

(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 9 月 10 日 (10.09.2020)



(10) 国际公布号
WO 2020/177378 A1

(51) 国际专利分类号:
G06F 17/27 (2006.01)

(21) 国际申请号: PCT/CN2019/117424

(22) 国际申请日: 2019 年 11 月 12 日 (12.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910168231.6 2019年3月6日 (06.03.2019) 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: 赵峰(ZHAO, Feng); 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)。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)。

(74) 代理人: 深圳众鼎专利商标代理事务所(普通合伙)(SHENZHEN ZHONGDING INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市龙岗区龙城街道中心城清林路 546 号城投商务中心 4 层/B, Guangdong 518172 (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, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,

(54) Title: TEXT INFORMATION FEATURE EXTRACTION METHOD AND DEVICE, COMPUTER APPARATUS, AND STORAGE MEDIUM

(54) 发明名称: 文本信息的特征提取方法、装置、计算机设备及存储介质

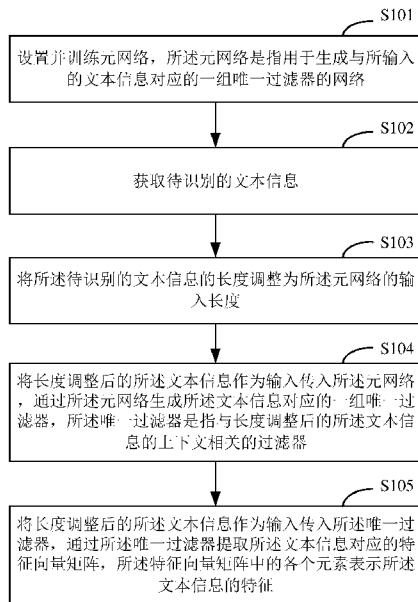


图 1

- S101 Provide and train a unit network, wherein the unit network is used to generate a group of unique filters corresponding to inputted text information
- S102 Acquire text information to undergo recognition
- S103 Adjust the length of the text information to be an input length of the unit network
- S104 Introduce the length-adjusted text information into the unit network, and generate a group of unique filters corresponding to the text information by means of the unit network, wherein a unique filter refers to a filter related to a length-adjusted text information context
- S105 Introduce the length-adjusted text information to the unique filter and extract a feature vector matrix corresponding to the text information by means of the unique filter, wherein each element in the feature vector matrix represents a feature of the text information

(57) Abstract: The present application discloses a text information feature extraction method and device, a computer apparatus, and a storage medium. The method comprises: providing and training a unit network, wherein the unit network is used to generate a group of unique filters corresponding to inputted text information; adjusting the length of text information to undergo recognition to be the unit network input length; introducing the length-adjusted text information to the unit network, and generating a group of unique filters corresponding to the text information by means of the unit network, wherein a unique filter refers to a filter related to a length-adjusted text information context; and introducing the length-adjusted text information to the unique filter, and extracting a feature vector matrix corresponding to the text information by means of the unique filter. The present application resolves the issue in which existing text

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))。

recognition techniques cannot be adapted for context, and thus, text recognition accuracy is poor.

(57) 摘要: 本申请公开了一种文本信息的特征提取方法、装置、计算机设备及存储介质, 所述方法包括: 设置并训练元网络, 所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络; 将待识别的文本信息的长度调整为所述元网络的输入长度; 将长度调整后的所述文本信息传入所述元网络, 通过所述元网络生成所述文本信息对应的一组唯一过滤器, 所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器; 将长度调整后的所述文本信息传入所述唯一过滤器, 通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵。本申请解决了现有文本识别技术无法适应上下文语境、文本识别准确率欠佳的问题。

文本信息的特征提取方法、装置、计算机设备及存储介质

本申请以 2019 年 3 月 6 日提交的申请号为 201910168231.6，名称为“文本信息的特征提取方法、装置、计算机设备及存储介质”的中国发明专利申请为基础，并要求其优先权。

技术领域

本申请涉及信息技术领域，尤其涉及一种文本信息的特征提取方法、装置、计算机设备及存储介质。

背景技术

卷积神经网络在近来逐渐成为自然语言处理的一种基础模块，尽管得到了成功，但是大部分现存的卷积神经网络都采用的是对所有输入句子应用习得的相同的静态过滤器。静态过滤器最大的不足在于它不是与文本相关的，也就是它同等地对待所有类型的文本。比如我们人在阅读一篇科普文章和一篇时政新闻时，阅读方式一般是不同的，阅读的重点也通常是不同的；对于时政新闻，我们应当主要提取时间、地点、人物、事件等信息，而对于科普文章，应该在概念、逻辑、因果等关系上给予更大的权重。而静态过滤器只能用同样的权重来对待所有的上下文信息，因而在文本识别的准确率这一方面上受到了限制。

