(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 10 月 17 日 (17.10.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/196236 A1

(51) 国际专利分类号:
G06F 17/27 (2006.01)

(21) 国际申请号: PCT/CN2018/096258

(22) 国际申请日: 2018 年 7 月 19 日 (19.07.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810309685.6 2018年4月9日 (09.04.2018) 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)。

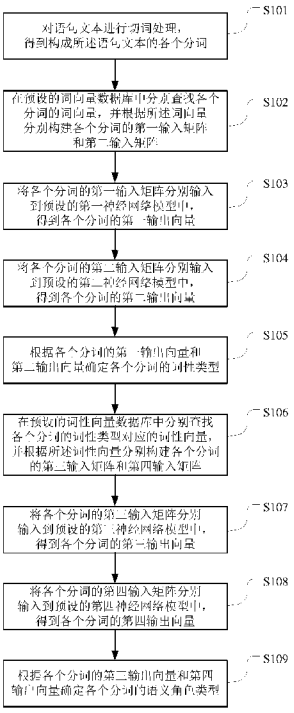
(72) 发明人: 张依 (ZHANG, Yi); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 汪伟 (WANG, Wei); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 肖京 (XIAO, Jing); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳中一专利商标事务所 (SHENZHEN ZHONGYI PATENT AND TRADEMARK OFFICE); 中国广东省深圳市福田区深南中路1014号老特区报社四楼（5号信箱）, Guangdong 518028 (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,

(54) **Title:** SEMANTIC ROLE ANALYSIS METHOD, READABLE STORAGE MEDIUM, TERMINAL DEVICE AND APPARATUS

(54) **发明名称:** 语义角色分析方法、可读存储介质、终端设备及装置



S101 PERFORM WORD-CUTTING ON A SENTENCE TEXT TO OBTAIN EACH WORD SEGMENT COMPRISED BY THE SENTENCE TEXT
S102 SEARCH FOR WORD VECTORS OF EACH WORD SEGMENT IN A PRESET WORD VECTOR DATABASE, AND RESPECTIVELY CONSTRUCT A FIRST INPUT MATRIX AND A SECOND INPUT MATRIX OF THE EACH WORD SEGMENT ACCORDING TO THE WORD VECTOR
S103 INPUT THE FIRST INPUT MATRIX OF THE EACH WORD SEGMENT INTO A PRESET FIRST NEURAL NETWORK MODEL SO AS TO OBTAIN A FIRST OUTPUT VECTOR OF THE EACH WORD SEGMENT
S104 INPUT THE SECOND INPUT MATRIX OF THE EACH WORD SEGMENT INTO A PRESET SECOND NEURAL NETWORK MODEL SO AS TO OBTAIN A SECOND OUTPUT VECTOR OF THE EACH WORD SEGMENT
S105 DETERMINE THE PART OF SPEECH TYPE OF THE EACH WORD SEGMENT ACCORDING TO THE FIRST OUTPUT VECTOR AND THE SECOND OUTPUT VECTOR OF THE EACH WORD SEGMENT
S106 SEARCH FOR A PART OF SPEECH VECTOR CORRESPONDING TO THE PART OF SPEECH TYPE OF THE EACH WORD SEGMENT IN THE PRESET PART OF SPEECH VECTOR DATABASE, AND CONSTRUCT A THIRD INPUT MATRIX AND A FOURTH INPUT MATRIX OF THE EACH WORD SEGMENT RESPECTIVELY ACCORDING TO THE PART OF SPEECH VECTOR
S107 INPUT THE THIRD INPUT MATRIX OF THE EACH WORD SEGMENT RESPECTIVELY INTO A PRESET THIRD NEURAL NETWORK MODEL SO AS TO OBTAIN A THIRD OUTPUT VECTOR OF THE EACH WORD SEGMENT
S108 INPUT THE FOURTH INPUT MATRIX OF THE EACH WORD SEGMENT RESPECTIVELY INTO A PRESET FOURTH NEURAL NETWORK MODEL SO AS TO OBTAIN A FOURTH OUTPUT VECTOR OF THE EACH WORD SEGMENT
S109 DETERMINE THE SEMANTIC ROLE TYPE OF THE EACH WORD SEGMENT ACCORDING TO THE THIRD OUTPUT VECTOR AND THE FOURTH OUTPUT VECTOR OF THE EACH WORD SEGMENT

图 1

(57) **Abstract:** The present application belongs to the technical field of computers, and relates in particular to a semantic role analysis method, a computer-readable storage medium, a terminal device and an apparatus. The method comprises: in the process of part-of-speech analysis, one neural network model is used for forward part-of-speech analysis, and another neural network model is used for reverse part-of-speech analysis. In the process of semantic role analysis, one neural network model is used for forward semantic role analysis, and another neural network model is used for reverse semantic role analysis. Thus, an originally complex neural network

WO 2019/196236 A1

[见续页]

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, 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))。

model is divided into relatively simple neural network models, and then the output of each neural network model is comprehensively processed to obtain a result. Due to the simplification of the neural network model structure, the computational load is greatly reduced and the analysis efficiency is improved.

(57) 摘要：本申请属于计算机技术领域，尤其涉及一种语义角色分析方法、计算机可读存储介质、终端设备及装置。所述方法在进行词性分析的过程中，使用一个神经网络模型进行正序词性分析，而使用另一个神经网络模型进行逆序词性分析。在进行语义角色分析的过程中，使用一个神经网络模型进行正序语义角色分析，而使用另一个神经网络模型进行逆序语义角色分析。也即将原本较为复杂的神经网络模型拆分为相对简单的神经网络模型，再对各个神经网络模型的输出进行综合处理得到结果，由于神经网络模型结构的简化，大大减少了计算量，提升了分析效率。

语义角色分析方法、可读存储介质、终端设备及装置

本申请要求于 2018 年 4 月 9 日提交中国专利局、申请号为 201810309685.6、发明名称为“一种语义角色分析方法、计算机可读存储介质及终端设备”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。

技术领域

本申请属于计算机技术领域，尤其涉及一种语义角色分析方法、计算机可读存储介质、终端设备及装置。

背景技术

目前，主流的语义角色分析研究主要集中于使用各种机器学习技术，利用多种语言学特征，进行了语义角色的识别和分类，通常的做法是首先使用一个神经网络模型进行各个分词的词性的确定，然后再通过一个神经网络模型进行各个分词的语义角色的确定，由于在计算过程中，需要在单个的神经网络模型中考虑整个语句对分词判定结果的影响，神经网络模型往往构造的非常复杂，计算量巨大，效率低下。

技术问题

有鉴于此，本申请实施例提供了一种语义角色分析方法、计算机可读存储介质、终端设备及装置，以解决目前的语义角色分析方法需要在单个的神经网络模型中考虑整个语句对分词判定结果的影响，神经网络模型往往构造的非常复杂，计算量巨大，效率低下的问题。

技术解决方案

本申请实施例的第一方面提供了一种语义角色分析方法，可以包括：

对语句文本进行切词处理，得到构成所述语句文本的各个分词；

在预设的词向量数据库中分别查找各个分词的词向量，并根据所述词向量分别构建各个分词的第一输入矩阵和第二输入矩阵，所述词向量数据库为记录词语与词向量之间的对应关系的数据库；

将各个分词的第一输入矩阵分别输入到预设的第一神经网络模型中，得到各个分词的第一输出向量，所述第一神经网络模型为进行正序词性分析的神经网络模型；

将各个分词的第二输入矩阵分别输入到预设的第二神经网络模型中，得到各个分词的第二输出向量，所述第二神经网络模型为进行逆序词性分析的神经网络模型；

根据各个分词的第一输出向量和第二输出向量确定各个分词的词性类型；

在预设的词性向量数据库中分别查找各个分词的词性类型对应的词性向量，并根据所述词性向量分别构建各个分词的第三输入矩阵和第四输入矩阵，所述词性向量数据库为记录词性类型与词性向量之间的对应关系的数据库；

将各个分词的第三输入矩阵分别输入到预设的第三神经网络模型中，得到各个分词的第三输出向量，所述第三神经网络模型为进行正序语义角色分析的神经网络模型；

将各个分词的第四输入矩阵分别输入到预设的第四神经网络模型中，得到各个分词的第四输出向量，所述第四神经网络模型为进行逆序语义角色分析的神经网络模型；

根据各个分词的第三输出向量和第四输出向量确定各个分词的语义角色类型。

本申请实施例的第二方面提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机可读指令，所述计算机可读指令被处理器执行时实现上述语义角色分析方法的步骤。

本申请实施例的第三方面提供了一种语义角色分析终端设备，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上述语义角色分析方法的步骤。

本申请实施例的第四方面提供了一种语义角色分析装置，可以包括用于实现上述语义角色分析方法的步骤的模块。

有益效果

本申请实施例与现有技术相比存在的有益效果是：本申请实施例将原本较为复杂的神经网络模型拆分为相对简单的神经网络模型，再对各个神经网络模型的输出进行综合处理得到结果，由于神经网络模型结构的简化，大大减少了计算量，提升了分析效率。

附图说明

图 1 为本申请实施例中一种语义角色分析方法的一个实施例流程图；

图 2 为第一神经网络模型的处理过程的示意流程图；

图 3 为第二神经网络模型的处理过程的示意流程图；
图 4 为第三神经网络模型的处理过程的示意流程图；
图 5 为第四神经网络模型的处理过程的示意流程图；
图 6 为本申请实施例中一种语义角色分析装置的一个实施例结构图；
图 7 为本申请实施例中一种语义角色分析终端设备的示意框图。

本发明的实施方式

请参阅图 1，本申请实施例中一种语义角色分析方法的一个实施例可以包括：

步骤 S101，对语句文本进行切词处理，得到构成所述语句文本的各个分词。

切词处理是指将一个语句文本切分成一个一个单独的词，也即各个所述分词，在本实施例中，可以根据通用词典对语句文本进行切分，保证分出的词语都是正常词汇，如词语不在词典内则分出单字。当前后方向都可以成词时，例如“要求神”，会根据统计词频的大小划分，如“要求”词频高则分出“要求/神”，如“求神”词频高则分出“要/求神”。

