

(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
2018 年 12 月 27 日 (27.12.2018)



(10) 国际公布号
WO 2018/233647 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01) **G06F 17/27** (2006.01)
- (21) 国际申请号: PCT/CN2018/092078
- (22) 国际申请日: 2018 年 6 月 21 日 (21.06.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201710481703.4 2017年6月22日 (22.06.2017) CN
- (71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。
- (72) 发明人: 孔行 (KONG, Hang); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

- (74) 代理人: 北京德琦知识产权代理有限公司 (DEQI INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区知春路 1 号学院国际大厦 7 层, Beijing 100083 (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, 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,

(54) **Title:** ABSTRACT GENERATION METHOD, DEVICE AND COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 摘要生成方法、装置、计算机设备以及存储介质

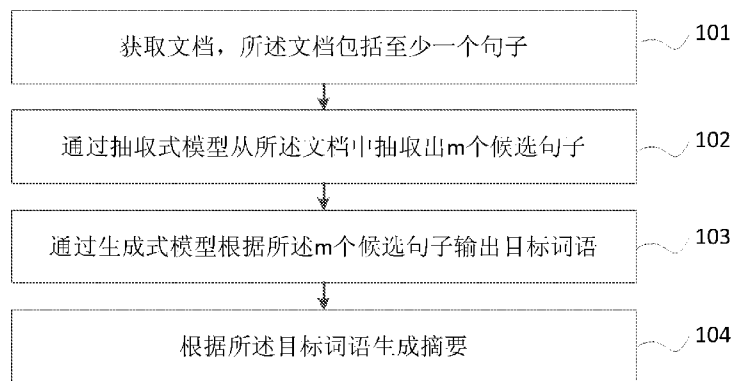


图 1

101 Acquire a document, the document comprising at least one sentence
102 Extract a number m of candidate sentences from the document by using an extractive model
103 Output target words according to the m candidate sentences by using a generative model
104 Generate an abstract according to the target words

(57) **Abstract:** An abstract generation method, device and computer device, relating to the technical field of natural language processing. The method comprises: acquiring a document, the document comprising at least one sentence (101); extracting a number m of candidate sentences from the document by using an extractive model (102); outputting target words according to the m candidate sentences by using a generative model (103); and generating an abstract according to the target words (104).

(57) **摘要:** 一种摘要生成方法、装置及计算机设备, 属于自然语言处理领域。所述方法包括: 获取文档, 所述文档包括至少一个句子 (101); 通过抽取式模型从所述文档中抽取 m 个候选句子 (102); 通过生成式模型根据所述 m 个候选句子输出目标词语 (103); 根据所述目标词语生成摘要 (104)。

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))。

摘要生成方法、装置、计算机设备以及存储介质

本申请要求于 2017 年 06 月 22 日提交中国专利局、申请号为 201710481703.4、名称为“摘要生成方法、装置及计算机设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本申请实施例涉及自然语言处理领域，特别涉及一种摘要生成方法、装置、计算机设备以及存储介质。

背景

自动文本摘要（Automatic Text Summarization）用于将一篇文档
10 进行提炼，生成简洁、流畅且包含文章主旨的摘要。自动文本摘要是自然语言处理领域的一大难题。

相关技术中提供了一种基于生成式（Abstractive）模型的自动文本摘要技术，该生成式模型用于从一篇文档中的每个句子中提取出词语，然后将提取出的词语重新组合为一个个句子，从而形成摘要。

15 技术内容

本申请实施例提供了一种摘要生成方法所述方法包括：

获取文档，所述文档包括至少一个句子；

通过抽取式模型从所述文档中抽取出 m 个候选句子；

通过生成式模型根据所述 m 个候选句子输出目标词语，根据所述
20 目标词语生成摘要。

本申请实施例提供了一种摘要生成装置，所述装置包括：存储器、

处理器；其中，所述存储器中存储有计算机可读指令，所述处理器执行所述存储中的计算机可读指令，用于：

获取文档，所述文档包括至少一个句子，每个句子包括至少一个词语；

5 通过抽取式模型从所述文档中抽取出 m 个候选句子；

通过生成式模型根据所述 m 个候选句子输出目标词语，根据所述目标词语生成摘要。

在一些实施例中，所述处理器进一步执行所述计算机可读指令，用于：

10 计算所述文档中每个句子的句子级编码向量；

根据所述每个句子的句子级编码向量计算所述文档的文档级编码向量；

根据所述句子级编码向量和所述文档级编码向量计算所述文档中的每个句子的概率 P ，将所述文档中所述概率 P 大于预设阈值的句子抽取为所述 m 个候选句子，每个所述句子的概率 P 的取值为 1 或 0。

在一些实施例中，所述处理器进一步执行所述计算机可读指令，用于：

将所述句子中的词语沿前向传播方向输入第一双向循环神经网络，得到第一前向编码向量；

20 将所述句子中的词语沿后向传播方向输入所述第一双向循环神经网络，得到第一后向编码向量；

根据所述第一前向编码向量和所述第一后向编码向量，得到所述句子的句子级编码向量。

在一些实施例中，所述处理器进一步执行所述计算机可读指令，
25 用于：将 n 个句子的句子级编码向量沿前向传播方向输入第二双向循

环神经网络，得到第二前向编码向量，其中， n 为所述文档中的句子的个数；

将所述 n 个句子的句子级编码向量沿后向传播方向输入所述第二双向循环神经网络，得到第二后向编码向量，其中， n 为所述文档中的句子的个数；

根据所述第二前向编码向量和所述第二后向编码向量，得到所述文档的文档级编码向量。

在一些实施例中，所述处理器进一步执行所述计算机可读指令，用于：

10 将所述 m 个候选句子的句子级编码向量作为输入序列输入所述生成式模型，计算第 i 个解码时刻的输出概率， i 为正整数；

当所述输出概率大于所述预设阈值时，从所述文档对应的词典中生成一个词语，作为所述摘要中的第 i 个词语；所述词典中包括所述文档中出现频率高于预设条件的词语；

15 当所述输出概率小于所述预设阈值时，从所述 m 个候选句子中的目标句子中抽取出目标词语，作为所述摘要中的第 i 个词语。

在一些实施例中，所述处理器进一步执行所述计算机可读指令，用于：

通过所述编码器根据所述 m 个候选句子的句子级编码向量计算第 i 个解码时刻的上下文向量 c_i ；

将第 $i-1$ 个解码时刻的隐含状态、所述上下文向量 c_i 和所述摘要中的第 $i-1$ 个词语输入所述解码器，计算得到第 i 个解码时刻的隐含状态，所述解码时刻的隐含状态用于表示所述解码器在当前解码时刻下已经处理的部分输入序列；

25 根据所述摘要中的第 $i-1$ 个词语、所述上下文向量 c_i 和所述第 i

个解码时刻的隐含状态，计算第 i 个解码时刻的输出概率，其中， i 为正整数，且当 $i-1=0$ 时，所述第 $i-1$ 个解码时刻的隐含状态取默认值，所述第 $i-1$ 个词语取默认值。

5 在一些实施例中，所述处理器进一步执行所述计算机可读指令，用于::

通过第一最大似然估计函数从所述 m 个候选句子中确定出所述第 i 个解码时刻下的目标句子；

通过第二最大似然估计函数从所述目标句子的词语中确定出目标词语。

10 本申请实施例提供了一种计算机设备，所述设备包括处理器和存储器，所述存储器中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行本申请实施例所述的摘要生成方法。

本申请实施例提供了一种计算机可读存储介质，所述存储介质中
15 存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行本申请实施例所述的摘要生成方法。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例
20 描述中所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 是本申请一个示例性实施例提供的摘要生成方法的流程图；