由此可见，寻找一种能够适应上下文语境、提高文本识别准确率的方法成为本领域亟需解决的技术问题。

发明内容

本申请实施例提供了一种文本信息的特征提取方法、装置、计算机设备及存储介质，以解决现有文本识别技术无法适应上下文语境、文本识别准确率欠佳的问题。

一种文本信息的特征提取方法，包括：

设置并训练元网络，所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络；

获取待识别的文本信息；

将所述待识别的文本信息的长度调整为所述元网络的输入长度；

将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器；

将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

可选地，所述将所述待识别的文本信息的长度调整为所述元网络的输入长度包括：

获取所述元网络的输入长度，判断所述待识别的文本信息的长度是否达到所述输入长度；

若否时，将预设字符填充至所述待识别的文本信息末尾，以将所述待识别的文本信息的长度调整为所述输入长度。

可选地，所述将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网

络生成所述文本信息对应的一组唯一过滤器包括：

对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等；

通过所述元网络对所述向量矩阵执行卷积运算，得到指定长度的隐藏层向量；

对所述隐藏层向量执行转置卷积运算，得到长度调整后的所述文本信息对应的一组唯一过滤器。

可选地，所述将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵包括：

对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等；

通过所述唯一过滤器对所述向量矩阵执行卷积运算，提取所述文本信息对应的特征图；

对所述特征图执行池化操作，提取特征图中每一行的最大值作为主要特征，得到所述文本信息对应的特征向量矩阵。

可选地，在通过所述唯一过滤器生成所述文本信息对应的特征向量矩阵后，还包括：

将所述特征向量矩阵作为输入传入全连接层，然后将全连接层的输出作为输入传入预设的 Softmax 分类器；

根据所述 Softmax 分类器的输出获取所述文本信息对应的类别。

一种文本信息的特征提取装置，包括：

训练模块，用于设置并训练元网络，所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络；

信息获取模块，用于获取待识别的文本信息；

长度调整模块，用于将所述待识别的文本信息的长度调整为所述元网络的输入长度；

过滤器生成模块，用于将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器；

特征提取模块，用于将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

可选地，所述长度调整模块包括：

长度获取单元，用于获取所述元网络的输入长度，判断所述待识别的文本信息的长度是否达到所述输入长度；

长度调整单元，用于若所述待识别的文本信息的长度未达到所述输入长度时，将预设字符填充至所述待识别的文本信息末尾，以将所述待识别的文本信息的长度调整为所述输入长度。

可选地，所述过滤器生成模块包括：

第一向量化单元，用于对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等；

第一卷积单元，用于通过所述元网络对所述向量矩阵执行卷积运算，得到指定长度的隐藏层向量；

转置卷积单元，用于对所述隐藏层向量执行转置卷积运算，得到所述长度调整后的文本信息对应的一组唯一过滤器。

一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现如下步骤：

设置并训练元网络，所述元网络是指用于生成与所输入的文本信息对应的一组唯一

过滤器的网络；

获取待识别的文本信息；

将所述待识别的文本信息的长度调整为所述元网络的输入长度；

将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器；

将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

一个或多个存储有计算机可读指令的非易失性可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行如下步骤：

设置并训练元网络，所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络；

获取待识别的文本信息；

将所述待识别的文本信息的长度调整为所述元网络的输入长度；

将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器；

将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

本申请的一个或多个实施例的细节在下面的附图和描述中提出，本申请的其他特征和优点将从说明书、附图以及权利要求变得明显。

附图说明

为了更清楚地说明本申请实施例的技术方案，下面将对本申请实施例的描述中需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图1是本申请一实施例中文本信息的特征提取方法的一流程图；

图2是本申请一实施例中文本信息的特征提取方法中步骤S103的一流程图；

图3是本申请一实施例中文本信息的特征提取方法中步骤S104的一流程图；

图4是本申请一实施例中文本信息的特征提取方法中步骤S105的一流程图；

图5是本申请一实施例中文本信息的特征提取方法的一流程图；

图6是本申请一实施例中文本信息的特征提取装置的一原理框图；

图7是本申请一实施例中计算机设备的一示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作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。

本申请实施例提供的文本信息的特征提取方法应用于服务器。所述服务器可以用独立的服务器或者是多个服务器组成的服务器集群来实现。在一实施例中，如图1所示，提

供一种文本信息的特征提取方法，包括如下步骤：

在步骤 S101 中，设置并训练元网络，所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络。

在这里，为了解决大部分现存的卷积神经网络采用对所有待识别的文本信息应用习得的相同的静态过滤器的问题，本申请实施例提出了一种使用元网络来学习上下文语境的卷积神经网络，并应用于文本处理。其中，所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络，能够根据所输入的文本信息生成与所输入的文本信息的上下文语境相关的权重矩阵，是一个关于网络的网络。元网络所生成的过滤器是针对不同类型的文本信息定制的，适用于该不同类型的文本信息，从而改变了以往的卷积神经网络中使用同一过滤器同等对待所有类型的文本的现状。由于所述过滤器与文本信息的上下文语境相关，使得提取到的特征可以更准确。

在本申请实施例中，所述元网络可以是任意可微分的深度网络。所述元网络预先通过在大量的训练文本集上训练得到。在进行文本信息的特征提取时，获取预先训练得到的元网络，并对所述元网络的参数进行微调，得到适用于本次文本信息的元网络。

在步骤 S102 中，获取待识别的文本信息。

在本申请实施例中，所述文本信息是根据指定的语言衔接和语义连贯规则组成的一个语句，包括但不限于话术文本信息、文学文本信息。服务器可以根据实际需要或者应用场景的需要获取待识别的文本信息。例如，服务器从预设数据库中获取待识别的文本信息，所述预设数据库中预先收集了大量的文本信息。或者，服务器通过客户端的麦克风获取用户输入的语音信息，然后将所述语音信息转换为文字，得到待识别的文本信息。又或者，服务器通过客户端的拍照功能获取图像信息，然后对所述图像信息进行 OCR 文本识别，得到待识别的文本信息。可以理解的是，服务器还可以通过多种方式获取到待识别的文本信息，此处不再过多赘述。

为了提高元网络输出结果的价值，提高文本特征提取的准确率，所述待识别的文本信息优选为与所述元网络的训练文本及上下文分布是相同或相似的，即所述待识别的文本信息与所述训练文本相同或者为训练文本中的一部分，两者在风格、类型、语义等层面上是相同或相似的。

在步骤 S103 中，将所述待识别的文本信息的长度调整为所述元网络的输入长度。

在这里，所述文本信息的长度是指文本信息的字符串长度。为了使得元网络根据不同的输入文本生成的唯一过滤器的大小是统一的，本申请实施例在将所述待识别的文本信息输入元网络之前，调整所述待识别的文本信息的长度为元网络的输入长度。所述输入长度是预先设置的元网络的输入参数的字符串长度。可选地，可以在训练元网络时，预先统计所有训练文本的字符串长度的分布，然后选取一个最大长度作为元网络的输入长度，以进行统一。如图 2 所示，所述步骤 S103 将所述待识别的文本信息的长度调整为所述元网络的输入长度包括：

在步骤 S1031 中，获取所述元网络的输入长度，判断所述待识别的文本信息的长度是否达到所述输入长度。

在这里，本申请实施例通过计算所述待识别的文本信息的字符个数，得到所述待识别的文本信息的字符串长度。然后将所述字符串长度与所述输入长度进行比较，以判断所述待识别的文本信息的长度是否达到所述输入长度。

若是时，即所述待识别的文本信息的长度达到所述输入长度，则跳转至步骤 S104，通过所述元网络生成所述待识别的文本信息对应的一组唯一过滤器。

在步骤 S1032 中，若否时，将预设字符填充至所述待识别的文本信息末尾，以将所述待识别的文本信息的长度调整为所述输入长度。然后跳转至步骤 S104。

在这里，对于字符串长度未达到输入长度的待识别的文本信息，本申请实施例采用预

设字符补齐所述待识别的文本信息，以将所述待识别的文本信息调整到输入长度。所述预设字符对于元网络、卷积神经网络来说是一个代表空白的特殊字符，比如 NUL。

为了便于理解，以下对上述步骤 S103 调整文本信息的长度为所述元网络的输入长度进行举例说明。假设预先统计所有训练文本的字符串长度的分布，然后选取一个最大长度作为所述输入长度，比如 7。若待识别的文本信息为“今天天气真好”，通过步骤 S1031 得到其字符串长度为 6，未达到输入长度 7。则在步骤 S1033 中采用预设字符 NUL 来对所述待识别的文本信息“今天天气真好”进行补齐，得到长度调整后的文本信息“今天天气真好 NUL”。

在步骤 S104 中，将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器。

如前所述，所述元网络能够根据输入参数学习得到经过调整的卷积神经网络过滤器。在本申请实施例中，所述输入参数为长度调整后的待识别的文本信息，所述过滤器为所述待识别的文本信息对应的一组唯一过滤器。所述唯一过滤器与待识别的文本信息的上下文相关，从而使用所述唯一过滤器能够对不同的待识别的文本信息精炼及提取到不同的特征。