步骤 S102，在预设的词向量数据库中分别查找各个分词的词向量，并根据所述词向量分别构建各个分词的第一输入矩阵和第二输入矩阵。

所述词向量数据库为记录词语与词向量之间的对应关系的数据库，所述词向量可以是根据 word2vec 模型训练词语所得到对应的词向量。

具体地，可以根据下式分别构建各个分词的第一输入矩阵：

$$FwWdMatrix_n=\begin{bmatrix} FwWdEm_{1,1} & FwWdEm_{1,2} & \dots & FwWdEm_{1,wvl} & \dots & FwWdEm_{1,wVecLen} \\ FwWdEm_{2,1} & FwWdEm_{2,2} & \dots & FwWdEm_{2,wvl} & \dots & FwWdEm_{2,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ FwWdEm_{cl,1} & FwWdEm_{cl,2} & \dots & FwWdEm_{cl,wvl} & \dots & FwWdEm_{cl,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ FwWdEm_{CoupLen,1} & FwWdEm_{CoupLen,2} & \dots & FwWdEm_{CoupLen,wvl} & \dots & FwWdEm_{CoupLen,wVecLen} \end{bmatrix}$$

其中，n 为所述分词按照前后顺序依次排列的序号，$1\le n\le N$，N 为所述语句文本的分词总数，cl 为所述第一输入矩阵的行号，$1\le cl\le CoupLen$，$CoupLen$ 为预设的耦合长度，wvl 为所述第一输入矩阵的列号，$1\le wvl\le wVecLen$，$wVecLen$ 为任意一个所述分词的词向量的长度，第 n 个分词的词向量为 $WordVec_n$，且

$$WordVec_n=(WdVecEm_{n,1},WdVecEm_{n,2},.....,WdVecEm_{n,vl},.....,WdVecEm_{n,wVecLen})$$，

$$FwWdEm_{cl,wvl}=\begin{cases} WdVecEm_{n-cl+1,wvl}, 1\le cl\le n \\ 0 \qquad\qquad ,n<cl \end{cases}$$，$FwWdMatrix_n$ 为第 n 个分词的第一输入矩阵。

根据下式分别构建各个分词的第二输入矩阵：

$$BkWdMatrix_n=\begin{bmatrix} BkWdEm_{1,1} & BkWdEm_{1,2} & \dots & BkWdEm_{1,wvl} & \dots & BkWdEm_{1,wVecLen} \\ BkWdEm_{2,1} & BkWdEm_{2,2} & \dots & BkWdEm_{2,wvl} & \dots & BkWdEm_{2,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ BkWdEm_{cl,1} & BkWdEm_{cl,2} & \dots & BkWdEm_{cl,wvl} & \dots & BkWdEm_{cl,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ BkWdEm_{CoupLen,1} & BkWdEm_{CoupLen,2} & \dots & BkWdEm_{CoupLen,wvl} & \dots & BkWdEm_{CoupLen,wVecLen} \end{bmatrix}$$

$$BkWdEm_{cl,wvl}=\begin{cases} WdVecEm_{cl,wvl}, 1\le cl\le n \\ 0 \qquad\qquad ,n<cl \end{cases}$$，$BkWdMatrix_n$ 为第 n 个分词的第二输入矩阵。

步骤 S103，将各个分词的第一输入矩阵分别输入到预设的第一神经网络模型中，得到各个分词的第一输出向量。

所述第一神经网络模型为进行正序词性分析的神经网络模型，所述第一神经网络模型的处理过程具体可以包括如图 2 所示的步骤：

步骤 S1031，分别计算各个分词的第一复合向量。

具体地，可以根据下式分别计算各个分词的第一复合向量：

$$FwWdCpVec_n = (FwWdCpEm_{n,1}, FwWdCpEm_{n,2}, ..., FwWdCpEm_{n,wvl}, ..., FwWdCpEm_{n,wVecLen})$$

其中，

$$FwWdCpEm_{n,wvl} = \tanh\{\sum_{n=2}^{CoupLen}[FwWdEm_{n,wvl} * \frac{ln(n)}{\sum_{n=2}^{CoupLen} ln(n)} * FwWdWt_{wvl}] + FwWdEm_{1,wvl} * FwWdWt'_{wvl}\}$$

ln 为自然对数函数，tanh 为双曲正切函数，$FwWdWt_{wvl}$、$FwWdWt'_{wvl}$ 均为预设的权重系数。

步骤 S1032，分别计算各个词性类型的第一概率值。

具体地，可以根据下式分别计算各个词性类型的第一概率值：

$$FwWdProb_{n,m} = \frac{sigmoid(FwWdCpVec_n * FwWdWtVec_m^T)}{\sum_{m=1}^{M} sigmoid(FwWdCpVec_n * FwWdWtVec_m^T)}$$

其中，m 为词性类型的序号，$1 \le m \le M$，M 为词性类型的个数，$FwWdWtVec_m$ 为预设的与第 m 个词性类型对应的权值向量，$sigmoid(x) = \frac{1}{1+e^{-x}}$，$FwWdProb_{n,m}$ 为第 n 个分词是第 m 个词性类型的第一概率值。

步骤 S1033，构建各个分词的第一输出向量。

具体地，可以根据下式构建各个分词的第一输出向量：

$$FwWdVec_n = (FwWdProb_{n,1}, FwWdProb_{n,2}, ..., FwWdProb_{n,m}, ..., FwWdProb_{n,M})$$

其中，$FwWdVec_n$ 为第 n 个分词的第一输出向量。

步骤 S104，将各个分词的第二输入矩阵分别输入到预设的第二神经网络模型中，得到各个分词的第二输出向量。

所述第二神经网络模型为进行逆序词性分析的神经网络模型，所述第二神经网络模型的处理过程具体可以包括如图 3 所示的步骤：

步骤 S1041，分别计算各个分词的第二复合向量。

具体地，可以根据下式分别计算各个分词的第二复合向量：

$$BkWdCpVec_n = (BkWdCpEm_{n,1}, BkWdCpEm_{n,2}, ..., BkWdCpEm_{n,wvl}, ..., BkWdCpEm_{n,wVecLen})$$

其中，

$$BkWdCpEm_{n,wvl} = \tanh\{\sum_{n=2}^{CoupLen}[BkWdEm_{n,wvl} * \frac{ln(n)}{\sum_{n=2}^{CoupLen} ln(n)} * BkWdWt_{wvl}] + BkWdEm_{1,wvl} * BkWdWt'_{wvl}\}$$

$BkWdWt_{wvl}$、$BkWdWt'_{wvl}$ 均为预设的权重系数。

步骤 S1042，分别计算各个词性类型的第二概率值。

具体地，可以根据下式分别计算各个词性类型的第二概率值：

$$BkWdProb_{n,m} = \frac{sigmoid(BkWdCpVec_n * BkWdWtVec_m^T)}{\sum_{m=1}^{M} sigmoid(BkWdCpVec_n * BkWdWtVec_m^T)}$$

其中，$BkWdWtVec_m$ 为预设的与第 m 个词性类型对应的权值向量，$BkWdProb_{n,m}$ 为第 n 个分词是第 m 个词性类型的第二概率值。

步骤 S1043，构建各个分词的第二输出向量。

具体地，可以根据下式构建各个分词的第二输出向量：

$$BkWdVec_n = (BkWdProb_{n,1}, BkWdProb_{n,2}, ..., BkWdProb_{n,m}, ..., BkWdProb_{n,M})$$

其中，$BkWdVec_n$ 为第 n 个分词的第二输出向量。

步骤 S105，根据各个分词的第一输出向量和第二输出向量确定各个分词的词性类型。

具体地，可以根据下式分别计算各个分词的词性概率向量：

$$WdProbVec_n = (WdProb_{n,1}, WdProb_{n,2}, ..., WdProb_{n,m}, ..., WdProb_{n,M})$$

其中，$WdProb_{n,m} = \eta_1 * FwWdProb_{n,m} + \eta_2 * BkWdProb_{n,m}$，$\eta_1$、$\eta_2$ 均为预设的权重系数，$WdProbVec_n$ 为第 n 个分词的词性概率向量。

根据下式分别确定各个分词的词性类型：

$$CharSeq_n = \arg\max(WdProbVec_n)$$

其中，$\arg\max$ 为最大自变量函数，$CharSeq_n$ 为第 n 个分词的词性类型序号。也即将第 n 个分词的词性概率向量中取值最大的元素所对应的词性类型确定为第 n 个分词的词性类型。

步骤 S106，在预设的词性向量数据库中分别查找各个分词的词性类型对应的词性向量，并根据所述词性向量分别构建各个分词的第三输入矩阵和第四输入矩阵。

所述词性向量数据库为记录词性类型与词性向量之间的对应关系的数据库。所述词性向量为各个词性类型所对应的向量形式，即根据词性类型的上下文信息来表示该词性类型出现的概率。词性向量的训练先将每个词性类型表示成一个 0-1 向量（one-hot）形式，再进行模型训练，用 n-1 个词的词性类型来预测第 n 个词的词性类型，神经网络模型预测后得到的中间过程作为词性向量。具体地，如词性类型“名词”的 one-hot 向量假设定为[1,0,0,0,……,0]，词性类型“形容词”的 one-hot 向量为[0,1,0,0,……,0]，词性类型“动词”的 one-hot 向量为[0,0,1,0, ……,0]，词性类型“副词”的向量[0,0,0,1,……,0]，模型经过训练会生成隐藏层的系数矩阵 W，每个词性类型的 one-hot 向量和系数矩阵的乘积为该词性类型的词性向量，最后的形式将是类似于“[-0.11,0.26,-0.03,…...,0.71]”这样的一个多维向量。

具体地，可以根据下式分别构建各个分词的第三输入矩阵：

$$FwCrMatrix_n = \begin{bmatrix} FwCrEm_{1,1} & FwCrEm_{1,2} & ... & FwCrEm_{1,cvl} & ... & FwCrEm_{1,cVecLen} \\ FwCrEm_{2,1} & FwCrEm_{2,2} & ... & FwCrEm_{2,cvl} & ... & FwCrEm_{2,cVecLen} \\ ... & ... & ... & ... & ... & ... \\ FwCrEm_{cl,1} & FwCrEm_{cl,2} & ... & FwCrEm_{cl,cvl} & ... & FwCrEm_{cl,cVecLen} \\ ... & ... & ... & ... & ... & ... \\ FwCrEm_{CoupLen,1} & FwCrEm_{CoupLen,2} & ... & FwCrEm_{CoupLen,cvl} & ... & FwCrEm_{CoupLen,cVecLen} \end{bmatrix}$$

其中，n 为所述分词按照前后顺序依次排列的序号，$1 \le n \le N$，N 为所述语句文本的分词总数，cl 为所述第三输入矩阵的行号，$1 \le cl \le CoupLen$，$CoupLen$ 为预设的耦合长度，cvl 为所述第三输入矩阵的列号，$1 \le cvl \le cVecLen$，$cVecLen$ 为任意一个所述分词的词性向量的长度，第 n 个分词的词性向量为 $CharVec_n$，且

$$CharVec_n = (CrVecEm_{n,1}, CrVecEm_{n,2}, ..., CrVecEm_{n,cvl}, ..., CrVecEm_{n,cVecLen}),$$

$$FwCrEm_{cl,cvl} = \begin{cases} CrVecEm_{n-cl+1,cvl}, & 1 \le cl \le n \\ 0, & n < cl \end{cases}$$，$FwCrMatrix_n$ 为第 n 个分词的第三输入矩阵。