图 2a 至图 2c 是本申请另一个示例性实施例提供的摘要生成方法

的流程图；

图 3a 至图 3c 是本申请另一个示例性实施例提供的摘要生成方法的流程图；

图 4 是本申请另一个示例性实施例提供的摘要生成方法的流程图；

5 图 5 是本申请另一个示例性实施例提供的摘要生成方法的流程图；

图 6 是本申请一个示例性实施例提供的摘要生成装置的框图；

图 7 是本申请一个示例性实施例提供的计算机设备的框图。

实施方式

下面为使本申请的目的、技术方案和优点更加清楚，下面将结合
10 附图对本申请实施方式作进一步地详细描述。

图 1 是本申请一个示例性实施例提供的摘要生成方法的流程图。
该摘要生成方法由计算机设备执行，包括：

在步骤 101 中，获取文档 D，该文档 D 中包括至少一个句子。

在步骤 102 中，通过抽取式模型从文档 D 中抽取出 m 个候选句
15 子。

可选地，该抽取式模型包括基于注意力机制 (Attention) 的模型。

基于注意力机制的模型用于计算文档 D 中的每个句子的概率值，
将概率值大于预设阈值的句子抽取为候选句子，将概率小于预设阈值
的句子抛弃。其中，抽取式模型用于从文档 D 中抽取出适合作为摘要
20 的 m 个候选句子。

在步骤 103 中，通过生成式模型根据 m 个候选句子输出目标词语，
根据目标词语生成摘要。

可选地，该生成式模型包括基于指针生成网络 (Pointer-Generator
Networks) 的模型，生成式模型中包括编码器与解码器，其中，编码

器对应应有编码时刻，解码器对应应有解码时刻。

通过该生成式模型对每个解码时刻的输出概率进行计算，将该输出概率与预设阈值进行比较，若输出概率大于预设阈值，则将该输出概率标注为 1，并从文档 D 对应的词典中生成一个词语进行输出；若
5 输出概率小于预设阈值，则将该输出概率标注为 0，并从 m 个候选句子中抽取目标词语进行输出。可选地，该词典由文档 D 中出现频率最高的 s 个词语组成，s 为正整数。

在步骤 104 中，根据目标词语生成摘要。

综上所述，本申请实施例中先通过抽取式模型从文档 D 中抽取
10 适合作为摘要的 m 个候选句子，从而减少生成式模型所需要处理的文本序列长度，再通过生成式模型根据候选句子生成或提取出目标词语，根据目标词语合成文档 D 的摘要，提高了最终生成的摘要的可读性和信息量。

在一些实施例中，抽取式模型是基于注意力机制的模型，抽取式
15 模型通过双向循环神经网络对文档 D 中的每个句子进行编码，得到每个句子的句子级编码向量，再根据每个句子的句子级编码向量从文档 D 中抽取 m 个适合作为摘要的候选句子，基于图 1 所示的摘要生成方法的流程图，步骤 102 可替代实现为步骤 102a 至步骤 102d，如图 2a 所示：

20 在步骤 102a 中，计算文档 D 中每个句子的句子级编码向量。

可选地，该句子级编码向量可由双向循环神经网络或单向循环神经网络计算得到，本申请实施例对此不加以限定。

以每个句子的句子级编码向量由双向循环神经网络计算为例，本步骤可以包括三个步骤，如图 2b 所示：

25 11，将文档 D 中的每个句子中的词语沿前向传播方向输入第一双

向循环神经网络得到第一前向编码向量;

12, 将文档 D 中的每个句子中的词语沿后向传播方向输入第一双向循环神经网络得到第一后向编码向量;

13, 根据第一前向编码向量和第一后向编码向量得到句子的句子级编码向量。

可选地, 针对步骤 11, 设单个句子= $\{w_1, w_2, w_3 \dots w_t\}$, w_i 是表示句子中的第 i 个词语, 第一双向循环神经网络为 RNN;

将句子中每个词语按照由前到后的顺序沿前向传播方向输入 RNN, 采用如下计算公式计算:

$$Sh_i^f = RNN (Sh_{i-1}^f, w_i)$$

其中, Sh_i^f 为句子中第 i 个词语沿前向传播方向输入 RNN 后的编码结果, Sh_{i-1}^f 为句子中第 $i-1$ 个词语沿前向传播方向输入 RNN 后的编码结果, w_i 为句子中第 i 个的词语, RNN 为第一双向循环神经网络, 其中, i 为正整数, 当 $i-1=0$ 时, 第 $i-1$ 个词语取空值或默认值;

也即, 将句子中第 $i-1$ 个词语的词语级编码结果与第 i 个词语输入第一双向循环神经网络得到第 i 个词语的词语级编码结果, 其中, 第 t 个词语输入 RNN 后的编码结果为该句子的第一前向编码向量, t 为该句子中的词语个数。

可选地, 针对步骤 12, 将句子中每个词语按照由后到前的顺序沿后向传播方向输入 RNN, 采用如下计算公式计算:

$$Sh_i^b = RNN (Sh_{i+1}^b, w_i)$$

其中, Sh_i^b 为句子中第 i 个词语沿后向传播方向输入 RNN 的编码结果, Sh_{i+1}^b 为句子中第 $i+1$ 个词语沿后向传播方向输入 RNN 的编码结果, w_i 为句子中第 i 个的词语, RNN 为第一双向循环神经网络;

也即, 将句子中第 $i+1$ 个词语的词语级编码结果与第 i 个词语输

入第一双向循环神经网络得到第 i 个词语的词语级编码结果，其中第 1 个词语输入 RNN 后的编码结果为该句子的第一后向编码向量。

可选地，需要对文档 D 中的每个句子执行步骤 11 和步骤 12。

可选地，针对步骤 13，采用如下计算公式进行计算：

$$S_i = [Sh_{it}^f, Sh_{il}^b]$$

其中， Sh_{it}^f 为第 i 个句子中第 t 个词语（最后一个）沿前向传播方向输入 RNN 后得到的第一前向编码向量， t 为该句子中的词语个数， Sh_{il}^b 为第 i 个句子中第一个词语沿后向传播方向输入 RNN 后得到的第一后向编码向量， S_i 为第 i 个句子的句子级编码向量。

10 在步骤 102b 中，根据每个句子的句子级编码向量计算文档 D 的文档级编码向量。

可选地，该文档级编码向量可由双向循环神经网络或单向循环神经网络计算得到，本申请对此不加以限定。

以该文档级编码向量由双向循环神经网络计算为例，本步骤可以
15 包括三个步骤，如图 2c 所示：

21，将 n 个句子的句子级编码向量沿前向传播方向输入第二双向循环神经网络，得到第二前向编码向量，其中， n 为所述文档中的句子的个数；

22，将 n 个句子的句子级编码向量沿后向传播方向输入第二双向
20 循环神经网络，得到第二后向编码向量；

23，根据第二前向编码向量和第二后向编码向量得到文档 D 的文档级编码向量。

可选地，针对步骤 21，第二双向循环神经网络为 RNN，采用如下计算公式进行计算：

$$h_i^f = \text{RNN}(h_{i-1}^f, S_i)$$

其中, h_i^f 为文档 D 中第 i 个句子沿前向传播方向输入 RNN 后的编码结果, h_{i-1}^f 为文档 D 中第 i-1 个句子沿前向传播方向输入 RNN 后的编码结果, S_i 为第 i 个句子的句子级编码向量。

- 5 可选地, 针对步骤 22, 第二双向循环神经网络为 RNN, 采用如下计算公式进行计算:

$$h_i^b = \text{RNN}(h_{i+1}^b, S_i)$$

- 其中, h_i^b 为文档 D 中第 i 个句子沿后向传播方向输入 RNN 后的编码结果, h_{i+1}^b 为文档 D 中第 i+1 个句子沿前向传播方向输入 RNN 后的编码结果, S_i 为第 i 个句子的句子级编码向量。
- 10