可选地，为了解决输入长度可变的问题，在本申请实施例中，所述元网络生成预定义大小的唯一过滤器。如图 3 所示，所述步骤 S104 将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器包括：

在步骤 S1041 中，对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等。

在这里，本申请实施例对长度调整后的所述文本信息进行向量化处理，得到长度调整后的文本信息对应的向量矩阵，作为所述待识别的文本信息对应的向量矩阵。所述向量矩阵中包括若干个词嵌入向量。所述词嵌入向量是指将长度调整后的所述文本信息进行分词后每个单词的词向量，即长度调整后的文本信息中的每一个词映射为向量矩阵中的一个列向量。在本申请实施例中，所述词嵌入向量的长度是预先指定的，即对于不同长度的待识别的文本信息，其对应的词嵌入向量的长度都是相同的。尽管长度调整时采用了预设字符来填充所述待识别的文本信息，但是所述预设字符对于元网络、卷积神经网络来说是一个代表空白的特殊字符。本申请实施例将长度调整后的所述文本信息转化为向量矩阵，有利于方便了后续卷积神经网络的识别和学习，即方便了后续执行卷积运算和转置卷积运算。

在步骤 S1042 中，通过所述元网络对所述向量矩阵执行卷积运算，得到指定长度的隐藏层向量。

在得到长度调整后的所述文本信息对应的向量矩阵之后，通过元网络中预设的卷积层对所述向量矩阵执行卷积操作，即通过卷积层滤波器对所述向量矩阵进行卷积运算，计算滤波器和所述向量矩阵之间的点积，以提取更高层次的特征，得到指定长度的隐藏层向量。在这里，组成卷积层滤波器的参数可以通过损失函数进行优化。

在步骤 S1043 中，对所述隐藏层向量执行转置卷积运算，得到所述长度调整后的文本信息对应的一组唯一过滤器。

在这里，转置卷积(transpose convolution)运算，又称为解卷积(deconvolution)或者反卷积，类似于卷积的逆运算。本申请实施例在步骤 S1042 所述的隐藏层之上叠加了一个转置卷积层。在得到隐藏层向量之后，将所述隐藏层向量通过所述转置卷积层，进行转置卷积运算，生成一组卷积核，以所述卷积核作为长度调整后的所述文本信息对应的一组唯一过滤器，即所述待识别的文本信息对应的一组唯一过滤器。在这里，组成转置卷积层的参数可以通过损失函数进行优化。可以理解的是，所述唯一过滤器是与所述待识别的文本信息的上下文相关的过滤器，为所述待识别的文本信息定制的，适用于所述待识别的文本信

息。

在本申请实施例中，由于待识别的文本信息通过步骤 S103 调整到指定的输入长度，然后封装到一个相同长度的向量矩阵中，隐藏层向量根据所述向量矩阵得到，从而使得所述隐藏层向量与所述待识别的文本信息长度无关，保证了通过元网络生成的过滤器对每一个待识别的文本信息具有相同的维度和大小，即通过元网络生成的过滤器的大小保持一致。

可选地，在本申请实施例中，上述卷积层和转置卷积层中的参数均为联合可微分的，因此在训练元网络时可以将卷积层的参数和转置卷积层的参数一起通过梯度的反向传播算法进行优化、更新。在这里，反向传播算法（即 BP 算法）的思想是将通过卷积层和转置卷积层的输出进行误差计算，并将误差反向逐级传下去，主要由激励传播和权重更新两个环节反复循环迭代，直至训练文本的特征向量矩阵达到预定的误差期望值。通过反向传播算法可以进一步优化元网络的参数，提高元网络生成唯一过滤器的准确度。

在步骤 S105 中，将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

在通过元网络得到长度调整后的所述文本信息对应的唯一过滤器之后，使用所述唯一过滤器对长度调整后的所述文本信息进行识别。具体为：将长度调整后的所述文本信息作为输入传入所述唯一过滤器，然后获取经过所述唯一过滤器后的输出，以所述输出作为所述待识别的文本信息对应的特征向量矩阵。所述特征向量矩阵中包含所述待识别的文本信息的特征信息，即语义信息。

可选地，如图 4 所示，所述步骤 S105 将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵包括：

在步骤 S1051 中，对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等。

在这里，在将长度调整后的文本信息传入所述唯一过滤器之前，可以对长度调整后的所述文本信息进行向量化处理，将文本信息中的每一个词映射为向量矩阵中的一个列向量，得到每一个词的词嵌入向量，组合所述词嵌入向量得到长度调整后的所述文本信息对应的向量矩阵。其中，所述词嵌入向量的长度是预先指定的，即对于不同长度的待识别的文本信息，其对应的词嵌入向量的长度都是相同的。本申请实施例将所述待识别的文本信息转化为向量矩阵，有利于方便后续执行卷积运算。

示例性地，假设长度调整后的文本信息的长度为 T ，组成的单词为 $x_1, x_2, \dots, x_T$ 。对长度调整后的所述文本信息进行向量化处理后，得到一个向量矩阵 $X \in R^{d \times T}$ ，向量矩阵 X 中的每一列表示长度调整后的所述文本信息中的一个单词对应的 d 维度的词嵌入向量。

在步骤 S1052 中，通过所述唯一过滤器对所述向量矩阵执行卷积运算，提取所述文本信息对应的特征图。

在得到长度调整后的所述文本信息对应的向量矩阵之后，将所述向量矩阵作为输入传入所述唯一过滤器执行卷积操作，即通过所述唯一过滤器对所述向量矩阵进行卷积运算，计算过滤器和所述向量矩阵之间的点积，以提取更高层次的特征，得到所述文本信息对应的特征图。

示例性地，假设所述唯一过滤器的权重为 $W \in R^{K \times h \times d}$ ，将所述唯一过滤器和所述向量矩阵中的每个大小为 h 的窗口进行卷积运算，得到所述向量矩阵的特征图 P 。其中，所述特征图 P 中的每一个元素 p_i 由窗口大小为 h 的文本片段生成： $p_i = f(W \times x_{i:i+h-1} + b)$ 。

在上式中, $i=1,2,\dots,T-h+1$, 在这里, $\times$ 表示卷积运算符, b 表示维度为 K 的偏置向量, f 表示非线性激活函数, 比如 ReLU。

在步骤 S1053 中, 对所述特征图执行池化操作, 提取特征图中每一行的最大值作为主要特征, 得到所述文本信息对应的特征向量矩阵。

在本申请实施例中, 所述特征图随后作为输入传入最大池化层中, 通过所述最大池化层对所述特征图中的每一行提取最大值, 得到主要特征, 组合所有主要特征得到一个 K 维向量, 以所述 K 维向量作为长度调整后的所述文本信息对应的特征向量矩阵, 即所述待识别的文本信息对应的特征向量矩阵。本申请实施例通过最大池化层弃置不重要的特征, 仅保留了最突出的特征, 一方面可以使特征图变小, 简化计算复杂度, 一方面可以提高识别的准确度。