根据下式分别构建各个分词的第四输入矩阵：

$$BkCrMatrix_n = \begin{bmatrix} BkCrEm_{1,1} & BkCrEm_{1,2} & ... & BkCrEm_{1,cvl} & ... & BkCrEm_{1,cVecLen} \\ BkCrEm_{2,1} & BkCrEm_{2,2} & ... & BkCrEm_{2,cvl} & ... & BkCrEm_{2,cVecLen} \\ ... & ... & ... & ... & ... & ... \\ BkCrEm_{cl,1} & BkCrEm_{cl,2} & ... & BkCrEm_{cl,cvl} & ... & BkCrEm_{cl,cVecLen} \\ ... & ... & ... & ... & ... & ... \\ BkCrEm_{CoupLen,1} & BkCrEm_{CoupLen,2} & ... & BkCrEm_{CoupLen,cvl} & ... & BkCrEm_{CoupLen,cVecLen} \end{bmatrix}$$

$$BkCrEm_{cl,cvl} = \begin{cases} CrVecEm_{cl,cvl}, & 1 \le cl \le n \\ 0, & n < cl \end{cases}$$，$BkCrMatrix_n$ 为第 n 个分词的第四输入矩阵。

步骤 S107，将各个分词的第三输入矩阵分别输入到预设的第三神经网络模型中，得到各个分词的第三输出向量。

所述第三神经网络模型为进行正序语义角色分析的神经网络模型，所述第三神经网络模型的处理过程具体可以包括如图 4 所示的步骤：

步骤 S1071，分别计算各个分词的第三复合向量。

具体地，可以根据下式分别计算各个分词的第三复合向量：

$$FwCrCpVec_n = (FwCrCpEm_{n,1}, FwCrCpEm_{n,2}, ..., FwCrCpEm_{n,vl}, ..., FwCrCpEm_{n,cVecLen})$$

其中，

$$FwCrCpEm_{n,cvl} = \tanh\{\sum_{n=2}^{CoupLen}[FwCrEm_{n,cvl} * \frac{ln(n)}{\sum_{n=2}^{CoupLen} ln(n)} * FwCrWt_{cvl}] + FwCrEm_{1,cvl} * FwCrWt'_{cvl}\}$$

ln 为自然对数函数，tanh 为双曲正切函数，$FwCrWt_{cvl}$、$FwCrWt'_{cvl}$ 均为预设的权重系数。

步骤 S1072，分别计算各个语义角色类型的第一概率值。

具体地，可以根据下式分别计算各个语义角色类型的第一概率值：

$$FwCrProb_{n,l} = \frac{sigmoid(FwCrCpVec_n * FwCrWtVec_l^T)}{\sum_{l=1}^{L} sigmoid(FwCrCpVec_n * FwCrWtVec_l^T)}$$

其中，l 为语义角色类型的序号，$1 \le l \le L$，L 为语义角色类型的个数，$FwCrWtVec_l$ 为预设的与第 l 个语义角色类型对应的权值向量，$sigmoid(x) = \frac{1}{1+e^{-x}}$，$FwCrProb_{n,l}$ 为第 n 个分词是第 l 个语义角色类型的第一概率值。

步骤 S1073，构建各个分词的第三输出向量。

具体地，可以根据下式构建各个分词的第三输出向量：

$$FwCrVec_n = (FwCrProb_{n,1}, FwCrProb_{n,2}, ..., FwCrProb_{n,l}, ..., FwCrProb_{n,L})$$

其中，$FwCrVec_n$ 为第 n 个分词的第三输出向量。

步骤 S108，将各个分词的第四输入矩阵分别输入到预设的第四神经网络模型中，得到各个分词的第四输出向量。

所述第四神经网络模型为进行逆序语义角色分析的神经网络模型，所述第三神经网络模型的处理过程具体可以包括如图 5 所示的步骤：

步骤 S1081，分别计算各个分词的第四复合向量。

具体地，可以根据下式分别计算各个分词的第四复合向量：

$BkCrCpVec_n = (BkCrCpEm_{n,1}, BkCrCpEm_{n,2}, ..., BkCrCpEm_{n,cvl}, ..., BkCrCpEm_{n,cVecLen})$，其中，

$$BkCrCpEm_{n,cvl} = \tanh\{\sum_{n=2}^{CoupLen}[BkCrEm_{n,cvl} * \frac{ln(n)}{\sum_{n=2}^{CoupLen} ln(n)} * BkCrWt_{cvl}] + BkCrEm_{1,cvl} * BkCrWt'_{cvl}\}$$

$BkCrWt_{cvl}$、$BkCrWt'_{cvl}$ 均为预设的权重系数。

步骤 S1082，分别计算各个语义角色类型的第二概率值。

具体地，可以根据下式分别计算各个语义角色类型的第二概率值：

$$BkCrProb_{n,l} = \frac{sigmoid(BkCrCpVec_n * BkCrWtVec_l^T)}{\sum_{l=1}^{L} sigmoid(BkCrCpVec_n * BkCrWtVec_l^T)}$$

其中，$BkCrWtVec_l$ 为预设的与第 l 个语义角色类型对应的权值向量，$BkCrProb_{n,l}$ 为第 n 个分

词是第l个语义角色类型的第二概率值。

步骤 S1083，构建各个分词的第四输出向量。

具体地，可以根据下式构建各个分词的第四输出向量：

$$BkCrVec_n = (BkCrProb_{n,1}, BkCrProb_{n,2}, ..., BkCrProb_{n,l}, ..., BkCrProb_{n,L})$$

其中，$BkCrVec_n$为第 n 个分词的第四输出向量。

步骤 S109，根据各个分词的第三输出向量和第四输出向量确定各个分词的语义角色类型。

具体地，可以根据下式分别计算各个分词的语义角色概率向量：

$$CrProbVec_n = (CrProb_{n,1}, CrProb_{n,2}, ..., CrProb_{n,l}, ..., CrProb_{n,L})$$

其中，$CrProb_{n,l} = \xi_1 * FwCrProb_{n,l} + \xi_2 * BkCrProb_{n,l}$，$\xi_1$、$\xi_2$均为预设的权重系数，$CrProbVec_n$为第 n 个分词的语义角色概率向量。

根据下式分别确定各个分词的语义角色类型：

$$RoleSeq_n = \arg\max(CrProbVec_n)$$

其中，$\arg\max$为最大自变量函数，$RoleSeq_n$为第 n 个分词的语义角色类型序号。也即将第 n 个分词的语义角色概率向量中取值最大的元素所对应的语义角色类型确定为第 n 个分词的语义角色类型。也即将第 n 个分词的语义角色概率向量中取值最大的元素所对应的语义角色类型确定为第 n 个分词的语义角色类型。

综上所述，本申请实施例在两个最关键的处理过程中，都采用了两个神经网络模型进行处理，将原本较为复杂的神经网络模型拆分为相对简单的神经网络模型，再对各个神经网络模型的输出进行综合处理得到结果，由于神经网络模型结构的简化，大大减少了计算量，提升了分析效率。

对应于上文实施例所述的一种语义角色分析方法，图 6 示出了本申请实施例提供的一种语义角色分析装置的一个实施例结构图。

本实施例中，一种语义角色分析装置可以包括：

切词处理模块 601，用于对语句文本进行切词处理，得到构成所述语句文本的各个分词；

词向量查找模块 602，用于在预设的词向量数据库中分别查找各个分词的词向量，所述词向量数据库为记录词语与词向量之间的对应关系的数据库；

词向量矩阵构建模块 603，用于根据所述词向量分别构建各个分词的第一输入矩阵和第二输入矩阵；

第一处理模块 604，用于将各个分词的第一输入矩阵分别输入到预设的第一神经网络模型中，得到各个分词的第一输出向量，所述第一神经网络模型为进行正序词性分析的神经网络模型；

第二处理模块 605，用于将各个分词的第二输入矩阵分别输入到预设的第二神经网络模型中，得到各个分词的第二输出向量，所述第二神经网络模型为进行逆序词性分析的神经网络模型；

词性类型确定模块 606，用于根据各个分词的第一输出向量和第二输出向量确定各个分词的词性类型；

词性向量查找模块 607，用于在预设的词性向量数据库中分别查找各个分词的词性类型对应的词性向量，所述词性向量数据库为记录词性类型与词性向量之间的对应关系的数据库；

词性向量矩阵构建模块 608，用于根据所述词性向量分别构建各个分词的第三输入矩阵和第四输入矩阵；

第三处理模块 609，用于将各个分词的第三输入矩阵分别输入到预设的第三神经网络模型中，得到各个分词的第三输出向量，所述第三神经网络模型为进行正序语义角色分析的神经网络模型；

第四处理模块 610，用于将各个分词的第四输入矩阵分别输入到预设的第四神经网络模型中，得到各个分词的第四输出向量，所述第四神经网络模型为进行逆序语义角色分析的神经网络模型；

语义角色类型确定模块 611，用于根据各个分词的第三输出向量和第四输出向量确定各个分词的语义角色类型。

所述语义角色分析装置的具体实施例与上述语义角色分析方法各实施例基本相同，可以参考前述方法实施例中的相关描述，在此不作赘述。

图 7 示出了本申请实施例提供的一种语义角色分析终端设备的示意框图，为了便于说明，仅示出了与本申请实施例相关的部分。

在本实施例中，所述语义角色分析终端设备 7 可以是手机、平板电脑、桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。该语义角色分析终端设备 7 可包括：处理器 70、存储器 71 以及存储在所述存储器 71 中并可在所述处理器 70 上运行的计算机可读指令 72，例如执行上述的语义角色分析方法的计算机可读指令。所述处理器 70 执行所述计算机可读指令 72 时实现上述各个语义角色分析方法实施例中的步骤。

在本申请各个实施例中的各功能单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读存储介质中。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或者该技术方案的全部或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一个存储介质中，包括若干计算机可读指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本申请各个实施例所述方法的全部或部分步骤。

权利要求书

1. 一种语义角色分析方法，其特征在于，包括：

对语句文本进行切词处理，得到构成所述语句文本的各个分词；

在预设的词向量数据库中分别查找各个分词的词向量，并根据所述词向量分别构建各个分词的第一输入矩阵和第二输入矩阵，所述词向量数据库为记录词语与词向量之间的对应关系的数据库；