可选地, 针对步骤 23, 采用如下计算公式进行计算:

$$h_i = [h_i^f, h_i^b]$$

其中, h_i^f 为第 i 个句子的第二前向编码向量, h_i^b 为第 i 个句子的第二后向编码向量, h_i 泛指文档 D 中的各个句子。

- 15 在步骤 102c 中, 根据句子级的编码向量和文档级的编码向量计算文档中的每个句子的概率 p。

可选地, 该抽取式模型包括基于注意力机制的模型。

该基于注意力机制的模型用于根据每个句子的句子级编码向量以及文档级编码向量计算每个句子的概率值。

- 20 计算公式如下:

$$P(S_i = 1|D) = \sigma(a(S_i, h_i))$$

其中, S_i 为文档 D 中第 i 个句子的句子级编码向量, h_i 为所述文档 D 的文档级编码向量, a 为预设的前馈神经网络, σ 为预设的第一非线性函数, $p(S_i = 1|D)$ 代表在第 i 个句子的概率大于预设阈值时被标

注为 1，并抽取该第 i 个句子；在第 i 个句子的概率小于预设阈值时被标注为 0 并抛弃该第 i 个句子。

在步骤 102d 中，将文档 D 中概率 p 大于预设阈值的句子抽取为 m 个候选句子。

5 综上所述，本申请实施例中通过基于注意力机制的模型从文档 D 中抽取出适合作为摘要的 m 个候选句子，从而将文档 D 中不被注意的无用句子进行丢弃，从而减少输入至生成式模型的文本序列中的噪声，提高生成式模型生成或提取出目标词语的信息量，保证了生成的摘要符合预期结果，进而提高最终生成的摘要的可读性。

10 在一些实施例中，生成式模型是基于指针生成网络的模型，该生成模型包括编码器及解码器，解码器中包括指针生成网络，生成式模型通过 m 个候选句子的句子级编码向量计算第 i 个解码时刻的上下文向量，并得到该第 i 个解码时刻的隐含状态，根据第 i 个解码时刻的隐含状态来输出摘要中第 i 个词语。具体地，基于图 1 所示的摘要生成方法的流程图，步骤 103 可替代实现为如下步骤 310 至步骤 312，
15 如图 3a 所示：

在步骤 310 中，将 m 个候选句子的句子级编码向量作为输入序列输入生成式模型，计算第 i 个解码时刻的输出概率。

其中，计算第 i 个解码时刻的输出概率包括如下步骤，如图 3b
20 所示：

31，通过编码器根据 m 个候选句子的句子级编码向量计算第 i 个解码时刻的上下文向量 c_i ；

32，将第 $i-1$ 个解码时刻的隐含状态、上下文向量 c_i 和摘要中的第 $i-1$ 个词语输入指针生成网络，计算得到第 i 个解码时刻的隐含状态，其中， i 为正整数，当 $i-1=0$ 时，第 $i-1$ 个解码时刻的隐含状态取
25

默认值。默认值可以是全部初始化为 0。

33, 根据摘要中的第 $i-1$ 个词语、上下文向量和第 i 个解码时刻的隐含状态, 计算第 i 个解码时刻的输出概率。

可选地, 针对步骤 31, 采用如下计算公式进行计算:

$$\begin{aligned}
 e_{ij} &= a(h'_{i-1}, h_j) \\
 \alpha_{ij} &= \frac{\exp(s_{ij})}{\sum_{j=1}^T \exp(s_{ij})} \\
 c_i &= \sum_{j=1}^T \alpha_{ij} h_j
 \end{aligned}$$

其中, α_{ij} 为解码器在第 i 个解码时刻的权重值, h_j 表示编码器在第 j 个编码时刻的隐含状态, h'_{i-1} 表示第 $i-1$ 个解码时刻的隐含状态, T 表示输入序列的输入长度, a 表示预设的前馈神经网络。解码器在第 $i-1$ 个解码时刻的隐含状态是解码器在第 $i-1$ 个解码时刻已经处理的部分输入序列; 编码器在第 j 个编码时刻的隐含状态是编码器在第 j 个解码时刻已经处理的部分输入序列。

将解码器在第 $i-1$ 个解码时刻的隐含状态以及编码器在第 j 个编码时刻的隐含状态输入预设的前馈神经网络 a 得到一个输出值, 通过该输出值计算得到解码器在第 i 个解码时刻的权重值, 通过解码器在第 i 个解码时刻的权重值与编码器在第 j 个编码时刻的隐含状态计算得到第 i 个解码时刻的输出概率。

可选地, 针对步骤 32, 采用如下计算公式进行计算:

$$h'_i = RNN(h'_{i-1}, c_i, y_{i-1})$$

其中, h'_i 为解码器在第 i 个解码时刻的隐含状态, h'_{i-1} 为解码器在第 $i-1$ 个解码时刻的隐含状态, c_i 为上下文向量, y_{i-1} 为摘要中的第 $i-1$ 个词语, RNN 为预设的指针生成网络。

可选的, 针对步骤 33, 采用如下计算公式进行计算:

$$p(G=1|D)=\sigma(h'_i, c_i, y_{i-1})$$

其中, h'_i 为解码器在第 i 个解码时刻的隐含状态、 c_i 为上下文向量, y_{i-1} 为摘要中的第 $i-1$ 个词语, σ 表示预设的第二非线性函数。

在步骤 311 中, 当输出概率大于预设阈值时, 将该输出概率标注
5 为 1, 并从文档 D 对应的词典中生成一个词语作为摘要的第 i 个词语。

可选地, 该词典由文档 D 中出现频率最高的 s 个词语组成, s 为正整数。

在步骤 312 中, 当输出概率小于预设阈值时, 将该输出概率标注
为 0, 并从 m 个候选句子中的目标句子中抽取出目标词语作为摘要的
10 第 i 个词语。

抽取目标句子和目标词语的方式包括通过基于注意力机制的模型进行抽取。可选地, 本步骤包括如下步骤, 如图 3c 所示:

41, 通过第一最大似然估计函数从 m 个候选句子中确定第 i 个解码时刻下的目标句子。

15 计算公式如下:

$$p_{ij} = \operatorname{argmax}(e_{ij})$$

其中, P_{ij} 是 m 个候选句子在第 i 个解码时刻下针对第一最大似然估计函数的最优解, e_{ij} 的计算公式如步骤 31 所示。

42, 通过第二最大似然估计函数从目标句子中的词语中确定出目
20 标词语。

计算公式如下:

$$p_{ijk} = \operatorname{argmax}(e_{ijk})$$

其中, P_{ij} 是目标句子中的各个词语在第 i 个解码时刻下针对第二最大似然估计函数的最优解, e_{ijk} 的计算公式如下所示:

$$e_{ijk} = a(h'_{i-1}, sh_{jk})$$

其中， h'_{i-1} 代表第 $i-1$ 个解码时刻的隐含状态， a 代表预设的前馈神经网络， sh_{jk} 代表目标句子中第 k 个词语在第一双向循环神经网络中的编码结果。

- 5 可选地，第一最大似然估计函数和第二最大似然估计函数是基于注意力机制所确定的函数。

综上所述，本申请实施例中先通过抽取式模型从文档 D 中抽取适合作为摘要的 m 个候选句子，再通过生成式模型根据候选句子生成或提取出目标词语，该生成式模型中采用了前馈神经网络给每个编码
10 时刻的隐含状态进行加权，使得编码器输出的上下文向量更适合作为当前编码时刻时选择目标词语时的背景。

本申请实施例中还通过第一最大似然估计函数和第二最大似然估计函数，在解码过程中使用注意力机制选择出更为合理的目标词语，提高生成式模型生成或提取出目标词语的准确性，保证了生成的摘要
15 符合预期结果，进而提高最终生成的摘要的可读性和信息量。

在一个具体的例子中，结合图 4 和图 5 对本申请实施例提供的摘要生成方法进行说明：

在步骤 301 中，获取文档 D ，该文档 D 中包括 6 个句子，第三个句子中包括至少 4 个词语 $w1$ 、 $w2$ 、 $w3$ 、 $w4$ 。

- 20 在步骤 302 至步骤 303 中，如图 4 指针所示，将文档 D 中的每个句子中的词语沿前向传播方向输入第一双向循环神经网络得到第一前向编码向量；将文档 D 中的每个句子中的词语沿后向传播方向输入第一双向循环神经网络得到第一后向编码向量。