所述特征向量矩阵中的各个元素表示所述文本信息的特征, 即语义信息。在本申请实施例中, 经过元网络生成的唯一过滤器与所述待识别的文本信息的上下文相关, 不同的待识别的文本信息对应的唯一过滤器不相同, 即在卷积神经网络中的权重矩阵不相同。通过所述唯一过滤器获取所述待识别的文本信息的特征向量矩阵, 大大地提高了特征提取的准确率。

综上所述, 本申请实施例通过设置并训练元网络, 所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络; 在对文本信息进行识别时, 根据待识别的文本信息获取元网络, 并将所述待识别的文本信息的长度调整为所述元网络的输入长度; 然后将长度调整后的所述文本信息作为输入传入所述元网络, 通过所述元网络生成所述文本信息对应的一组唯一过滤器, 所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器; 将长度调整后的所述文本信息作为输入传入所述唯一过滤器, 通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵, 所述特征向量矩阵中的各个元素表示所述文本信息的特征; 从而实现了针对待识别的文本信息习得对应的一组过滤器用于识别所述文本信息, 解决了现有文本识别技术无法适应上下文语境、文本识别准确率欠佳的问题, 极大地提高了文本信息的特征提取的准确率。

在本申请实施例中, 所述特征向量矩阵与待识别的文本信息对应, 包括若干个元素, 每一个元素表示从所述文本信息中提取出来的特征, 即语义信息。相比于待识别的文本信息, 所述特征向量矩阵的维度大大缩小。基于所述特征向量矩阵可以进一步实现对所述待识别文本信息的分类。如图 5 所示, 所述方法还可以包括:

在步骤 S106 中, 将所述特征向量矩阵作为输入传入全连接层, 然后将全连接层的输出作为输入传入预设的 Softmax 分类器。

在步骤 S107 中, 根据所述 Softmax 分类器的输出获取所述文本信息对应的类别。

在这里, 本申请实施例通过全连接层对所述特征向量矩阵进行提纯, 以将特征向量矩阵转化为指定维度的向量, 方便后续 softmax 分类器执行分类操作。所述全连接层预先根据分类类别的数量 N , 设置 $K*N$ 个权重系数和 N 个偏置值, K 为全连接层前一层的最后一维的维度, 即输出的特征向量矩阵的维度。然后将所述特征向量矩阵与全连接层的权重矩阵相乘后加上一个偏置值, 所得的和组合成一维向量, 从而得到全连接层的输出。

将所述全连接层的输出作为输入传入预设的 softmax 分类器。在这里, 所述 Softmax 分类器用于处理多分类问题, 其输出需要经过 Softmax 函数进行数值处理。关于 Softmax 函数的定义如下:

$$S_n = \frac{e^{V_n}}{\sum_n^N e^{V_n}}$$

在上述中， V_n 表示全连接层输出的一维向量中的元素， n 表示类别索引， $n=1,2,3,\dots,N$ ，总的类别个数为 N 。 S_n 表示当前元素 V_n 的指数与所有元素指数和的比值。将所有的 S_n 组合成一维向量，得到 softmax 分类器的输出。通过上式可知，softmax 分类器将多分类的全连接层输出数值转化为相对概率，其元素表征了不同类别之间的相对概率，便于理解和比较。基于所述 softmax 分类器的输出，概率最大的元素对应的类别可能性最大，可以清晰地预测所述待识别的文本信息为概率最大的元素对应的类别。

可选地，在本申请实施例中，所述类别可以是意图类别，比如同意、拒绝、等待等，也可以是网页类别、情感类别、用户评论类别等等，此处不作限制。

在本申请实施例中，经过元网络生成的唯一过滤器与所述待识别的文本信息的上下文相关，不同的待识别的文本信息对应的唯一过滤器不相同，即权重矩阵不相同。通过所述唯一过滤器获取所述待识别的文本信息的特征向量矩阵，大大地提高了特征提取的准确率；基于所述特征向量矩阵进行分类，进一步提高了分类的准确率。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不对本申请实施例的实施过程构成任何限定。

在一实施例中，提供一种文本信息的特征提取装置，该中文本信息的特征提取装置与上述实施例中文本信息的特征提取方法一一对应。如图 6 所示，该文本信息的特征提取装置包括训练模块 61、信息获取模块 62、长度调整模块 63、过滤器生成模块 64、特征提取模块 65。各功能模块详细说明如下：

训练模块 61，用于设置并训练元网络，所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络；

信息获取模块 62，用于获取待识别的文本信息；

长度调整模块 63，用于将所述待识别的文本信息的长度调整为所述元网络的输入长度；

过滤器生成模块 64，用于将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器；

特征提取模块 65，用于将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

可选地，所述长度调整模块 63 包括：

长度获取单元，用于获取所述元网络的输入长度，判断所述待识别的文本信息的长度是否达到所述输入长度；

长度调整单元，用于若所述待识别的文本信息的长度未达到所述输入长度时，将预设字符填充至所述待识别的文本信息末尾，以将所述待识别的文本信息的长度调整为所述输入长度。

可选地，所述过滤器生成模块 64 包括：

第一向量化单元，用于对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，

所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等；

第一卷积单元，用于通过所述元网络对所述向量矩阵执行卷积运算，得到指定长度的隐藏层向量；

转置卷积单元，用于对所述隐藏层向量执行转置卷积运算，得到所述长度调整后的文本信息对应的一组唯一过滤器。

可选地，所述特征提取模块 65 包括：

第二向量化单元，用于对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等；

第二卷积单元，用于通过所述唯一过滤器对所述向量矩阵执行卷积运算，提取所述文本信息对应的特征图；

池化单元，用于对所述特征图执行池化操作，提取特征图中每一行的最大值作为主要特征，得到所述文本信息对应的特征向量矩阵。

可选地，在通过所述唯一过滤器生成所述文本信息对应的特征向量矩阵后，还包括：

分类模块，用于将所述特征向量矩阵作为输入传入全连接层，然后将全连接层的输出作为输入传入预设的 Softmax 分类器；根据所述 Softmax 分类器的输出获取所述文本信息对应的类别。

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

在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 7 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种文本信息的特征提取方法。

在一个实施例中，提供了一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机可读指令，处理器执行计算机可读指令时实现以下步骤：

设置并训练元网络，所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络；

获取待识别的文本信息；

将所述待识别的文本信息的长度调整为所述元网络的输入长度；

将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器；

将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

在一个实施例中，提供了一个或多个存储有计算机可读指令的非易失性可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行如下步骤：

设置并训练元网络，所述元网络是指用于生成与所输入的文本信息对应的一组唯一过

滤器的网络；

获取待识别的文本信息；

将所述待识别的文本信息的长度调整为所述元网络的输入长度；

将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器；

将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

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

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，仅以上述各功能单元、模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能单元、模块完成，即将所述装置的内部结构划分成不同的功能单元或模块，以完成以上描述的全部或者部分功能。