将各个分词的第一输入矩阵分别输入到预设的第一神经网络模型中，得到各个分词的第一输出向量，所述第一神经网络模型为进行正序词性分析的神经网络模型；

将各个分词的第二输入矩阵分别输入到预设的第二神经网络模型中，得到各个分词的第二输出向量，所述第二神经网络模型为进行逆序词性分析的神经网络模型；

根据各个分词的第一输出向量和第二输出向量确定各个分词的词性类型；

在预设的词性向量数据库中分别查找各个分词的词性类型对应的词性向量，并根据所述词性向量分别构建各个分词的第三输入矩阵和第四输入矩阵，所述词性向量数据库为记录词性类型与词性向量之间的对应关系的数据库；

将各个分词的第三输入矩阵分别输入到预设的第三神经网络模型中，得到各个分词的第三输出向量，所述第三神经网络模型为进行正序语义角色分析的神经网络模型；

将各个分词的第四输入矩阵分别输入到预设的第四神经网络模型中，得到各个分词的第四输出向量，所述第四神经网络模型为进行逆序语义角色分析的神经网络模型；

根据各个分词的第三输出向量和第四输出向量确定各个分词的语义角色类型。

2. 根据权利要求 1 所述的语义角色分析方法，其特征在于，所述根据所述词向量分别构建各个分词的第一输入矩阵和第二输入矩阵包括：

根据下式分别构建各个分词的第一输入矩阵：

$$FwWdMatrix_n=\begin{bmatrix} FwWdEm_{1,1} & FwWdEm_{1,2} & \dots & FwWdEm_{1,wvl} & \dots & FwWdEm_{1,wVecLen} \\ FwWdEm_{2,1} & FwWdEm_{2,2} & \dots & FwWdEm_{2,wvl} & \dots & FwWdEm_{2,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ FwWdEm_{cl,1} & FwWdEm_{cl,2} & \dots & FwWdEm_{cl,wvl} & \dots & FwWdEm_{cl,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ FwWdEm_{CoupLen,1} & FwWdEm_{CoupLen,2} & \dots & FwWdEm_{CoupLen,wvl} & \dots & FwWdEm_{CoupLen,wVecLen} \end{bmatrix}$$

其中，n 为所述分词按照前后顺序依次排列的序号，$1\le n\le N$，N 为所述语句文本的分词总数，cl 为所述第一输入矩阵的行号，$1\le cl\le CoupLen$，$CoupLen$ 为预设的耦合长度，wvl 为所述第一输入矩阵的列号，$1\le wvl\le wVecLen$，$wVecLen$ 为任意一个所述分词的词向量的长度，第 n 个分词的词向量为 $WordVec_n$，且

$WordVec_n=(WdVecEm_{n,1},WdVecEm_{n,2},.....,WdVecEm_{n,vl},.....,WdVecEm_{n,wVecLen})$，

$FwWdEm_{cl,wvl}=\begin{cases} WdVecEm_{n-cl+1,wvl}, 1\le cl\le n \\ 0 \qquad\qquad ,n<cl \end{cases}$，$FwWdMatrix_n$ 为第 n 个分词的第一输入矩阵；

根据下式分别构建各个分词的第二输入矩阵：

$$BkWdMatrix_n=\begin{bmatrix} BkWdEm_{1,1} & BkWdEm_{1,2} & \dots & BkWdEm_{1,wvl} & \dots & BkWdEm_{1,wVecLen} \\ BkWdEm_{2,1} & BkWdEm_{2,2} & \dots & BkWdEm_{2,wvl} & \dots & BkWdEm_{2,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ BkWdEm_{cl,1} & BkWdEm_{cl,2} & \dots & BkWdEm_{cl,wvl} & \dots & BkWdEm_{cl,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ BkWdEm_{CoupLen,1} & BkWdEm_{CoupLen,2} & \dots & BkWdEm_{CoupLen,wvl} & \dots & BkWdEm_{CoupLen,wVecLen} \end{bmatrix}$$

$BkWdEm_{cl,wvl}=\begin{cases} WdVecEm_{cl,wvl}, 1\le cl\le n \\ 0 \qquad\qquad ,n<cl \end{cases}$，$BkWdMatrix_n$ 为第 n 个分词的第二输入矩阵。

3. 根据权利要求 2 所述的语义角色分析方法，其特征在于，所述第一神经网络模型的处理过程包括：

根据下式分别计算各个分词的第一复合向量：

$$FwWdCpVec_n=(FwWdCpEm_{n,1},FwWdCpEm_{n,2},.....,FwWdCpEm_{n,wvl},.....,FwWdCpEm_{n,wVecLen})$$

其中，

$$FwWdCpEm_{n,wvl}=\tanh\{\sum_{n=2}^{CoupLen}[FwWdEm_{n,wvl}*\frac{ln(n)}{\sum_{n=2}^{CoupLen}ln(n)}*FwWdWt_{wvl}]+FwWdEm_{1,wvl}*FwWdWt'_{wvl}\}$$

ln 为自然对数函数，tanh 为双曲正切函数，$FwWdWt_{wvl}$、$FwWdWt'_{wvl}$ 均为预设的权重系数；

根据下式分别计算各个词性类型的第一概率值：

$$FwWdProb_{n,m}=\frac{sigmoid(FwWdCpVec_n*FwWdWtVec_m^T)}{\sum_{m=1}^{M}sigmoid(FwWdCpVec_n*FwWdWtVec_m^T)}$$

其中，m 为词性类型的序号，$1\le m\le M$，M 为词性类型的个数，$FwWdWtVec_m$ 为预设的与第 m 个词性类型对应的权值向量，$sigmoid(x)=\frac{1}{1+e^{-x}}$，$FwWdProb_{n,m}$ 为第 n 个分词是第 m 个词性类型的第一概率值；

根据下式构建各个分词的第一输出向量：

$$FwWdVec_n=(FwWdProb_{n,1},FwWdProb_{n,2},.....,FwWdProb_{n,m},.....,FwWdProb_{n,M})$$

其中，$FwWdVec_n$ 为第 n 个分词的第一输出向量；

所述第二神经网络模型的处理过程包括：

根据下式分别计算各个分词的第二复合向量：

$$BkWdCpVec_n=(BkWdCpEm_{n,1},BkWdCpEm_{n,2},.....,BkWdCpEm_{n,wvl},.....,BkWdCpEm_{n,wVecLen})$$

其中，

$$BkWdCpEm_{n,wvl}=\tanh\{\sum_{n=2}^{CoupLen}[BkWdEm_{n,wvl}*\frac{ln(n)}{\sum_{n=2}^{CoupLen}ln(n)}*BkWdWt_{wvl}]+BkWdEm_{1,wvl}*BkWdWt'_{wvl}\}$$

$BkWdWt_{wvl}$、$BkWdWt'_{wvl}$ 均为预设的权重系数；

根据下式分别计算各个词性类型的第二概率值：

$$BkWdProb_{n,m}=\frac{sigmoid(BkWdCpVec_n*BkWdWtVec_m^T)}{\sum_{m=1}^{M}sigmoid(BkWdCpVec_n*BkWdWtVec_m^T)}$$

其中，$BkWdWtVec_m$ 为预设的与第 m 个词性类型对应的权值向量，$BkWdProb_{n,m}$ 为第 n 个分词是第 m 个词性类型的第二概率值；

根据下式构建各个分词的第二输出向量：

$$BkWdVec_n=(BkWdProb_{n,1},BkWdProb_{n,2},.....,BkWdProb_{n,m},.....,BkWdProb_{n,M})$$

其中，$BkWdVec_n$ 为第 n 个分词的第二输出向量。

4. 根据权利要求 3 所述的语义角色分析方法，其特征在于，所述根据各个分词的第一输出向量和第二输出向量确定各个分词的词性类型包括：

根据下式分别计算各个分词的词性概率向量：

$$WdProbVec_n=(WdProb_{n,1},WdProb_{n,2},.....,WdProb_{n,m},.....,WdProb_{n,M})$$

其中，$WdProb_{n,m}=\eta_1*FwWdProb_{n,m}+\eta_2*BkWdProb_{n,m}$，$\eta_1$、$\eta_2$ 均为预设的权重系数，

$WdProbVec_n$ 为第 n 个分词的词性概率向量；

根据下式分别确定各个分词的词性类型：

$$CharSeq_n = \arg\max(WdProbVec_n)$$

其中，$\arg\max$ 为最大自变量函数，$CharSeq_n$ 为第 n 个分词的词性类型序号。

5. 根据权利要求 1 所述的语义角色分析方法，其特征在于，所述根据所述词性向量分别构建各个分词的第三输入矩阵和第四输入矩阵包括：

根据下式分别构建各个分词的第三输入矩阵：

$$FwCrMatrix_n = \begin{bmatrix} FwCrEm_{1,1} & FwCrEm_{1,2} & \dots & FwCrEm_{1,cvl} & \dots & FwCrEm_{1,cVecLen} \\ FwCrEm_{2,1} & FwCrEm_{2,2} & \dots & FwCrEm_{2,cvl} & \dots & FwCrEm_{2,cVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ FwCrEm_{cl,1} & FwCrEm_{cl,2} & \dots & FwCrEm_{cl,cvl} & \dots & FwCrEm_{cl,cVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ FwCrEm_{CoupLen,1} & FwCrEm_{CoupLen,2} & \dots & FwCrEm_{CoupLen,cvl} & \dots & FwCrEm_{CoupLen,cVecLen} \end{bmatrix}$$

其中，n 为所述分词按照前后顺序依次排列的序号，$1 \le n \le N$，N 为所述语句文本的分词总数，cl 为所述第三输入矩阵的行号，$1 \le cl \le CoupLen$，$CoupLen$ 为预设的耦合长度，cvl 为所述第三输入矩阵的列号，$1 \le cvl \le cVecLen$，$cVecLen$ 为任意一个所述分词的词性向量的长度，第 n 个分词的词性向量为 $CharVec_n$，且

$$CharVec_n = (CrVecEm_{n,1}, CrVecEm_{n,2}, ..., CrVecEm_{n,cvl}, ..., CrVecEm_{n,cVecLen})$$，

$$FwCrEm_{cl,cvl} = \begin{cases} CrVecEm_{n-cl+1,cvl}, 1 \le cl \le n \\ 0, n < cl \end{cases}$$，$FwCrMatrix_n$ 为第 n 个分词的第三输入矩阵；