在步骤 304 中，根据第一前向编码向量和第一后向编码向量得到

句子级编码向量 s_1 至 s_6 。

在步骤 305 至步骤 307 中，将 6 个句子的句子级编码向量沿前向传播方向输入第二双向循环神经网络，得到第二前向编码向量；将 6 个句子的句子级编码向量沿后向传播方向输入第二双向循环神经网络，得到第二后向编码向量；根据第二前向编码向量和第二后向编码向量得到文档 D 的文档级编码向量 h_1 至 h_6 。

在步骤 308 中，将句子级编码向量和文档级编码向量输入基于注意力机制的模型计算文档中的每个句子的概率 p 。

其中，第二句、第四句以及第五句的概率 p 大于预设阈值，被标注为 1；第一句、第三句以及第六句的概率 p 小于预设阈值，被标注为 0。

在步骤 309 中，将文档 D 中概率 p 大于预设阈值的句子抽取为候选句子，即抽取第二句、第四句以及第五句为候选句子 h'_2 、 h'_4 以及 h'_5 。

在步骤 310 中，将候选句子的句子级编码向量作为输入序列输入生成式模型，计算第 i 个解码时刻的输出概率。

如图 4 可知， h'_2 在第 i 个解码时刻的输出概率大于预设阈值，并被标注为 1， h'_4 在第 $i+k$ 个解码时刻的输出概率小于预设阈值，并被标注为 0， h'_5 在第 $i+l$ 个解码时刻的输出概率大于预设阈值，并被标注为 1，其中，第 i 个解码时刻、第 $i+k$ 个解码时刻以及第 $i+l$ 个解码时刻为三个不同的解码时刻，且 k 、 l 均为整数。

在步骤 311 中，当输出概率大于预设阈值时，从文档 D 对应的词典中生成一个词语作为摘要的词语，即 w_1 与 w_3 为从词典中生成的词语，

在步骤 312 中，当输出概率小于预设阈值时，从候选句子中的目

标句子中抽取出目标词语作为摘要的词语，即 w_2 为从目标句子第二句中抽取出的词语。

图 6 是本申请一个示例性实施例提供的摘要生成装置的框图。该装置可以通过软件、硬件或两者的结合实现成为计算机设备的全部或部分。所述装置包括：获取模块 52、抽取模块 54 以及输出模块 56。

获取模块 52，用于获取文档 D，所述文档 D 包括至少一个句子，每个句子包括至少一个词语；

抽取模块 54，用于通过抽取式模型从所述文档 D 中抽取出 m 个候选句子；

10 输出模块 56，用于通过生成式模型根据所述 m 个候选句子输出目标词语，根据所述目标词语生成摘要。

在一些实施例中，所述抽取模块，包括：

计算单元，用于计算所述文档 D 中每个句子的句子级编码向量；

所述计算单元，还用于根据所述每个句子的句子级编码向量计算
15 所述文档 D 的文档级编码向量；

所述计算单元，还用于根据所述句子级编码向量和所述文档级编码向量计算所述文档中的每个句子的概率 P ，将所述文档 D 中所述概率 P 大于预设阈值的句子抽取为所述 m 个候选句子，每个所述句子的概率 P 的取值为 1 或 0。

20 在一些实施例中，所述计算单元，包括：

第一输入子单元，用于将所述句子中的词语沿前向传播方向输入第一双向循环神经网络，得到第一前向编码向量；

所述第一输入子单元，还用于将所述句子中的词语沿后向传播方向输入所述第一双向循环神经网络，得到第一后向编码向量；

25 所述第一输入子单元，还用于根据所述第一前向编码向量和所述

第一后向编码向量，得到所述句子的句子级编码向量。

在一些实施例中，所述输入子单元，还用于将 n 个句子的句子级编码向量沿前向传播方向输入第二双向循环神经网络，得到第二前向编码向量，其中， n 为所述文档中的句子的个数；

- 5 所述第一输入子单元，还用于将所述 n 个句子的句子级编码向量沿后向传播方向输入所述第二双向循环神经网络，得到第二后向编码向量；

所述第一输入子单元，还用于根据所述第二前向编码向量和所述第二后向编码向量，得到所述文档 D 的文档级编码向量。

- 10 在一些实施例中，所述输出模块，包括：

输入子模块，用于将所述 m 个候选句子的句子级编码向量作为输入序列输入所述生成式模型，计算第 i 个解码时刻的输出概率， i 为正整数；

- 生成子模块，还用于当所述输出概率大于所述预设阈值时，从所述文档 D 对应的词典中生成一个词语，作为所述摘要中的第 i 个词语；
15 所述词典中包括所述文档 D 中出现频率高于预设条件的词语；

抽取子模块，还用于当所述输出概率小于所述预设阈值时，从所述 m 个候选句子中的目标句子中抽取出目标词语，作为所述摘要中的第 i 个词语。

- 20 在一些实施例中，所述输入子模块，包括：

计算子单元，用于通过所述编码器根据所述 m 个候选句子的句子级编码向量计算第 i 个解码时刻的上下文向量 c_i ；

- 第二输入子单元，还用于将第 $i-1$ 个解码时刻的隐含状态、所述上下文向量 c_i 和所述摘要中的第 $i-1$ 个词语输入所述解码器，计算得
25 到第 i 个解码时刻的隐含状态，所述解码时刻的隐含状态用于表示所

述解码器在当前解码时刻下已经处理的部分输入序列;

所述计算子单元,还用于根据所述摘要中的第 $i-1$ 个词语、所述上下文向量 c_i 和所述第 i 个解码时刻的隐含状态,计算第 i 个解码时刻的输出概率,其中, i 为正整数,且当 $i-1=0$ 时,所述第 $i-1$ 个解码时刻的隐含状态取默认值,所述第 $i-1$ 个词语取默认值。

在一些实施例中,所述计算子单元,还用于按照如下公式计算所述上下文向量 c_i :

$$c_i = \sum_{j=1}^T \alpha_{ij} h_j$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{j=1}^T \exp(e_{ij})}$$

10

$$e_{ij} = a(h'_{i-1}, h_j)$$

其中, h_j 代表第 j 个编码时刻的隐含状态, T 代表所述输入序列的长度, h'_{i-1} 代表第 $i-1$ 个解码时刻的隐含状态, a 代表预设的前馈神经网络,其中,所述编码时刻的隐含状态用于表示所述编码器在当前编码时刻下已经处理的部分输入序列。

15

在一些实施例中,所述输出模块,还包括:

句子子模块,用于通过第一最大似然估计函数从所述 m 个候选句子中确定出所述第 i 个解码时刻下的目标句子;

词语子模块,用于通过第二最大似然估计函数从所述目标句子的词语中确定出目标词语。

20

图 7 是本申请一个示例性实施例提供的计算机设备的框图。该计算机设备可以是手机、平板电脑、膝上型便携计算机、台式计算机和服务器等等,该计算机设备包括:处理器 62 和存储器 64,可选地,该计算机设备还包括显示屏 66。

处理器 62 可以是单核处理器、多核处理器、嵌入式芯片以及具有计算能力的处理器中的至少一种。

存储器 64 存储有处理器的可执行指令。示意性的，存储器 64 中一种计算机可读存储介质，所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器 62 加载并执行以实现如图 1 至图 5 所示的任意一种摘要生成方法。

可选地，显示屏 66 是用于计算机设备显示文档 D 的内容以及根据文档 D 的内容生成的自动文本摘要的内容。

10 可选地，本申请实施例还提供了一种计算机可读存储介质，该存储介质中一种计算机可读存储介质，所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现如图 1 至图 5 所示的任意一种摘要生成方法，可选地，该计算机可读存
15 储介质包括高速存取存储器、非易失性存储器。

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。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介
20 质可以是只读存储器，磁盘或光盘等。