以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

权 利 要 求 书

1. 一种文本信息的特征提取方法，其特征在于，包括：

设置并训练元网络，所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络；

获取待识别的文本信息；

将所述待识别的文本信息的长度调整为所述元网络的输入长度；

将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器；

将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

2. 如权利要求1所述的文本信息的特征提取方法，其特征在于，所述将所述待识别的文本信息的长度调整为所述元网络的输入长度包括：

获取所述元网络的输入长度，判断所述待识别的文本信息的长度是否达到所述输入长度；

若否时，将预设字符填充至所述待识别的文本信息末尾，以将所述待识别的文本信息的长度调整为所述输入长度。

3. 如权利要求1或2所述的文本信息的特征提取方法，其特征在于，所述将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器包括：

对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等；

通过所述元网络对所述向量矩阵执行卷积运算，得到指定长度的隐藏层向量；

对所述隐藏层向量执行转置卷积运算，得到长度调整后的所述文本信息对应的一组唯一过滤器。

4. 如权利要求1或2所述的文本信息的特征提取方法，其特征在于，所述将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵包括：

对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，所述向量矩阵中包括

若干个词嵌入向量，每一个词嵌入向量的长度相等；

通过所述唯一过滤器对所述向量矩阵执行卷积运算，提取所述文本信息对应的特征图；

对所述特征图执行池化操作，提取特征图中每一行的最大值作为主要特征，得到所述文本信息对应的特征向量矩阵。

5. 如权利要求 1 或 2 所述的文本信息的特征提取方法，其特征在于，在通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵后，还包括：

将所述特征向量矩阵作为输入传入全连接层，然后将全连接层的输出作为输入传入预设的 Softmax 分类器；

根据所述 Softmax 分类器的输出获取所述文本信息对应的类别。

6. 一种文本信息的特征提取装置，其特征在于，包括：

训练模块，用于设置并训练元网络，所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络；

信息获取模块，用于获取待识别的文本信息；

长度调整模块，用于将所述待识别的文本信息的长度调整为所述元网络的输入长度；

过滤器生成模块，用于将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器；

特征提取模块，用于将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

7. 如权利要求 6 所述的文本信息的特征提取装置，其特征在于，所述长度调整模块包括：

长度获取单元，用于获取所述元网络的输入长度，判断所述待识别的文本信息的长度是否达到所述输入长度；

长度调整单元，用于若所述待识别的文本信息的长度未达到所述输入长度时，将预设字符填充至所述待识别的文本信息末尾，以将所述待识别的文本信息的长度调整为所述输入长度。

8. 如权利要求 6 或 7 所述的文本信息的特征提取装置，其特征在于，所述过滤器生成模块包括：

第一向量化单元，用于对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，

所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等；

第一卷积单元，用于通过所述元网络对所述向量矩阵执行卷积运算，得到指定长度的隐藏层向量；

转置卷积单元，用于对所述隐藏层向量执行转置卷积运算，得到所述长度调整后的文本信息对应的一组唯一过滤器。

9. 如权利要求 6 或 7 所述的文本信息的特征提取装置，其特征在于，所述特征提取模块包括：

第二向量化单元，用于对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等；

第二卷积单元，用于通过所述唯一过滤器对所述向量矩阵执行卷积运算，提取所述文本信息对应的特征图；

池化单元，用于对所述特征图执行池化操作，提取特征图中每一行的最大值作为主要特征，得到所述文本信息对应的特征向量矩阵。

10. 如权利要求 6 或 7 所述的文本信息的特征提取装置，其特征在于，所述特征提取装置还包括：

分类模块，用于将所述特征向量矩阵作为输入传入全连接层，然后将全连接层的输出作为输入传入预设的 Softmax 分类器；根据所述 Softmax 分类器的输出获取所述文本信息对应的类别。

11. 一种计算机设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现如下步骤：

设置并训练元网络，所述元网络是指用于生成与所输入的文本信息对应的一组唯一过滤器的网络；

获取待识别的文本信息；

将所述待识别的文本信息的长度调整为所述元网络的输入长度；

将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器；

将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

12. 如权利要求 11 所述的计算机设备, 其特征在于, 所述将所述待识别的文本信息的长度调整为所述元网络的输入长度包括:

获取所述元网络的输入长度, 判断所述待识别的文本信息的长度是否达到所述输入长度;

若否时, 将预设字符填充至所述待识别的文本信息末尾, 以将所述待识别的文本信息的长度调整为所述输入长度。

13. 如权利要求 11 或 12 所述的计算机设备, 其特征在于, 所述将长度调整后的所述文本信息作为输入传入所述元网络, 通过所述元网络生成所述文本信息对应的一组唯一过滤器包括:

对长度调整后的所述文本信息进行向量化处理, 得到向量矩阵, 所述向量矩阵中包括若干个词嵌入向量, 每一个词嵌入向量的长度相等;

通过所述元网络对所述向量矩阵执行卷积运算, 得到指定长度的隐藏层向量;

对所述隐藏层向量执行转置卷积运算, 得到长度调整后的所述文本信息对应的一组唯一过滤器。

14. 如权利要求 11 或 12 所述的计算机设备, 其特征在于, 所述将长度调整后的所述文本信息作为输入传入所述唯一过滤器, 通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵包括:

对长度调整后的所述文本信息进行向量化处理, 得到向量矩阵, 所述向量矩阵中包括若干个词嵌入向量, 每一个词嵌入向量的长度相等;

通过所述唯一过滤器对所述向量矩阵执行卷积运算, 提取所述文本信息对应的特征图;

对所述特征图执行池化操作, 提取特征图中每一行的最大值作为主要特征, 得到所述文本信息对应的特征向量矩阵。

15. 如权利要求 11 或 12 所述的计算机设备, 其特征在于, 所述处理器执行所述计算机可读指令时还实现如下步骤:

将所述特征向量矩阵作为输入传入全连接层, 然后将全连接层的输出作为输入传入预设的 Softmax 分类器;

根据所述 Softmax 分类器的输出获取所述文本信息对应的类别。

16. 一个或多个存储有计算机可读指令的非易失性可读存储介质, 所述计算机可读指令被一个或多个处理器执行时, 使得所述一个或多个处理器执行如下步骤:

设置并训练元网络, 所述元网络是指用于生成与所输入的文本信息对应的一组唯一过

滤器的网络；

获取待识别的文本信息；

将所述待识别的文本信息的长度调整为所述元网络的输入长度；

将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器，所述唯一过滤器是指与长度调整后的所述文本信息的上下文相关的过滤器；

将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵，所述特征向量矩阵中的各个元素表示所述文本信息的特征。