根据下式分别构建各个分词的第四输入矩阵：

$$BkCrMatrix_n = \begin{bmatrix} BkCrEm_{1,1} & BkCrEm_{1,2} & \dots & BkCrEm_{1,cvl} & \dots & BkCrEm_{1,cVecLen} \\ BkCrEm_{2,1} & BkCrEm_{2,2} & \dots & BkCrEm_{2,cvl} & \dots & BkCrEm_{2,cVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ BkCrEm_{cl,1} & BkCrEm_{cl,2} & \dots & BkCrEm_{cl,cvl} & \dots & BkCrEm_{cl,cVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ BkCrEm_{CoupLen,1} & BkCrEm_{CoupLen,2} & \dots & BkCrEm_{CoupLen,cvl} & \dots & BkCrEm_{CoupLen,cVecLen} \end{bmatrix}$$

$$BkCrEm_{cl,cvl} = \begin{cases} CrVecEm_{cl,cvl}, 1 \le cl \le n \\ 0, n < cl \end{cases}$$，$BkCrMatrix_n$ 为第 n 个分词的第四输入矩阵。

6. 根据权利要求 5 所述的语义角色分析方法，其特征在于，所述第三神经网络模型的处理过程包括：

根据下式分别计算各个分词的第三复合向量：

$$FwCrCpVec_n = (FwCrCpEm_{n,1}, FwCrCpEm_{n,2}, ..., FwCrCpEm_{n,vl}, ..., FwCrCpEm_{n,cVecLen})$$

其中，

$$FwCrCpEm_{n,cvl} = \tanh\{\sum_{n=2}^{CoupLen}[FwCrEm_{n,cvl} * \frac{ln(n)}{\sum_{n=2}^{CoupLen} ln(n)} * FwCrWt_{cvl}] + FwCrEm_{1,cvl} * FwCrWt'_{cvl}\}$$

ln 为自然对数函数，tanh 为双曲正切函数，$FwCrWt_{cvl}$、$FwCrWt'_{cvl}$ 均为预设的权重系数；

根据下式分别计算各个语义角色类型的第一概率值：

$$FwCrProb_{n,l}=\frac{sigmoid(FwCrCpVec_n * FwCrWtVec_l^T)}{\sum_{l=1}^{L} sigmoid(FwCrCpVec_n * FwCrWtVec_l^T)}$$

其中，l为语义角色类型的序号，$1\le l\le L$，L 为语义角色类型的个数，$FwCrWtVec_l$为预设的与第l个语义角色类型对应的权值向量，$sigmoid(x)=\frac{1}{1+e^{-x}}$，$FwCrProb_{n,l}$为第 n 个分词是第$l$个语义角色类型的第一概率值；

根据下式构建各个分词的第三输出向量：

$$FwCrVec_n=(FwCrProb_{n,1},FwCrProb_{n,2},.....,FwCrProb_{n,l},.....,FwCrProb_{n,L})$$

其中，$FwCrVec_n$为第 n 个分词的第三输出向量；

所述第四神经网络模型的处理过程包括：

根据下式分别计算各个分词的第四复合向量：

$$BkCrCpVec_n=(BkCrCpEm_{n,1},BkCrCpEm_{n,2},.....,BkCrCpEm_{n,cvl},.....,BkCrCpEm_{n,cVecLen})$$ 其中，

$$BkCrCpEm_{n,cvl}=\tanh\{\sum_{n=2}^{CoupLen}[BkCrEm_{n,cvl}*\frac{ln(n)}{\sum_{n=2}^{CoupLen} ln(n)}*BkCrWt_{cvl}]+BkCrEm_{1,cvl}*BkCrWt'_{cvl}\}$$

$BkCrWt_{cvl}$、$BkCrWt'_{cvl}$均为预设的权重系数；

根据下式分别计算各个语义角色类型的第二概率值：

$$BkCrProb_{n,l}=\frac{sigmoid(BkCrCpVec_n * BkCrWtVec_l^T)}{\sum_{l=1}^{L} sigmoid(BkCrCpVec_n * BkCrWtVec_l^T)}$$

其中，$BkCrWtVec_l$为预设的与第l个语义角色类型对应的权值向量，$BkCrProb_{n,l}$为第 n 个分词是第l个语义角色类型的第二概率值；

根据下式构建各个分词的第四输出向量：

$$BkCrVec_n=(BkCrProb_{n,1},BkCrProb_{n,2},.....,BkCrProb_{n,l},.....,BkCrProb_{n,L})$$

其中，$BkCrVec_n$为第 n 个分词的第四输出向量。

7. 根据权利要求 6 所述的语义角色分析方法，其特征在于，所述根据各个分词的第三输出向量和第四输出向量确定各个分词的语义角色类型包括：

根据下式分别计算各个分词的语义角色概率向量：

$$CrProbVec_n=(CrProb_{n,1},CrProb_{n,2},.....,CrProb_{n,l},.....,CrProb_{n,L})$$

其中，$CrProb_{n,l}=\xi_1 * FwCrProb_{n,l}+\xi_2 * BkCrProb_{n,l}$，$\xi_1$、$\xi_2$均为预设的权重系数，$CrProbVec_n$为第 n 个分词的语义角色概率向量；

根据下式分别确定各个分词的语义角色类型：

$$RoleSeq_n=\arg\max(CrProbVec_n)$$

其中，$\arg\max$为最大自变量函数，$RoleSeq_n$为第 n 个分词的语义角色类型序号。

8. 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现如下步骤：

对语句文本进行切词处理，得到构成所述语句文本的各个分词；

在预设的词向量数据库中分别查找各个分词的词向量，并根据所述词向量分别构建各个分词的第一输入矩阵和第二输入矩阵，所述词向量数据库为记录词语与词向量之间的对应关系的数据库；

将各个分词的第一输入矩阵分别输入到预设的第一神经网络模型中，得到各个分词的第一输出向量，所述第一神经网络模型为进行正序词性分析的神经网络模型；

将各个分词的第二输入矩阵分别输入到预设的第二神经网络模型中，得到各个分词的第二输出向量，所述第二神经网络模型为进行逆序词性分析的神经网络模型；

根据各个分词的第一输出向量和第二输出向量确定各个分词的词性类型；

在预设的词性向量数据库中分别查找各个分词的词性类型对应的词性向量，并根据所述词性向量分别构建各个分词的第三输入矩阵和第四输入矩阵，所述词性向量数据库为记录词性类型与词性向量之间的对应关系的数据库；

将各个分词的第三输入矩阵分别输入到预设的第三神经网络模型中，得到各个分词的第三输出向量，所述第三神经网络模型为进行正序语义角色分析的神经网络模型；

将各个分词的第四输入矩阵分别输入到预设的第四神经网络模型中，得到各个分词的第四输出向量，所述第四神经网络模型为进行逆序语义角色分析的神经网络模型；

根据各个分词的第三输出向量和第四输出向量确定各个分词的语义角色类型。

9. 根据权利要求 8 所述的计算机可读存储介质，其特征在于，所述根据所述词向量分别构建各个分词的第一输入矩阵和第二输入矩阵包括：

根据下式分别构建各个分词的第一输入矩阵：

$$FwWdMatrix_n=\begin{bmatrix} FwWdEm_{1,1} & FwWdEm_{1,2} & \dots & FwWdEm_{1,wvl} & \dots & FwWdEm_{1,wVecLen} \\ FwWdEm_{2,1} & FwWdEm_{2,2} & \dots & FwWdEm_{2,wvl} & \dots & FwWdEm_{2,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ FwWdEm_{cl,1} & FwWdEm_{cl,2} & \dots & FwWdEm_{cl,wvl} & \dots & FwWdEm_{cl,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ FwWdEm_{CoupLen,1} & FwWdEm_{CoupLen,2} & \dots & FwWdEm_{CoupLen,wvl} & \dots & FwWdEm_{CoupLen,wVecLen} \end{bmatrix}$$

其中，n 为所述分词按照前后顺序依次排列的序号，$1\le n\le N$，N 为所述语句文本的分词总数，cl 为所述第一输入矩阵的行号，$1\le cl\le CoupLen$，$CoupLen$ 为预设的耦合长度，wvl 为所述第一输入矩阵的列号，$1\le wvl\le wVecLen$，$wVecLen$ 为任意一个所述分词的词向量的长度，第 n 个分词的词向量为 $WordVec_n$，且

$WordVec_n=(WdVecEm_{n,1},WdVecEm_{n,2},.....,WdVecEm_{n,vl},.....,WdVecEm_{n,wVecLen})$，

$FwWdEm_{cl,wvl}=\begin{cases} WdVecEm_{n-cl+1,wvl}, 1\le cl\le n \\ 0 \qquad\qquad ,n<cl \end{cases}$，$FwWdMatrix_n$ 为第 n 个分词的第一输入矩阵；

根据下式分别构建各个分词的第二输入矩阵：

$$BkWdMatrix_n=\begin{bmatrix} BkWdEm_{1,1} & BkWdEm_{1,2} & \dots & BkWdEm_{1,wvl} & \dots & BkWdEm_{1,wVecLen} \\ BkWdEm_{2,1} & BkWdEm_{2,2} & \dots & BkWdEm_{2,wvl} & \dots & BkWdEm_{2,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ BkWdEm_{cl,1} & BkWdEm_{cl,2} & \dots & BkWdEm_{cl,wvl} & \dots & BkWdEm_{cl,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ BkWdEm_{CoupLen,1} & BkWdEm_{CoupLen,2} & \dots & BkWdEm_{CoupLen,wvl} & \dots & BkWdEm_{CoupLen,wVecLen} \end{bmatrix}$$

$BkWdEm_{cl,wvl}=\begin{cases} WdVecEm_{cl,wvl}, 1\le cl\le n \\ 0 \qquad\qquad ,n<cl \end{cases}$，$BkWdMatrix_n$ 为第 n 个分词的第二输入矩阵。

10. 根据权利要求 9 所述的计算机可读存储介质，其特征在于，所述第一神经网络模型的处理过程包括：

根据下式分别计算各个分词的第一复合向量：

$FwWdCpVec_n=(FwWdCpEm_{n,1},FwWdCpEm_{n,2},.....,FwWdCpEm_{n,wvl},.....,FwWdCpEm_{n,wVecLen})$

其中，

$$FwWdCpEm_{n,wvl}=\tanh\{\sum_{n=2}^{CoupLen}[FwWdEm_{n,wvl}*\frac{ln(n)}{\sum_{n=2}^{CoupLen}ln(n)}*FwWdWt_{wvl}]+FwWdEm_{1,wvl}*FwWdWt'_{wvl}\}$$

ln 为自然对数函数，tanh 为双曲正切函数，$FwWdWt_{wvl}$、$FwWdWt'_{wvl}$ 均为预设的权重系数；

根据下式分别计算各个词性类型的第一概率值：

$$FwWdProb_{n,m}=\frac{sigmoid(FwWdCpVec_n*FwWdWtVec_m^T)}{\sum_{m=1}^{M}sigmoid(FwWdCpVec_n*FwWdWtVec_m^T)}$$