以上所述仅为本申请的较佳实施例，并不用以限制本申请，凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

权利要求书

- 1、一种摘要生成方法，由计算机设备执行，所述方法包括：
获取文档，所述文档包括至少一个句子；
通过抽取式模型从所述文档中抽取出 m 个候选句子；
5 通过生成式模型根据所述 m 个候选句子输出目标词语，根据所述目标词语生成摘要。
- 2、根据权利要求 1 所述的方法，其中，所述通过抽取式模型从所述文档中抽取出 m 个候选句子，包括：
计算所述文档中每个句子的句子级编码向量；
10 根据所述每个句子的句子级编码向量计算所述文档的文档级编码向量；
根据所述句子级编码向量和所述文档级编码向量计算所述文档中的每个句子的概率 P，将所述文档中所述概率 P 大于预设阈值的句子抽取为所述 m 个候选句子。
- 15 3、根据权利要求 2 所述的方法，其中，所述抽取式模型包括：
第一双向循环神经网络；
所述计算所述文档中每个句子的句子级编码向量，包括：
将所述句子中的词语沿前向传播方向输入所述第一双向循环神经网络，得到第一前向编码向量；
20 将所述句子中的词语沿后向传播方向输入所述第一双向循环神经网络，得到第一后向编码向量；
根据所述第一前向编码向量和所述第一后向编码向量，得到所述句子的句子级编码向量。
- 4、根据权利要求 2 所述的方法，其中，所述抽取式模型包括：
25 第二双向循环神经网络；

所述根据所述每个句子的句子级编码向量计算所述文档的文档级编码向量，包括：

将n个句子的句子级编码向量沿前向传播方向输入所述第二双向循环神经网络，得到第二前向编码向量，其中，n为所述文档中的句子的个数；

将所述n个句子的句子级编码向量沿后向传播方向输入所述第二双向循环神经网络，得到第二后向编码向量；

根据所述第二前向编码向量和所述第二后向编码向量，得到所述文档的文档级编码向量。

5 5、根据权利要求2所述的方法，其中，所述抽取式模型包括基于注意力机制的模型；

所述基于注意力机制的模型为：

$$p(S_i=1|D)=\sigma(a(S_i,h_i));$$

其中，D为所述文档； S_i 为所述文档D中第i个句子的句子级编码向量，所述 h_i 为所述文档D的文档级编码向量，a为预设的前馈神经网络， σ 为预设的第一非线性函数， $p(S_i=1|D)$ 代表在第i个句子的概率大于所述预设阈值时将所述第i个句子标注为1，并抽取所述第i个句子，i为正整数。

20 6、根据权利要求2至5任一所述的方法，其中，所述通过生成式模型根据所述m个候选句子输出目标词语，根据所述目标词语生成摘要，包括：

将所述m个候选句子的句子级编码向量作为输入序列输入所述生成式模型，计算第i个解码时刻的输出概率，i为正整数；

25 当所述输出概率大于所述预设阈值时，从所述文档对应的词典中生成一个词语，作为所述摘要中的第i个词语；所述词典中包括所述

文档中出现频率高于预设条件的词语；

当所述输出概率小于所述预设阈值时，从所述 m 个候选句子中的目标句子中抽取出目标词语，作为所述摘要中的第 i 个词语。

7、根据权利要求 6 所述的方法，其中，所述生成式模型包括编
5 码器和解码器；

所述将所述 m 个候选句子的句子级编码向量作为输入序列输入所述生成式模型，计算第 i 个解码时刻的输出概率，包括：

通过所述编码器根据所述 m 个候选句子的句子级编码向量计算第 i 个解码时刻的上下文向量 c_i ；

10 将第 $i-1$ 个解码时刻的隐含状态、所述上下文向量 c_i 和所述摘要中的第 $i-1$ 个词语输入所述解码器，计算得到第 i 个解码时刻的隐含状态，所述第 i 个解码时刻的隐含状态用于表示所述解码器在所述第 i 个解码时刻下已经处理的部分输入序列；

根据所述摘要中的第 $i-1$ 个词语、所述上下文向量 c_i 和所述第 i
15 个解码时刻的隐含状态，计算第 i 个解码时刻的输出概率。

8、根据权利要求 6 所述的方法，其中，所述生成式模型还包括：
第一最大似然估计函数和第二最大似然估计函数；

所述从所述 m 个候选句子中的目标句子中抽取出目标词语，作为所述摘要中的第 i 个词语，包括：

20 通过所述第一最大似然估计函数从所述 m 个候选句子中确定出所述第 i 个解码时刻下的目标句子；

通过所述第二最大似然估计函数从所述目标句子的词语中确定出目标词语。

9、一种摘要生成装置，所述装置包括：存储器、处 理器；其中，
25 所述存储器中存储有计算机可读指令，所述处理器执行所述存储中的

计算机可读指令，用于：

获取文档，所述文档包括至少一个句子，每个句子包括至少一个词语；

通过抽取式模型从所述文档中抽取出 m 个候选句子；

- 5 通过生成式模型根据所述 m 个候选句子输出目标词语，根据所述目标词语生成摘要。

- 10 10、一种计算机设备，所述设备包括处理器和存储器，所述存储器中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器
10 加载并执行以实现如权利要求 1 至 8 任一所述的摘要生成方法。

11、一种计算机可读存储介质，所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现如权利要求 1 至 8 任一所述的摘要生成方法。