17. 如权利要求 16 所述的非易失性可读存储介质，其特征在于，所述将所述待识别的文本信息的长度调整为所述元网络的输入长度包括：

获取所述元网络的输入长度，判断所述待识别的文本信息的长度是否达到所述输入长度；

若否时，将预设字符填充至所述待识别的文本信息末尾，以将所述待识别的文本信息的长度调整为所述输入长度。

18. 如权利要求 16 或 17 所述的非易失性可读存储介质，其特征在于，所述将长度调整后的所述文本信息作为输入传入所述元网络，通过所述元网络生成所述文本信息对应的一组唯一过滤器包括：

对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等；

通过所述元网络对所述向量矩阵执行卷积运算，得到指定长度的隐藏层向量；

对所述隐藏层向量执行转置卷积运算，得到长度调整后的所述文本信息对应的一组唯一过滤器。

19. 如权利要求 16 或 17 所述的非易失性可读存储介质，其特征在于，所述将长度调整后的所述文本信息作为输入传入所述唯一过滤器，通过所述唯一过滤器提取所述文本信息对应的特征向量矩阵包括：

对长度调整后的所述文本信息进行向量化处理，得到向量矩阵，所述向量矩阵中包括若干个词嵌入向量，每一个词嵌入向量的长度相等；

通过所述唯一过滤器对所述向量矩阵执行卷积运算，提取所述文本信息对应的特征图；

对所述特征图执行池化操作，提取特征图中每一行的最大值作为主要特征，得到所述

文本信息对应的特征向量矩阵。

20. 如权利要求 16 或 17 所述的非易失性可读存储介质，其特征在于，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器还执行如下步骤：