其中，m 为词性类型的序号，$1\le m\le M$，M 为词性类型的个数，$FwWdWtVec_m$ 为预设的与第 m 个词性类型对应的权值向量，$sigmoid(x)=\frac{1}{1+e^{-x}}$，$FwWdProb_{n,m}$ 为第 n 个分词是第 m 个词性类型的第一概率值；

根据下式构建各个分词的第一输出向量：

$$FwWdVec_n=(FwWdProb_{n,1},FwWdProb_{n,2},.....,FwWdProb_{n,m},.....,FwWdProb_{n,M})$$

其中，$FwWdVec_n$ 为第 n 个分词的第一输出向量；

所述第二神经网络模型的处理过程包括：

根据下式分别计算各个分词的第二复合向量：

$$BkWdCpVec_n=(BkWdCpEm_{n,1},BkWdCpEm_{n,2},.....,BkWdCpEm_{n,wvl},.....,BkWdCpEm_{n,wVecLen})$$

其中，

$$BkWdCpEm_{n,wvl}=\tanh\{\sum_{n=2}^{CoupLen}[BkWdEm_{n,wvl}*\frac{ln(n)}{\sum_{n=2}^{CoupLen}ln(n)}*BkWdWt_{wvl}]+BkWdEm_{1,wvl}*BkWdWt'_{wvl}\}$$

$BkWdWt_{wvl}$、$BkWdWt'_{wvl}$ 均为预设的权重系数；

根据下式分别计算各个词性类型的第二概率值：

$$BkWdProb_{n,m}=\frac{sigmoid(BkWdCpVec_n*BkWdWtVec_m^T)}{\sum_{m=1}^{M}sigmoid(BkWdCpVec_n*BkWdWtVec_m^T)}$$

其中，$BkWdWtVec_m$ 为预设的与第 m 个词性类型对应的权值向量，$BkWdProb_{n,m}$ 为第 n 个分词是第 m 个词性类型的第二概率值；

根据下式构建各个分词的第二输出向量：

$$BkWdVec_n=(BkWdProb_{n,1},BkWdProb_{n,2},.....,BkWdProb_{n,m},.....,BkWdProb_{n,M})$$

其中，$BkWdVec_n$ 为第 n 个分词的第二输出向量。

11. 一种语义角色分析终端设备，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：

对语句文本进行切词处理，得到构成所述语句文本的各个分词；

在预设的词向量数据库中分别查找各个分词的词向量，并根据所述词向量分别构建各个分词的第一输入矩阵和第二输入矩阵，所述词向量数据库为记录词语与词向量之间的对应关系的数据库；

将各个分词的第一输入矩阵分别输入到预设的第一神经网络模型中，得到各个分词的第一输出向量，所述第一神经网络模型为进行正序词性分析的神经网络模型；

将各个分词的第二输入矩阵分别输入到预设的第二神经网络模型中，得到各个分词的第二输出向量，所述第二神经网络模型为进行逆序词性分析的神经网络模型；

根据各个分词的第一输出向量和第二输出向量确定各个分词的词性类型；

在预设的词性向量数据库中分别查找各个分词的词性类型对应的词性向量，并根据所述词性向量分别构建各个分词的第三输入矩阵和第四输入矩阵，所述词性向量数据库为记录词性类型与词性向量之间的对应关系的数据库；

将各个分词的第三输入矩阵分别输入到预设的第三神经网络模型中，得到各个分词的第三输出向量，所述第三神经网络模型为进行正序语义角色分析的神经网络模型；

将各个分词的第四输入矩阵分别输入到预设的第四神经网络模型中，得到各个分词的第四输出向量，所述第四神经网络模型为进行逆序语义角色分析的神经网络模型；

根据各个分词的第三输出向量和第四输出向量确定各个分词的语义角色类型。

12. 根据权利要求 11 所述的语义角色分析终端设备，其特征在于，所述根据所述词向量分别构建各个分词的第一输入矩阵和第二输入矩阵包括：

根据下式分别构建各个分词的第一输入矩阵：

$$FwWdMatrix_n=\begin{bmatrix} FwWdEm_{1,1} & FwWdEm_{1,2} & \dots & FwWdEm_{1,wvl} & \dots & FwWdEm_{1,wVecLen} \\ FwWdEm_{2,1} & FwWdEm_{2,2} & \dots & FwWdEm_{2,wvl} & \dots & FwWdEm_{2,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ FwWdEm_{cl,1} & FwWdEm_{cl,2} & \dots & FwWdEm_{cl,wvl} & \dots & FwWdEm_{cl,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ FwWdEm_{CoupLen,1} & FwWdEm_{CoupLen,2} & \dots & FwWdEm_{CoupLen,wvl} & \dots & FwWdEm_{CoupLen,wVecLen} \end{bmatrix}$$

其中，n 为所述分词按照前后顺序依次排列的序号，$1\le n\le N$，N 为所述语句文本的分词总数，cl 为所述第一输入矩阵的行号，$1\le cl\le CoupLen$，$CoupLen$ 为预设的耦合长度，wvl 为所述第一输入矩阵的列号，$1\le wvl\le wVecLen$，$wVecLen$ 为任意一个所述分词的词向量的长度，第 n 个分词的词向量为 $WordVec_n$，且

$$WordVec_n=(WdVecEm_{n,1},WdVecEm_{n,2},.....,WdVecEm_{n,vl},.....,WdVecEm_{n,wVecLen}),$$

$$FwWdEm_{cl,wvl}=\begin{cases} WdVecEm_{n-cl+1,wvl}, 1\le cl\le n \\ 0 \qquad\qquad , n<cl \end{cases},$$ $FwWdMatrix_n$ 为第 n 个分词的第一输入矩阵；

根据下式分别构建各个分词的第二输入矩阵：

$$BkWdMatrix_n=\begin{bmatrix} BkWdEm_{1,1} & BkWdEm_{1,2} & \dots & BkWdEm_{1,wvl} & \dots & BkWdEm_{1,wVecLen} \\ BkWdEm_{2,1} & BkWdEm_{2,2} & \dots & BkWdEm_{2,wvl} & \dots & BkWdEm_{2,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ BkWdEm_{cl,1} & BkWdEm_{cl,2} & \dots & BkWdEm_{cl,wvl} & \dots & BkWdEm_{cl,wVecLen} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ BkWdEm_{CoupLen,1} & BkWdEm_{CoupLen,2} & \dots & BkWdEm_{CoupLen,wvl} & \dots & BkWdEm_{CoupLen,wVecLen} \end{bmatrix}$$

$$BkWdEm_{cl,wvl}=\begin{cases} WdVecEm_{cl,wvl}, 1\le cl\le n \\ 0 \qquad\qquad , n<cl \end{cases},$$ $BkWdMatrix_n$ 为第 n 个分词的第二输入矩阵。

13. 根据权利要求 12 所述的语义角色分析终端设备，其特征在于，所述第一神经网络模型的处理过程包括：

根据下式分别计算各个分词的第一复合向量：

$$FwWdCpVec_n=(FwWdCpEm_{n,1},FwWdCpEm_{n,2},.....,FwWdCpEm_{n,wvl},.....,FwWdCpEm_{n,wVecLen})$$

其中，

$$FwWdCpEm_{n,wvl}=\tanh\{\sum_{n=2}^{CoupLen}[FwWdEm_{n,wvl}*\frac{ln(n)}{\sum_{n=2}^{CoupLen}ln(n)}*FwWdWt_{wvl}]+FwWdEm_{1,wvl}*FwWdWt'_{wvl}\}$$

ln 为自然对数函数，tanh 为双曲正切函数，$FwWdWt_{wvl}$、$FwWdWt'_{wvl}$ 均为预设的权重系数；

根据下式分别计算各个词性类型的第一概率值：

$$FwWdProb_{n,m}=\frac{sigmoid(FwWdCpVec_n*FwWdWtVec_m^T)}{\sum_{m=1}^{M}sigmoid(FwWdCpVec_n*FwWdWtVec_m^T)}$$

其中，m 为词性类型的序号，$1\le m\le M$，M 为词性类型的个数，$FwWdWtVec_m$ 为预设的与第 m 个词性类型对应的权值向量，$sigmoid(x)=\frac{1}{1+e^{-x}}$，$FwWdProb_{n,m}$ 为第 n 个分词是第 m 个词性类型的第一概率值；

根据下式构建各个分词的第一输出向量：

$$FwWdVec_n=(FwWdProb_{n,1},FwWdProb_{n,2},.....,FwWdProb_{n,m},.....,FwWdProb_{n,M})$$

其中，$FwWdVec_n$ 为第 n 个分词的第一输出向量；

所述第二神经网络模型的处理过程包括：

根据下式分别计算各个分词的第二复合向量：

$$BkWdCpVec_n=(BkWdCpEm_{n,1},BkWdCpEm_{n,2},.....,BkWdCpEm_{n,wvl},.....,BkWdCpEm_{n,wVecLen})$$

其中，

$$BkWdCpEm_{n,wvl}=\tanh\{\sum_{n=2}^{CoupLen}[BkWdEm_{n,wvl}*\frac{ln(n)}{\sum_{n=2}^{CoupLen}ln(n)}*BkWdWt_{wvl}]+BkWdEm_{1,wvl}*BkWdWt'_{wvl}\}$$

$BkWdWt_{wvl}$、$BkWdWt'_{wvl}$ 均为预设的权重系数；

根据下式分别计算各个词性类型的第二概率值：

$$BkWdProb_{n,m}=\frac{sigmoid(BkWdCpVec_n*BkWdWtVec_m^T)}{\sum_{m=1}^{M}sigmoid(BkWdCpVec_n*BkWdWtVec_m^T)}$$

其中，$BkWdWtVec_m$ 为预设的与第 m 个词性类型对应的权值向量，$BkWdProb_{n,m}$ 为第 n 个分词是第 m 个词性类型的第二概率值；

根据下式构建各个分词的第二输出向量：

$$BkWdVec_n=(BkWdProb_{n,1},BkWdProb_{n,2},.....,BkWdProb_{n,m},.....,BkWdProb_{n,M})$$

其中，$BkWdVec_n$ 为第 n 个分词的第二输出向量。

14. 根据权利要求 13 所述的语义角色分析终端设备，其特征在于，所述根据各个分词的第一输出向量和第二输出向量确定各个分词的词性类型包括：

根据下式分别计算各个分词的词性概率向量：

$$WdProbVec_n=(WdProb_{n,1},WdProb_{n,2},.....,WdProb_{n,m},.....,WdProb_{n,M})$$