15

1/7

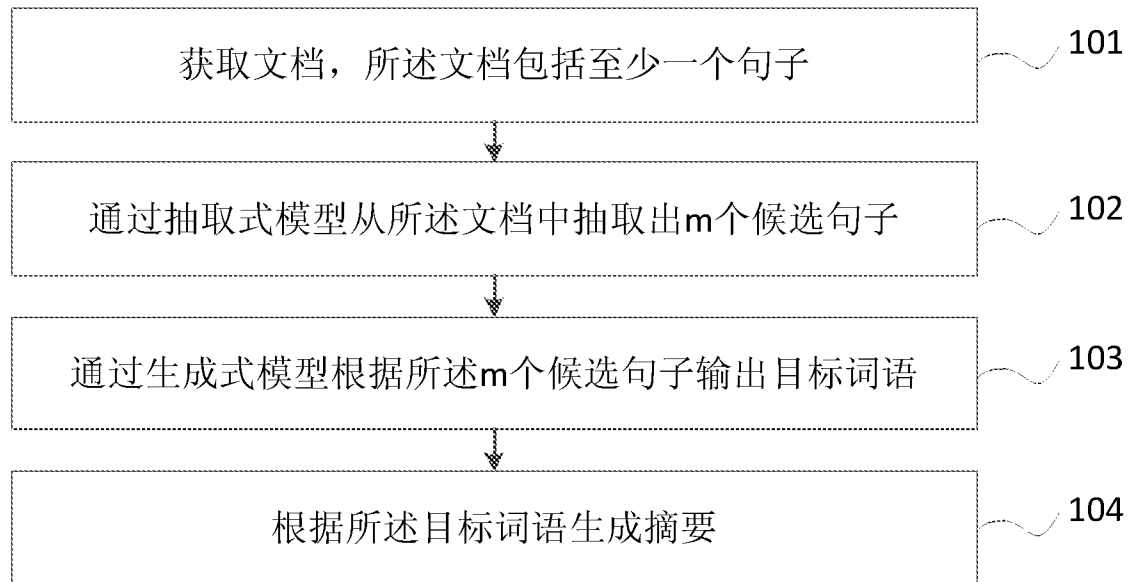


图 1

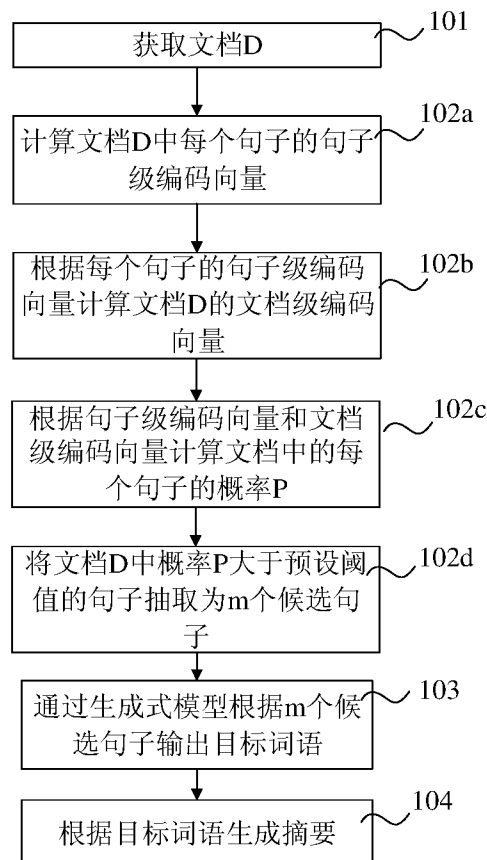


图 2a

2/7

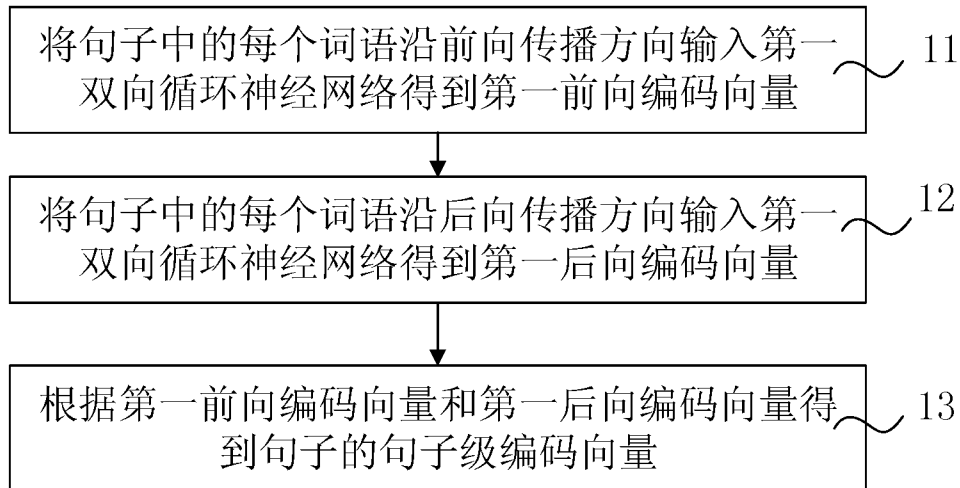


图 2b

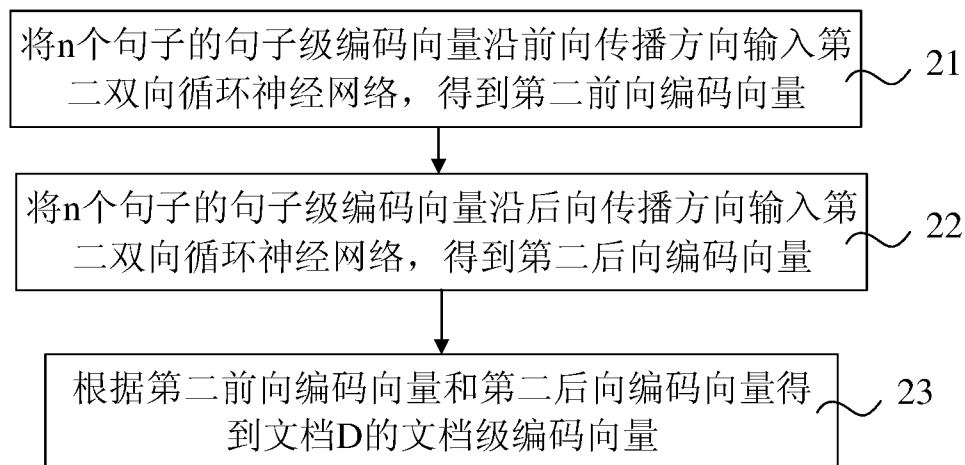


图 2c

3/7

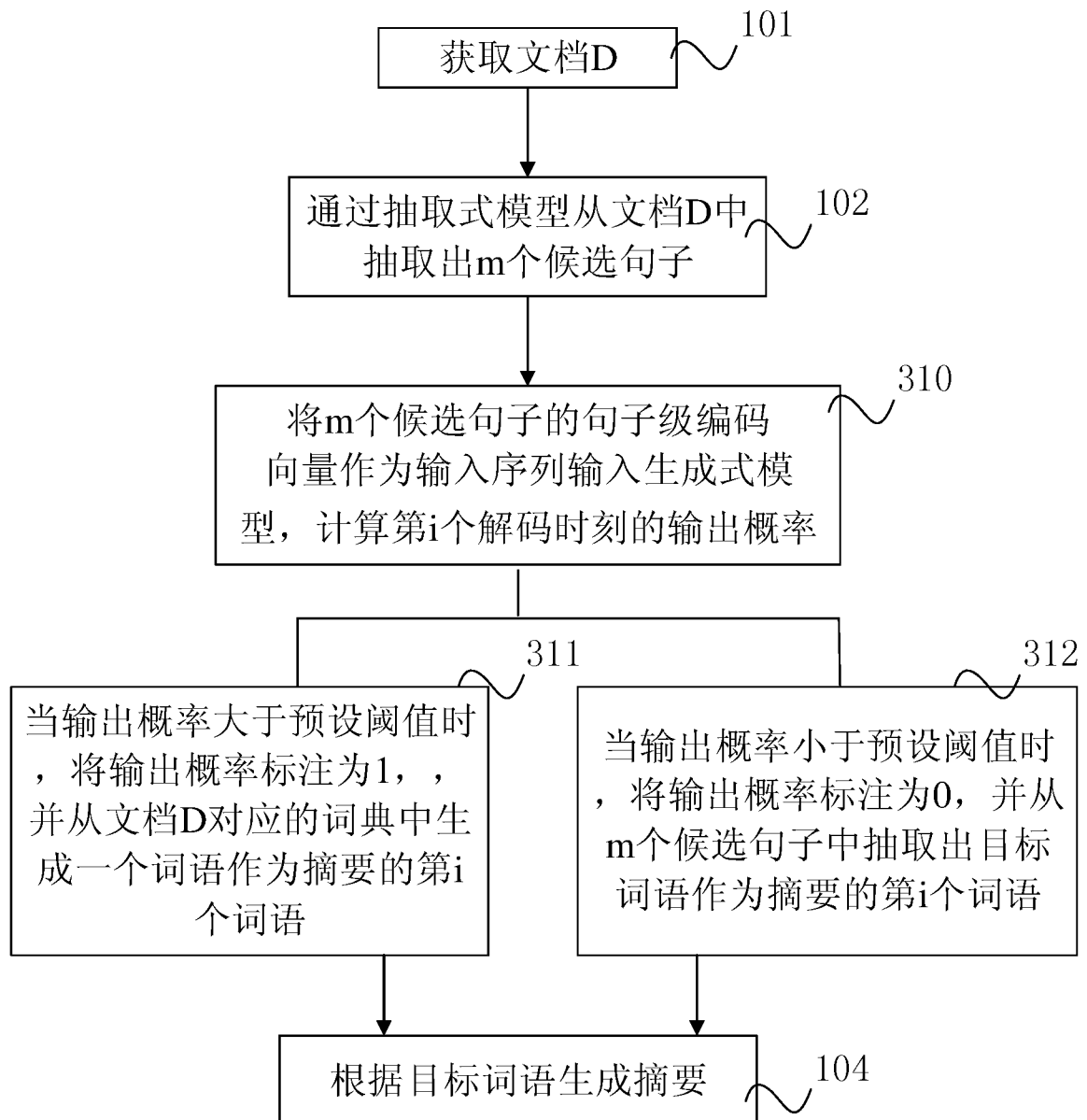


图 3a

4/7

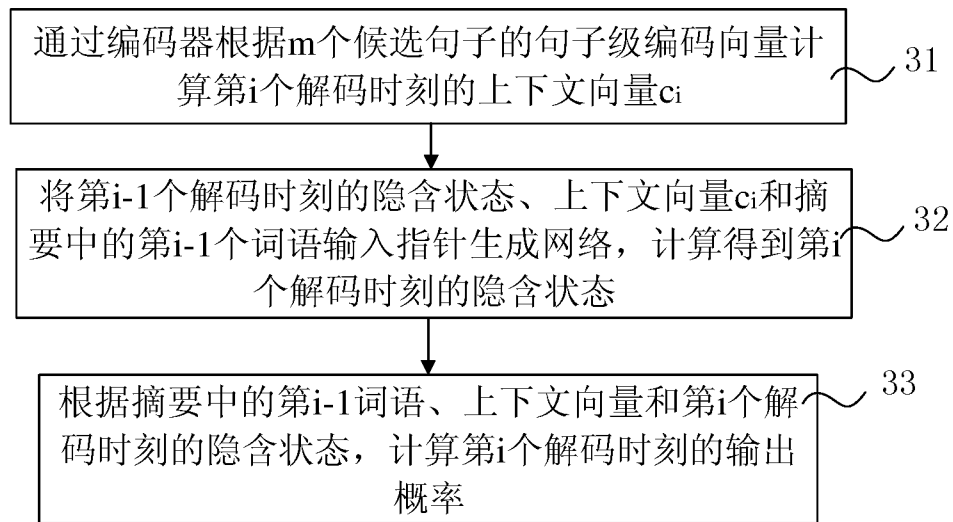


图 3b

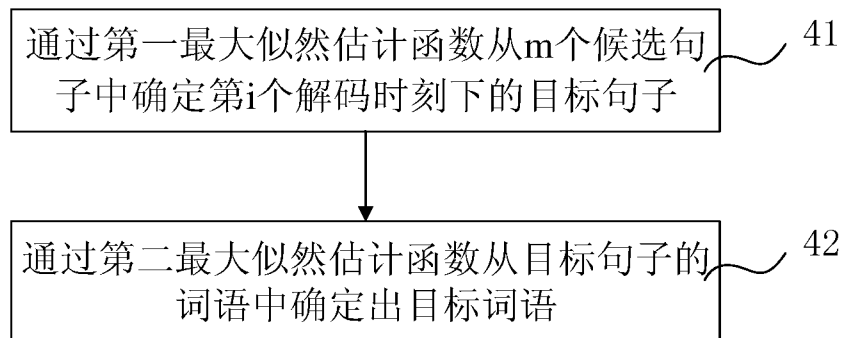


图 3c

5/7

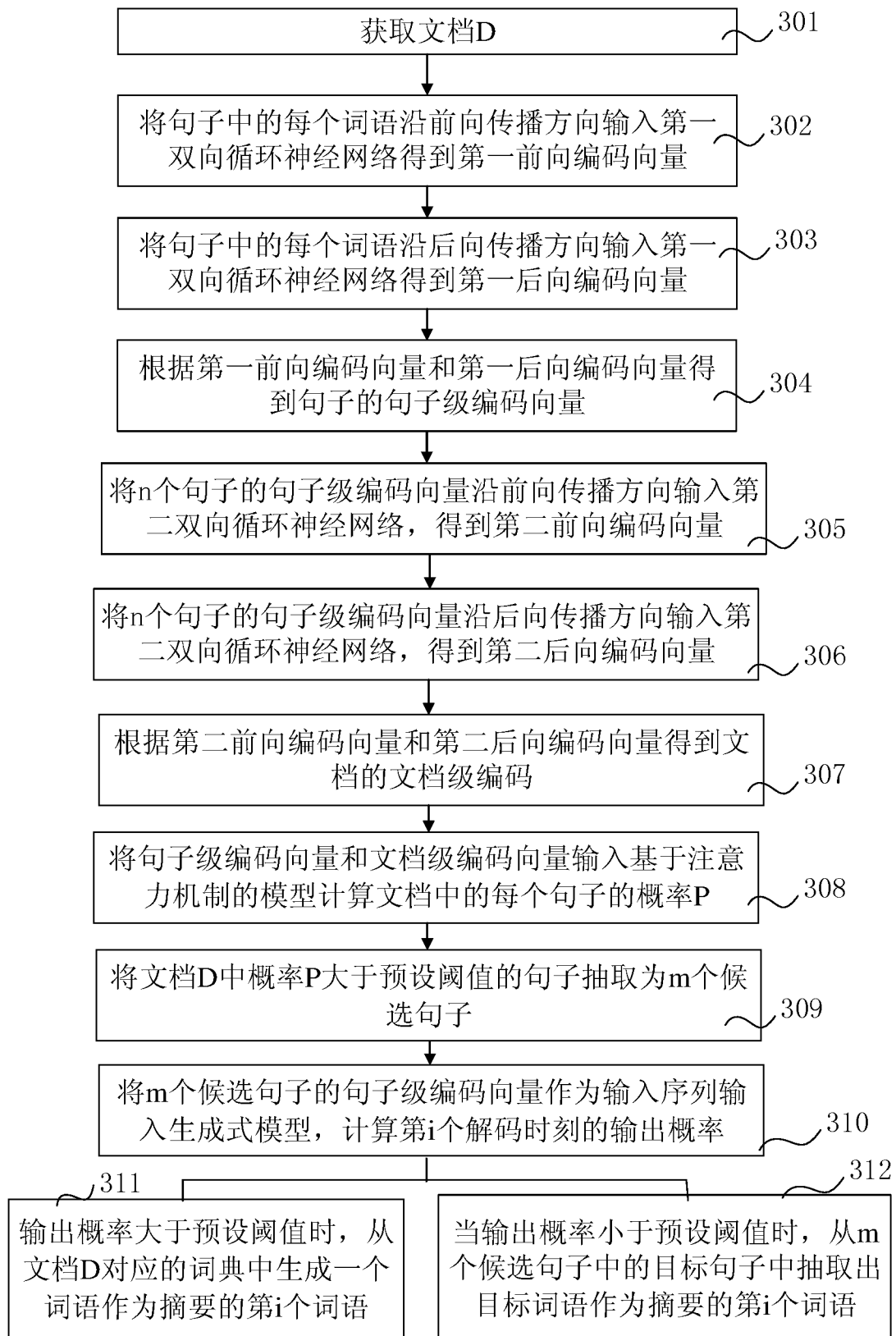


图 4

6/7

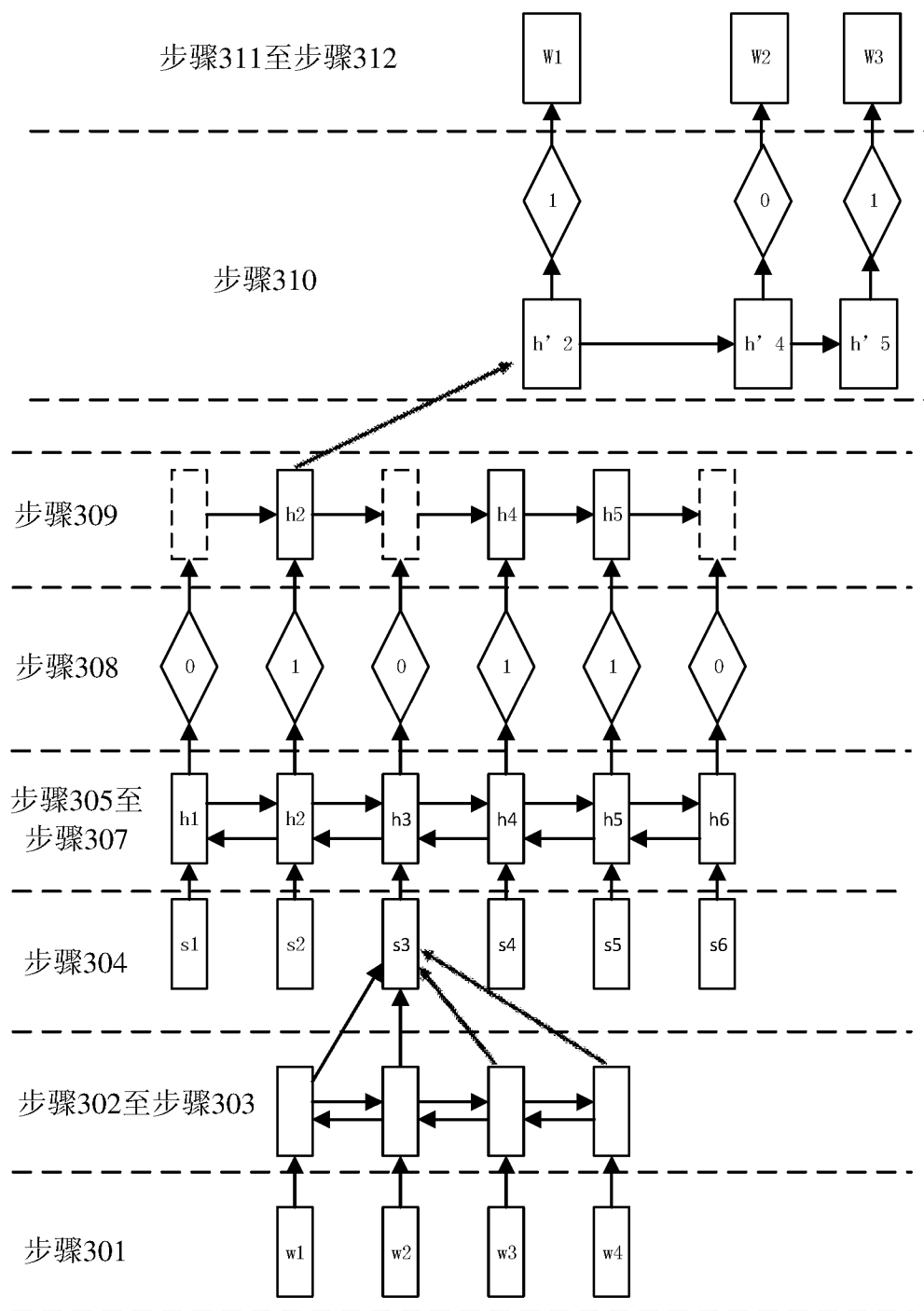


图 5

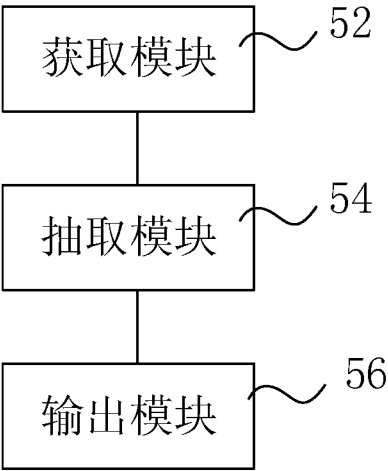


图 6

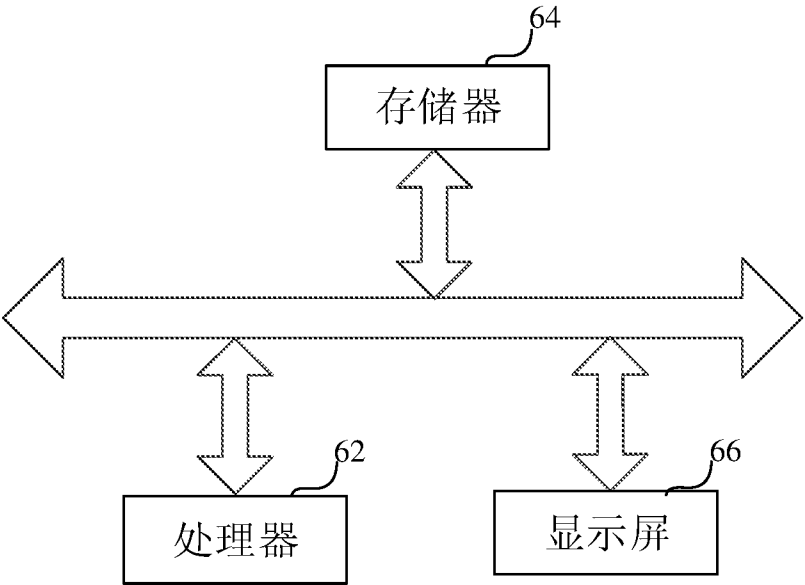


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/092078

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(2006.01)i; 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)

CNPAT, WPI, EPODOC, CNKI: 文档, 文本, 摘要, 文摘, 生成, 句子, 抽取, 提取, 模型, 候选, 目标, 编码, 向量, 矢量, 概率, 阈值, 双向, 循环, 神经网络, 前向, 后向, 注意力, 解码, 上下文, 似然, 估计, document?, text, summar+, abstract?, generat+, abstractive, sentence?, extract+, model?, candidate?, target?, cod???, vector?, probability, threshold?, bidirection?!, cycle, neural network?, forward+, backward+, attention, decod???, context, likelihood, estimat+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	林晶 (LIN, Jing). "一种基于混合神经网络的抽取式文本摘要方法 (On Extractive Summarization via Hybrid Neural Networks)" 怀化学院学报 (<i>Journal of Huaihua University</i>), Vol. 36, No. (5), 31 May 2017 (2017-05-31), pages 71-74	1-5, 9-11