将所述特征向量矩阵作为输入传入全连接层，然后将全连接层的输出作为输入传入预设的 Softmax 分类器；

根据所述 Softmax 分类器的输出获取所述文本信息对应的类别。

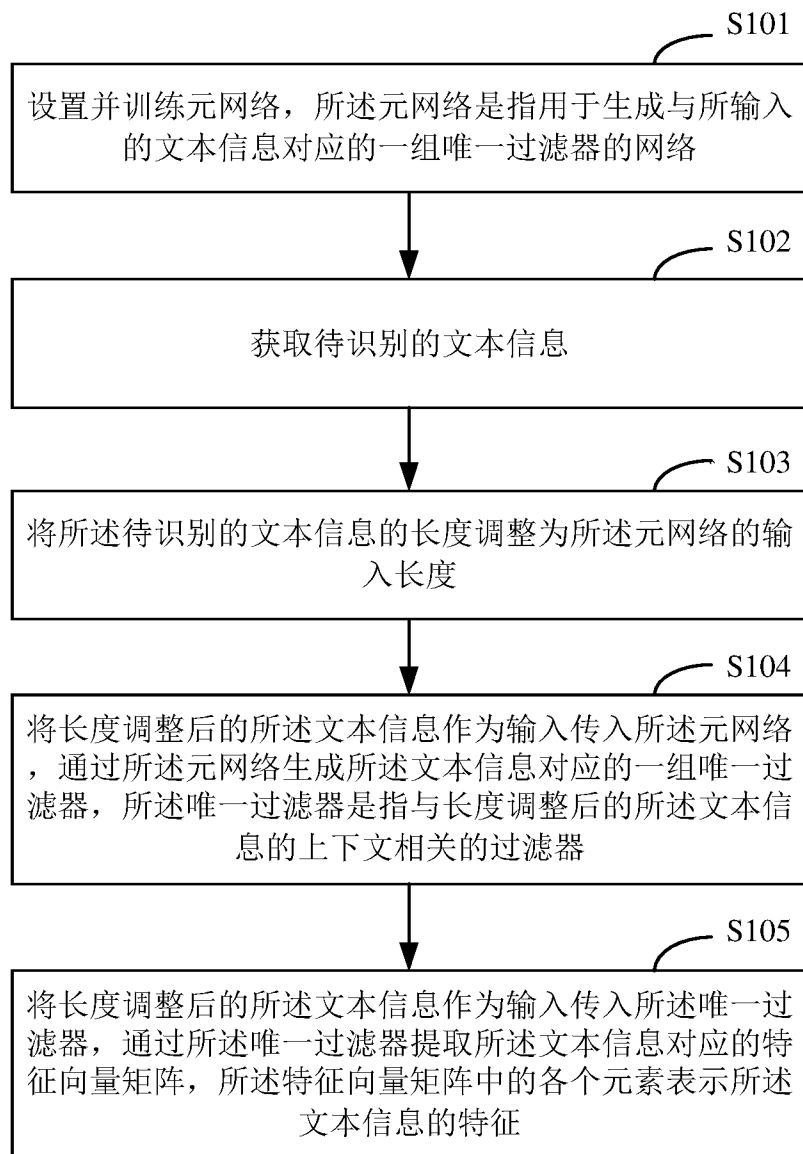


图 1

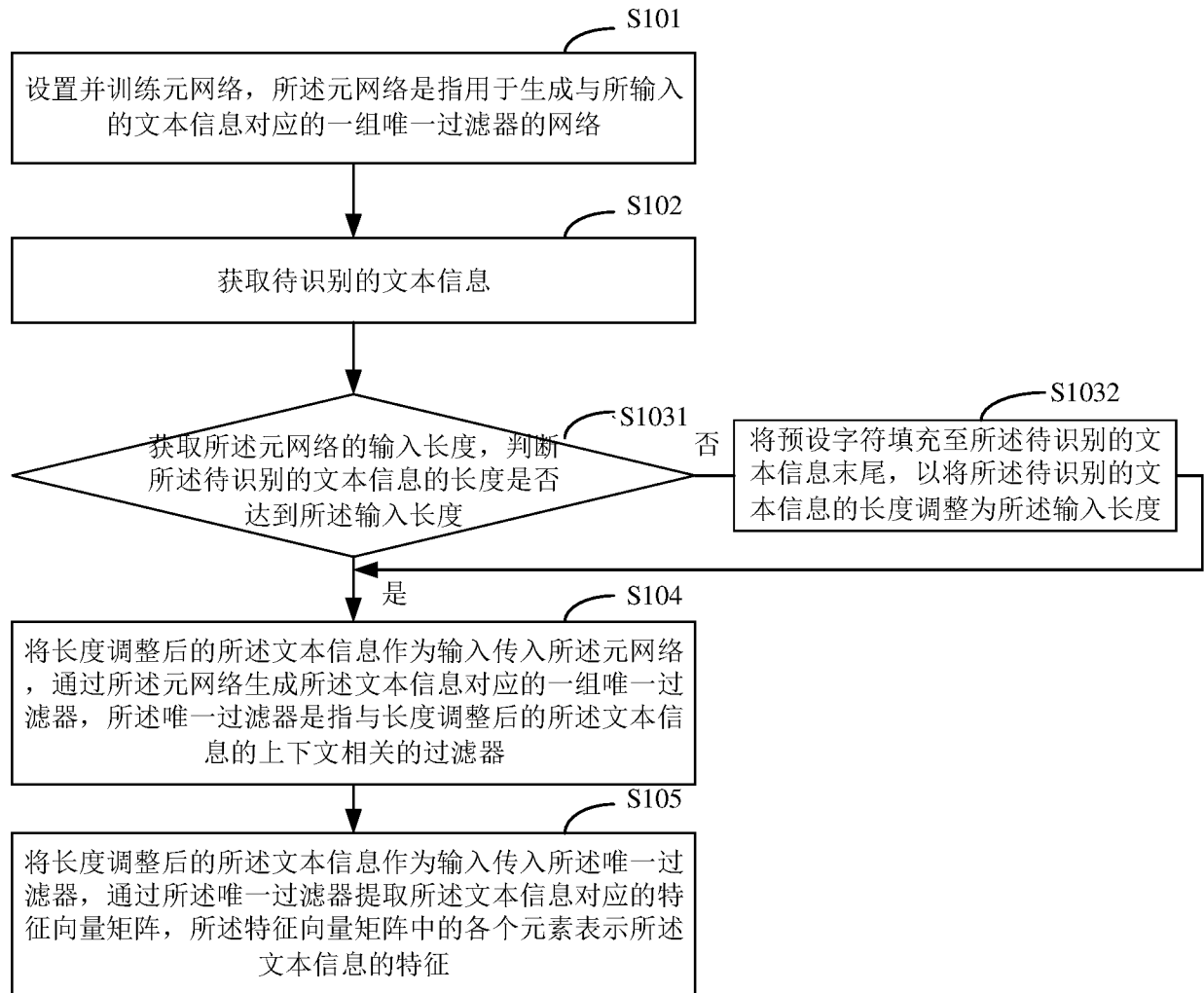


图 2

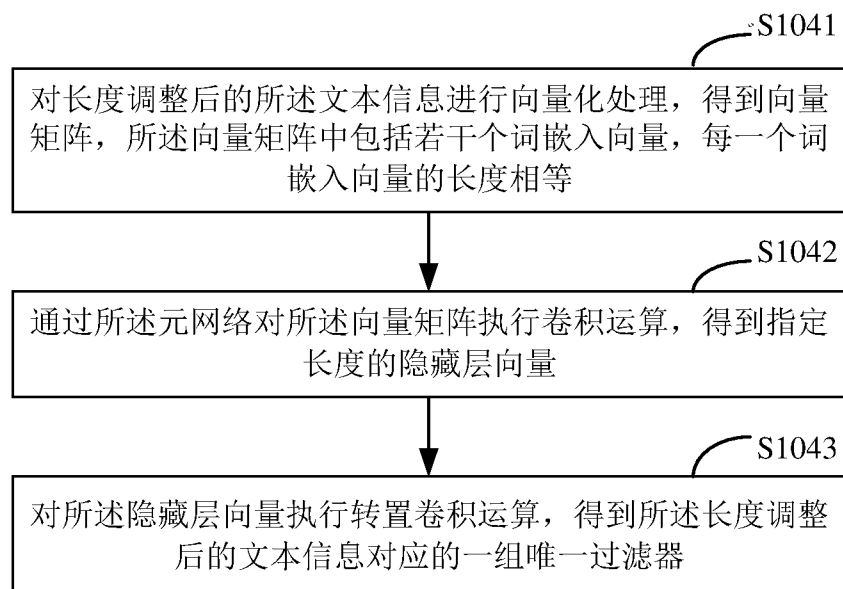


图 3

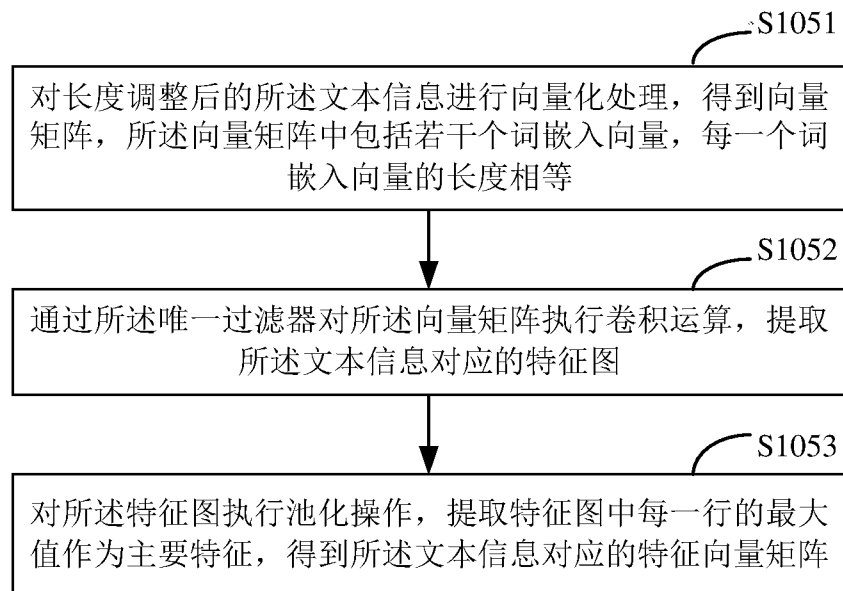


图 4

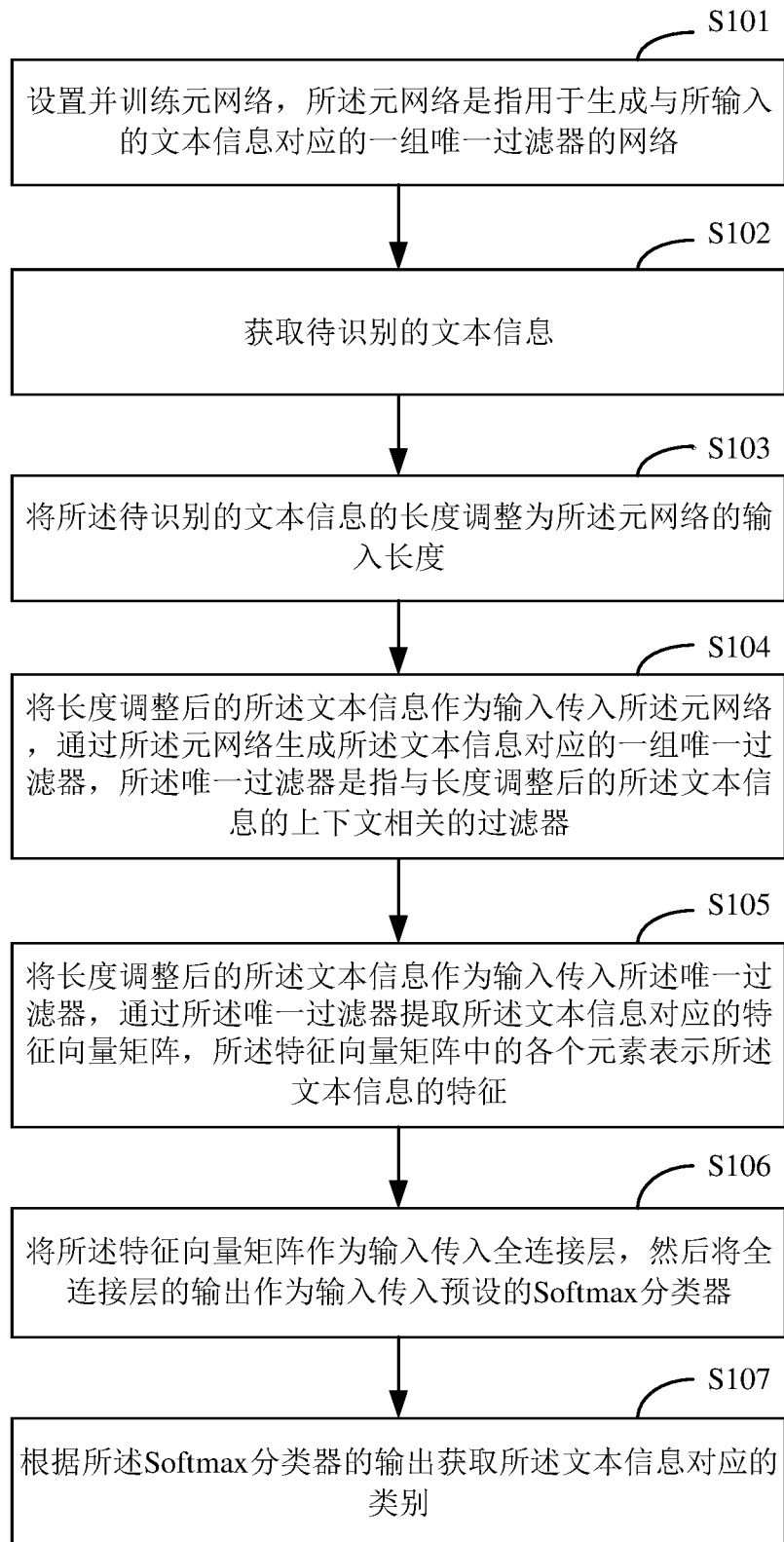


图 5

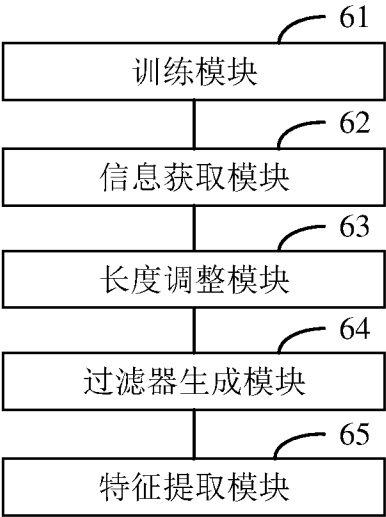


图 6

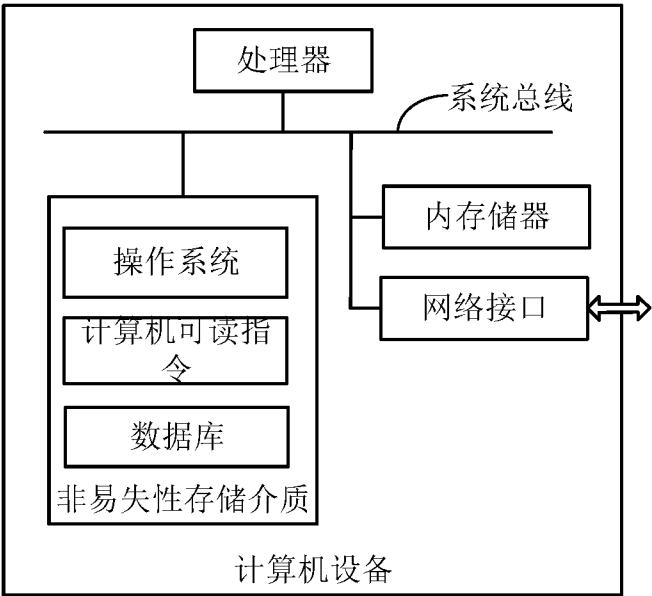


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/117424

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/27(2006.01)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, GOOGLE: 文本, 特征, 提取, 元网络, 过滤器, 滤波器, 卷积神经网络, 长度, 尺寸, 上下文, text, context, feature, CNN, character+, filter, length, element

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110020431 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 16 July 2019 (2019-07-16) claims 1-10, and description, paragraphs [0100]-[0132]	1-20
A	吴琼等 (WU, Qiong et al.). "多尺度卷积循环神经网络的情感分类技术 (Sentiment Classification with Multiscale Convolutional Recurrent Neural Network)" 华侨大学学报 (自然科学版) (<i>Journal of Huaqiao University (Natural Science)</i>), Vol. 38, No. 6, 20 November 2017 (2017-11-20), pp. 876 and 877	1-20
A	CN 108763319 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES et al.) 06 November 2018 (2018-11-06) entire document	1-20
A	CN 107797985 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 13 March 2018 (2018-03-13) entire document	1-20
A	US 2018373947 A1 (STRADVISION, INC.) 27 December 2018 (2018-12-27) entire document	1-20

☐ 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

