12日 (2019 - 02 - 12) 说明书第[0035]-[0080], [0119]段	1, 6-8, 13-15, 20
Y	CN 109325106 A (厦门快商通信息技术有限公司) 2019年 2月 12日 (2019 - 02 - 12) 说明书第[0035]-[0080], [0119]段	2-5, 9-12, 16-19
X	CN 109284406 A (四川长虹电器股份有限公司) 2019年 1月 29日 (2019 - 01 - 29) 说明书第[0036]-[0079]段	1, 6-8, 13-15, 20
Y	CN 109284406 A (四川长虹电器股份有限公司) 2019年 1月 29日 (2019 - 01 - 29) 说明书第[0036]-[0079]段	2-5, 9-12, 16-19
Y	王明亚. “基于词向量的文本分类算法研究与改进” 中国优秀硕士学位论文全文数据库 信息科技辑, 第10期, 2016年 10月 15日 (2016 - 10 - 15), ISSN: 1674-0246, 第23-26页	2-5, 9-12, 16-19
A	CN 109635117 A (零犀北京科技有限公司) 2019年 4月 16日 (2019 - 04 - 16) 全文	1-20
A	US 2008104037 A1 (INQUIRA, INC.) 2008年 5月 1日 (2008 - 05 - 01) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 3月 6日

国际检索报告邮寄日期

2020年 3月 16日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

王伟

电话号码 86-(10)-53961525

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/102207

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	109325106	A	2019年 2月 12日	无	
CN	109284406	A	2019年 1月 29日	无	
CN	109635117	A	2019年 4月 16日	无	
US	2008104037	A1	2008年 5月 1日	US 2012131033 A1	2012年 5月 24日