其中，$WdProb_{n,m}=\eta_1*FwWdProb_{n,m}+\eta_2*BkWdProb_{n,m}$，$\eta_1$、$\eta_2$ 均为预设的权重系数，$WdProbVec_n$ 为第 n 个分词的词性概率向量；

根据下式分别确定各个分词的词性类型：

$$CharSeq_n=\arg\max(WdProbVec_n)$$

其中，$\arg\max$ 为最大自变量函数，$CharSeq_n$ 为第 n 个分词的词性类型序号。

15. 根据权利要求 11 所述的语义角色分析终端设备，其特征在于，所述根据所述词性向量分别构建各个分词的第三输入矩阵和第四输入矩阵包括：

根据下式分别构建各个分词的第三输入矩阵：

$$FwCrMatrix_n=\begin{bmatrix} FwCrEm_{1,1} & FwCrEm_{1,2} & ... & FwCrEm_{1,cvl} & ... & FwCrEm_{1,cVecLen} \\ FwCrEm_{2,1} & FwCrEm_{2,2} & ... & FwCrEm_{2,cvl} & ... & FwCrEm_{2,cVecLen} \\ ... & ... & ... & ... & ... & ... \\ FwCrEm_{cl,1} & FwCrEm_{cl,2} & ... & FwCrEm_{cl,cvl} & ... & FwCrEm_{cl,cVecLen} \\ ... & ... & ... & ... & ... & ... \\ FwCrEm_{CoupLen,1} & FwCrEm_{CoupLen,2} & ... & FwCrEm_{CoupLen,cvl} & ... & FwCrEm_{CoupLen,cVecLen} \end{bmatrix}$$

其中，n 为所述分词按照前后顺序依次排列的序号，$1\le n\le N$，N 为所述语句文本的分词总数，cl 为所述第三输入矩阵的行号，$1\le cl\le CoupLen$，$CoupLen$ 为预设的耦合长度，cvl 为所述第三输入矩阵的列号，$1\le cvl\le cVecLen$，$cVecLen$ 为任意一个所述分词的词性向量的长度，第 n 个分词的词性向量为 $CharVec_n$，且

$CharVec_n=(CrVecEm_{n,1},CrVecEm_{n,2},.....,CrVecEm_{n,cvl},.....,CrVecEm_{n,cVecLen})$，

$FwCrEm_{cl,cvl}=\begin{cases} CrVecEm_{n-cl+1,cvl}, 1\le cl\le n \\ 0, n<cl \end{cases}$，$FwCrMatrix_n$ 为第 n 个分词的第三输入矩阵；

根据下式分别构建各个分词的第四输入矩阵：

$$BkCrMatrix_n=\begin{bmatrix} BkCrEm_{1,1} & BkCrEm_{1,2} & ... & BkCrEm_{1,cvl} & ... & BkCrEm_{1,cVecLen} \\ BkCrEm_{2,1} & BkCrEm_{2,2} & ... & BkCrEm_{2,cvl} & ... & BkCrEm_{2,cVecLen} \\ ... & ... & ... & ... & ... & ... \\ BkCrEm_{cl,1} & BkCrEm_{cl,2} & ... & BkCrEm_{cl,cvl} & ... & BkCrEm_{cl,cVecLen} \\ ... & ... & ... & ... & ... & ... \\ BkCrEm_{CoupLen,1} & BkCrEm_{CoupLen,2} & ... & BkCrEm_{CoupLen,cvl} & ... & BkCrEm_{CoupLen,cVecLen} \end{bmatrix}$$

$BkCrEm_{cl,cvl}=\begin{cases} CrVecEm_{cl,cvl}, 1\le cl\le n \\ 0, n<cl \end{cases}$，$BkCrMatrix_n$ 为第 n 个分词的第四输入矩阵。

16. 根据权利要求 15 所述的语义角色分析终端设备，其特征在于，所述第三神经网络模型的处理过程包括：

根据下式分别计算各个分词的第三复合向量：

$FwCrCpVec_n=(FwCrCpEm_{n,1},FwCrCpEm_{n,2},.....,FwCrCpEm_{n,vl},.....,FwCrCpEm_{n,cVecLen})$

其中，

$$FwCrCpEm_{n,cvl}=\tanh\{\sum_{n=2}^{CoupLen}[FwCrEm_{n,cvl}*\frac{ln(n)}{\sum_{n=2}^{CoupLen}ln(n)}*FwCrWt_{cvl}]+FwCrEm_{1,cvl}*FwCrWt'_{cvl}\}$$

ln 为自然对数函数，tanh 为双曲正切函数，$FwCrWt_{cvl}$、$FwCrWt'_{cvl}$ 均为预设的权重系数；

根据下式分别计算各个语义角色类型的第一概率值：

$$FwCrProb_{n,l}=\frac{sigmoid(FwCrCpVec_n*FwCrWtVec_l^T)}{\sum_{l=1}^{L}sigmoid(FwCrCpVec_n*FwCrWtVec_l^T)}$$

其中，l 为语义角色类型的序号，$1\le l\le L$，L 为语义角色类型的个数，$FwCrWtVec_l$ 为预设的与第 l 个语义角色类型对应的权值向量，$sigmoid(x)=\frac{1}{1+e^{-x}}$，$FwCrProb_{n,l}$ 为第 n 个分词是第 l 个语义角色类型的第一概率值；

根据下式构建各个分词的第三输出向量：

$$FwCrVec_n = (FwCrProb_{n,1}, FwCrProb_{n,2}, ..., FwCrProb_{n,l}, ..., FwCrProb_{n,L})$$

其中，$FwCrVec_n$ 为第 n 个分词的第三输出向量；

所述第四神经网络模型的处理过程包括：

根据下式分别计算各个分词的第四复合向量：

$$BkCrCpVec_n = (BkCrCpEm_{n,1}, BkCrCpEm_{n,2}, ..., BkCrCpEm_{n,cvl}, ..., BkCrCpEm_{n,cVecLen})$$ 其中，

$$BkCrCpEm_{n,cvl} = \tanh\{\sum_{n=2}^{CoupLen}[BkCrEm_{n,cvl} * \frac{ln(n)}{\sum_{n=2}^{CoupLen} ln(n)} * BkCrWt_{cvl}] + BkCrEm_{1,cvl} * BkCrWt'_{cvl}\}$$

$BkCrWt_{cvl}$、$BkCrWt'_{cvl}$ 均为预设的权重系数；

根据下式分别计算各个语义角色类型的第二概率值：

$$BkCrProb_{n,l} = \frac{sigmoid(BkCrCpVec_n * BkCrWtVec_l^T)}{\sum_{l=1}^{L} sigmoid(BkCrCpVec_n * BkCrWtVec_l^T)}$$

其中，$BkCrWtVec_l$ 为预设的与第 l 个语义角色类型对应的权值向量，$BkCrProb_{n,l}$ 为第 n 个分词是第 l 个语义角色类型的第二概率值；

根据下式构建各个分词的第四输出向量：

$$BkCrVec_n = (BkCrProb_{n,1}, BkCrProb_{n,2}, ..., BkCrProb_{n,l}, ..., BkCrProb_{n,L})$$

其中，$BkCrVec_n$ 为第 n 个分词的第四输出向量。

17. 根据权利要求 16 所述的语义角色分析终端设备，其特征在于，所述根据各个分词的第三输出向量和第四输出向量确定各个分词的语义角色类型包括：

根据下式分别计算各个分词的语义角色概率向量：

$$CrProbVec_n = (CrProb_{n,1}, CrProb_{n,2}, ..., CrProb_{n,l}, ..., CrProb_{n,L})$$

其中，$CrProb_{n,l} = \xi_1 * FwCrProb_{n,l} + \xi_2 * BkCrProb_{n,l}$，$\xi_1$、$\xi_2$ 均为预设的权重系数，$CrProbVec_n$ 为第 n 个分词的语义角色概率向量；

根据下式分别确定各个分词的语义角色类型：

$$RoleSeq_n = \arg\max(CrProbVec_n)$$

其中，$\arg\max$ 为最大自变量函数，$RoleSeq_n$ 为第 n 个分词的语义角色类型序号。

18. 一种语义角色分析装置，其特征在于，包括：

切词处理模块，用于对语句文本进行切词处理，得到构成所述语句文本的各个分词；

词向量查找模块，用于在预设的词向量数据库中分别查找各个分词的词向量，所述词向量数据库为记录词语与词向量之间的对应关系的数据库；

词向量矩阵构建模块，用于根据所述词向量分别构建各个分词的第一输入矩阵和第二输入矩阵；

第一处理模块，用于将各个分词的第一输入矩阵分别输入到预设的第一神经网络模型中，得到各个分词的第一输出向量，所述第一神经网络模型为进行正序词性分析的神经网络模型；

第二处理模块，用于将各个分词的第二输入矩阵分别输入到预设的第二神经网络模型中，得到各个分词的第二输出向量，所述第二神经网络模型为进行逆序词性分析的神经网络模型；

词性类型确定模块，用于根据各个分词的第一输出向量和第二输出向量确定各个分词的词性类型；

词性向量查找模块，用于在预设的词性向量数据库中分别查找各个分词的词性类型对应的词性向量，所述词性向量数据库为记录词性类型与词性向量之间的对应关系的数据库；

词性向量矩阵构建模块，用于根据所述词性向量分别构建各个分词的第三输入矩阵和第四输入矩阵；

第三处理模块，用于将各个分词的第三输入矩阵分别输入到预设的第三神经网络模型中，得到各个分词的第三输出向量，所述第三神经网络模型为进行正序语义角色分析的神经网络模型；

第四处理模块，用于将各个分词的第四输入矩阵分别输入到预设的第四神经网络模型中，得到各个分词的第四输出向量，所述第四神经网络模型为进行逆序语义角色分析的神经网络模型；

语义角色类型确定模块，用于根据各个分词的第三输出向量和第四输出向量确定各个分词的语义角色类型。

19. 根据权利要求 18 所述的语义角色分析装置，其特征在于，所述词向量矩阵构建模块包括：

第一输入矩阵构建单元，用于根据下式分别构建各个分词的第一输入矩阵：

$$FwWdMatrix_n=\begin{bmatrix} FwWdEm_{1,1} & FwWdEm_{1,2} & ... & FwWdEm_{1,wvl} & ... & FwWdEm_{1,wVecLen} \\ FwWdEm_{2,1} & FwWdEm_{2,2} & ... & FwWdEm_{2,wvl} & ... & FwWdEm_{2,wVecLen} \\ ... & ... & ... & ... & ... & ... \\ FwWdEm_{cl,1} & FwWdEm_{cl,2} & ... & FwWdEm_{cl,wvl} & ... & FwWdEm_{cl,wVecLen} \\ ... & ... & ... & ... & ... & ... \\ FwWdEm_{CoupLen,1} & FwWdEm_{CoupLen,2} & ... & FwWdEm_{CoupLen,wvl} & ... & FwWdEm_{CoupLen,wVecLen} \end{bmatrix}$$