30 December 2019

Date of mailing of the international search report

23 January 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

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/117424

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110020431	A	16 July 2019	None			
CN	108763319	A	06 November 2018	None			
CN	107797985	A	13 March 2018	US	2019095429	A1	28 March 2019
US	2018373947	A1	27 December 2018	None			

国际检索报告

国际申请号

PCT/CN2019/117424

A. 主题的分类

G06F 17/27(2006.01)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, EPODOC, GOOGLE: 文本, 特征, 提取, 元网络, 过滤器, 滤波器, 卷积神经网络, 长度, 尺寸, 上下文, text, context, feature, CNN, character+, filter, length, element

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
PX	CN 110020431 A (平安科技深圳有限公司) 2019年 7月 16日 (2019 - 07 - 16) 权利要求1-10, 说明书第0100-0132段	1-20
A	吴琼等. “多尺度卷积循环神经网络的情感分类技术” 华侨大学学报(自然科学版), 第38卷, 第6期, 2017年 11月 20日 (2017 - 11 - 20), 第876-877页	1-20
A	CN 108763319 A (中国科学院自动化研究所等) 2018年 11月 6日 (2018 - 11 - 06) 全文	1-20
A	CN 107797985 A (百度在线网络技术北京有限公司) 2018年 3月 13日 (2018 - 03 - 13) 全文	1-20
A	US 2018373947 A1 (STRADVISION, INC.) 2018年 12月 27日 (2018 - 12 - 27) 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

2019年 12月 30日

国际检索报告邮寄日期

2020年 1月 23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

授权官员

柴华

电话号码 86-(10)-53961630

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/117424

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	110020431	A	2019年 7月 16日	无			
CN	108763319	A	2018年 11月 6日	无			
CN	107797985	A	2018年 3月 13日	US	2019095429	A1	2019年 3月 28日
US	2018373947	A1	2018年 12月 27日	无			