

(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
2019年9月12日 (12.09.2019)



(10) 国际公布号
WO 2019/169719 A1

(51) 国际专利分类号:
G06F 17/27 (2006.01)

(21) 国际申请号: PCT/CN2018/085249

(22) 国际申请日: 2018年5月2日 (02.05.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810191506.3 2018年3月8日 (08.03.2018) 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八卦岭八卦三路平安大厦六楼, Guangdong 518000 (CN)。

(72) 发明人: 林林 (LIN, Lin); 中国广东省深圳市福田区八卦岭八卦三路平安大厦六楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市精英专利事务所 (SHENZHEN TALENT PATENT SERVICE); 中国广东省深圳市福田区深南中路6009号绿景广场B栋20层B, 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, 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,

(54) Title: AUTOMATIC ABSTRACT EXTRACTION METHOD AND APPARATUS, AND COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 文摘自动提取方法、装置、计算机设备及存储介质

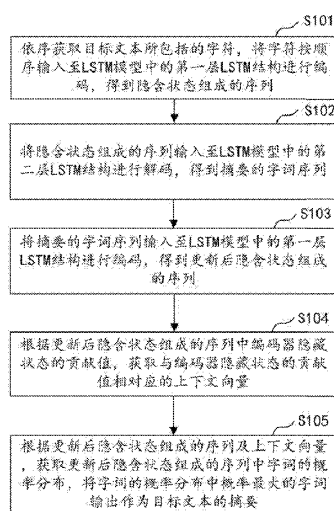


图 1

S101 Sequentially acquire characters included in a target text, and sequentially input the characters into a first-layer LSTM structure in an LSTM model for encoding to obtain a sequence composed of implicit states

S102 Input the sequence composed of implicit states into a second-layer LSTM structure in the LSTM model for decoding to obtain a word sequence of an abstract

S103 Input the word sequence of the abstract into the first-layer LSTM structure in the LSTM model for encoding to obtain an updated sequence composed of implicit states

S104 According to a contribution value of a hidden state, in the updated sequence composed of implicit states, of an encoder, acquire a context vector corresponding to the contribution value of the hidden state of the encoder

S105 According to the updated sequence composed of implicit states and the context vector, acquire the probability distribution of words in the updated sequence composed of implicit states, and output a word with the highest probability in the probability distribution of the words, and take same as an abstract of the target text

(57) Abstract: Disclosed are an automatic abstract extraction method and apparatus, and a computer device and a storage medium. The method comprises: sequentially acquiring characters of a target text, and sequentially inputting the characters into a first-layer LSTM structure in an LSTM model for encoding to obtain a sequence composed of implicit states; inputting the sequence composed of implicit states into a second-layer LSTM structure in the LSTM model for decoding to obtain a word sequence of an abstract; inputting the word sequence of the abstract into the first-layer LSTM structure for encoding to obtain an updated sequence composed of implicit states;

[见续页]

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))。

and according to a contribution value of a hidden state, in the updated sequence composed of implicit states, of an encoder, acquiring a context vector, acquiring the probability distribution of corresponding words, and taking a word with the highest probability as an abstract of the target text. In the method, after a target text is encoded and decoded by means of LSTM, an abstract of the target text is obtained by combining context variables; and acquiring an abstract in a generalized manner improves the accuracy of acquisition.

(57) 摘要: 本申请公开了一种文摘自动提取方法、装置、计算机设备及存储介质。该方法包括: 依序获取目标文本的字符并按顺序输入至LSTM模型中第一层LSTM结构进行编码, 得到隐含状态组成的序列; 将隐含状态组成的序列输入至LSTM模型中第二层LSTM结构进行解码得到摘要的字词序列; 将摘要的字词序列输入第一层LSTM结构进行编码得到更新后隐含状态组成的序列; 根据更新后隐含状态组成的序列中编码器隐藏状态的贡献值获取上下文向量, 并获取对应字词的概率分布, 将概率最大的字词作为目标文本的摘要。该方法采用LSTM对目标文本进行编码解码后, 结合上下文变量得到目标文本的摘要, 采取概括方式获取摘要, 提高获取准确性。