Y	林晶 (LIN, Jing). "一种基于混合神经网络的抽取式文本摘要方法 (On Extractive Summarization via Hybrid Neural Networks)" 怀化学院学报 (<i>Journal of Huaihua University</i>), Vol. 36, No. (5), 31 May 2017 (2017-05-31), pages 71-74	6-8, 10-11
Y	CN 106502985 A (TSINGHUA UNIVERSITY) 15 March 2017 (2017-03-15) description, paragraphs [0002]-[0004] and [0069]-[0132]	6-8, 10-11
A	CN 101751425 A (PEKING UNIVERSITY ET AL.) 23 June 2010 (2010-06-23) entire document	1-11
A	RUSH, A.M. et al. "A Neural Attention Model for Sentence Summarization" <i>Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing</i> , 21 September 2015 (2015-09-21), pages 379-389	1-11



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

03 August 2018

Date of mailing of the international search report

06 September 2018

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/092078

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	106502985	A	15 March 2017	None	
CN	101751425	A	23 June 2010	None	

国际检索报告

国际申请号

PCT/CN2018/092078

A. 主题的分类

G06F 17/30(2006.01)i; 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))

CNPAT, WPI, EPODOC, CNKI; 文档, 文本, 摘要, 文摘, 生成, 句子, 抽取, 提取, 模型, 候选, 目标, 编码, 向量, 矢量, 概率, 阈值, 双向, 循环, 神经网络, 前向, 后向, 注意力, 解码, 上下文, 似然, 估计, document?, text, summar+, abstract?, generat+, abstractive, sentence?, extract+, model?, candidate?, target?, cod???, vector?, probability, threshold?, bidirection?!, cycle, neural network?, forward+, backward+, attention, decod???, context, likelihood, estimat+

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	林晶. “一种基于混合神经网络的抽取式文本摘要方法” 怀化学院学报, 第36卷, 第5期, 2017年 5月 31日 (2017 - 05 - 31), 第71-74页	1-5、9-11
Y	林晶. “一种基于混合神经网络的抽取式文本摘要方法” 怀化学院学报, 第36卷, 第5期, 2017年 5月 31日 (2017 - 05 - 31), 第71-74页	6-8、10-11
Y	CN 106502985 A (清华大学) 2017年 3月 15日 (2017 - 03 - 15) 说明书第[0002]-[0004]、[0069]-[0132]段	6-8、10-11
A	CN 101751425 A (北京大学 等) 2010年 6月 23日 (2010 - 06 - 23) 全文	1-11
A	ALEXANDER M. Rush等. “A Neural Attention Model for Sentence Summarization” Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015年 9月 21日 (2015 - 09 - 21), 第379-389页	1-11

☐ 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

2018年 8月 3日

国际检索报告邮寄日期

2018年 9月 6日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(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

魏晶瑶

电话号码 86-(10)-53961340

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/092078

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	106502985	A	2017年 3月 15日	无	
CN	101751425	A	2010年 6月 23日	无	