其中，n 为所述分词按照前后顺序依次排列的序号，$1\le n\le N$，N 为所述语句文本的分词总数，cl 为所述第一输入矩阵的行号，$1\le cl\le CoupLen$，$CoupLen$ 为预设的耦合长度，wvl 为所述第一输入矩阵的列号，$1\le wvl\le wVecLen$，$wVecLen$ 为任意一个所述分词的词向量的长度，第 n 个分词的词向量为 $WordVec_n$，且

$WordVec_n=(WdVecEm_{n,1},WdVecEm_{n,2},.....,WdVecEm_{n,vl},.....,WdVecEm_{n,wVecLen})$，

$FwWdEm_{cl,wvl}=\begin{cases} WdVecEm_{n-cl+1,wvl}, 1\le cl\le n \\ 0 \qquad\qquad ,n<cl \end{cases}$，$FwWdMatrix_n$ 为第 n 个分词的第一输入矩阵；

第二输入矩阵构建单元，用于根据下式分别构建各个分词的第二输入矩阵：

$$BkWdMatrix_n=\begin{bmatrix} BkWdEm_{1,1} & BkWdEm_{1,2} & ... & BkWdEm_{1,wvl} & ... & BkWdEm_{1,wVecLen} \\ BkWdEm_{2,1} & BkWdEm_{2,2} & ... & BkWdEm_{2,wvl} & ... & BkWdEm_{2,wVecLen} \\ ... & ... & ... & ... & ... & ... \\ BkWdEm_{cl,1} & BkWdEm_{cl,2} & ... & BkWdEm_{cl,wvl} & ... & BkWdEm_{cl,wVecLen} \\ ... & ... & ... & ... & ... & ... \\ BkWdEm_{CoupLen,1} & BkWdEm_{CoupLen,2} & ... & BkWdEm_{CoupLen,wvl} & ... & BkWdEm_{CoupLen,wVecLen} \end{bmatrix}$$

$BkWdEm_{cl,wvl}=\begin{cases} WdVecEm_{cl,wvl}, 1\le cl\le n \\ 0 \qquad\qquad ,n<cl \end{cases}$，$BkWdMatrix_n$ 为第 n 个分词的第二输入矩阵。

20. 根据权利要求 19 所述的语义角色分析装置，其特征在于，所述第一处理模块包括：

第一复合向量计算单元，用于根据下式分别计算各个分词的第一复合向量：

$FwWdCpVec_n=(FwWdCpEm_{n,1},FwWdCpEm_{n,2},.....,FwWdCpEm_{n,wvl},.....,FwWdCpEm_{n,wVecLen})$

其中，

$$FwWdCpEm_{n,wvl}=\tanh\{\sum_{n=2}^{CoupLen}[FwWdEm_{n,wvl}*\frac{ln(n)}{\sum_{n=2}^{CoupLen}ln(n)}*FwWdWt_{wvl}]+FwWdEm_{1,wvl}*FwWdWt'_{wvl}\}$$

ln 为自然对数函数，tanh 为双曲正切函数，$FwWdWt_{wvl}$、$FwWdWt'_{wvl}$ 均为预设的权重系数；

词性第一概率值计算单元，用于根据下式分别计算各个词性类型的第一概率值：

$$FwWdProb_{n,m}=\frac{sigmoid(FwWdCpVec_n*FwWdWtVec_m^T)}{\sum_{m=1}^{M}sigmoid(FwWdCpVec_n*FwWdWtVec_m^T)}$$

其中，m 为词性类型的序号，$1 \le m \le M$，M 为词性类型的个数，$FwWdWtVec_m$ 为预设的与第 m 个词性类型对应的权值向量，$sigmoid(x) = \frac{1}{1+e^{-x}}$，$FwWdProb_{n,m}$ 为第 n 个分词是第 m 个词性类型的第一概率值；

第一输出向量构建单元，用于根据下式构建各个分词的第一输出向量：

$$FwWdVec_n = (FwWdProb_{n,1}, FwWdProb_{n,2}, ..., FwWdProb_{n,m}, ..., FwWdProb_{n,M})$$

其中，$FwWdVec_n$ 为第 n 个分词的第一输出向量；

所述第二处理模块包括：

第二复合向量计算单元，用于根据下式分别计算各个分词的第二复合向量：

$$BkWdCpVec_n = (BkWdCpEm_{n,1}, BkWdCpEm_{n,2}, ..., BkWdCpEm_{n,wvl}, ..., BkWdCpEm_{n,wVecLen})$$

其中，

$$BkWdCpEm_{n,wvl} = \tanh\{\sum_{n=2}^{CoupLen}[BkWdEm_{n,wvl} * \frac{ln(n)}{\sum_{n=2}^{CoupLen} ln(n)} * BkWdWt_{wvl}] + BkWdEm_{1,wvl} * BkWdWt'_{wvl}\}$$

$BkWdWt_{wvl}$、$BkWdWt'_{wvl}$ 均为预设的权重系数；

词性第二概率值计算单元，用于根据下式分别计算各个词性类型的第二概率值：

$$BkWdProb_{n,m} = \frac{sigmoid(BkWdCpVec_n * BkWdWtVec_m^T)}{\sum_{m=1}^{M} sigmoid(BkWdCpVec_n * BkWdWtVec_m^T)}$$

其中，$BkWdWtVec_m$ 为预设的与第 m 个词性类型对应的权值向量，$BkWdProb_{n,m}$ 为第 n 个分词是第 m 个词性类型的第二概率值；

第二输出向量构建单元，用于根据下式构建各个分词的第二输出向量：

$$BkWdVec_n = (BkWdProb_{n,1}, BkWdProb_{n,2}, ..., BkWdProb_{n,m}, ..., BkWdProb_{n,M})$$

其中，$BkWdVec_n$ 为第 n 个分词的第二输出向量。

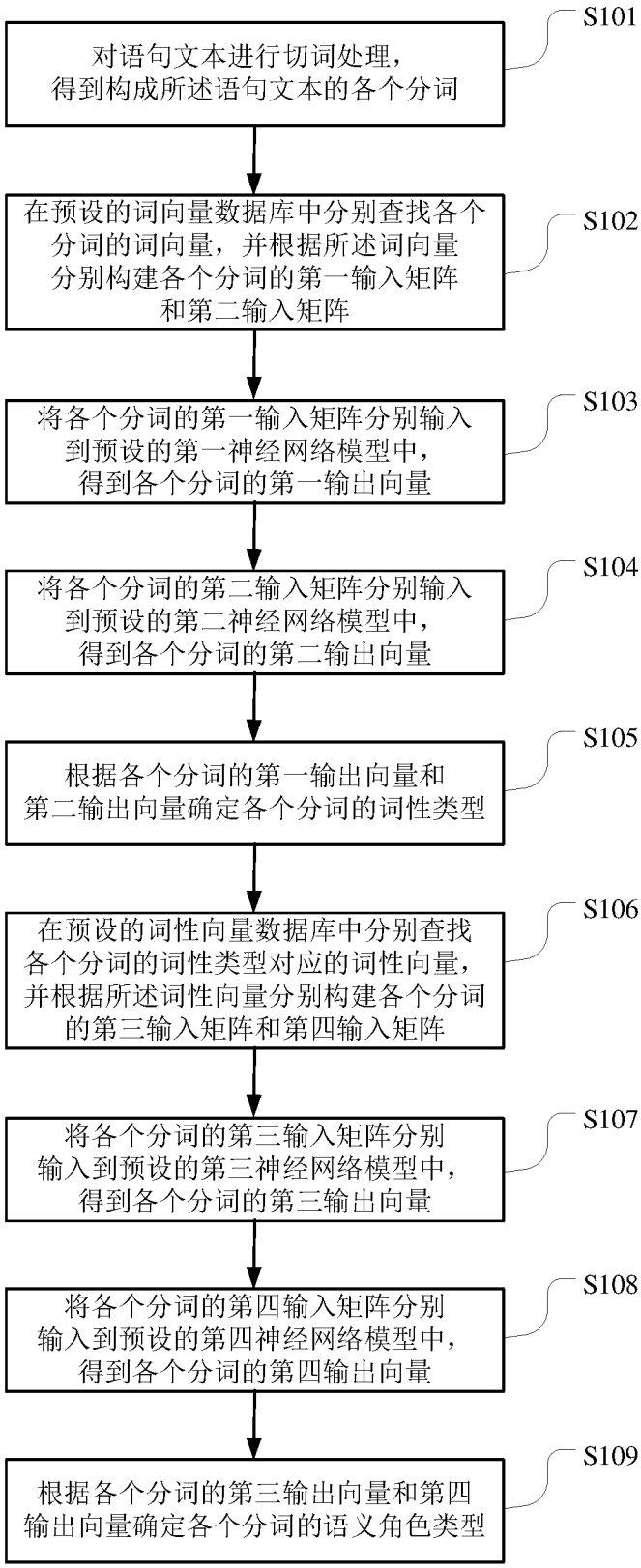

对语句文本进行切词处理，得到构成所述语句文本的各个分词
S101
在预设的词向量数据库中分别查找各个分词的词向量，并根据所述词向量分别构建各个分词的第一输入矩阵和第二输入矩阵
S102
将各个分词的第一输入矩阵分别输入到预设的第一神经网络模型中，得到各个分词的第一输出向量
S103
将各个分词的第二输入矩阵分别输入到预设的第二神经网络模型中，得到各个分词的第二输出向量
S104
根据各个分词的第一输出向量和第二输出向量确定各个分词的词性类型
S105
在预设的词性向量数据库中分别查找各个分词的词性类型对应的词性向量，并根据所述词性向量分别构建各个分词的第三输入矩阵和第四输入矩阵
S106
将各个分词的第三输入矩阵分别输入到预设的第三神经网络模型中，得到各个分词的第三输出向量
S107
将各个分词的第四输入矩阵分别输入到预设的第四神经网络模型中，得到各个分词的第四输出向量
S108
根据各个分词的第三输出向量和第四输出向量确定各个分词的语义角色类型
S109

图 1

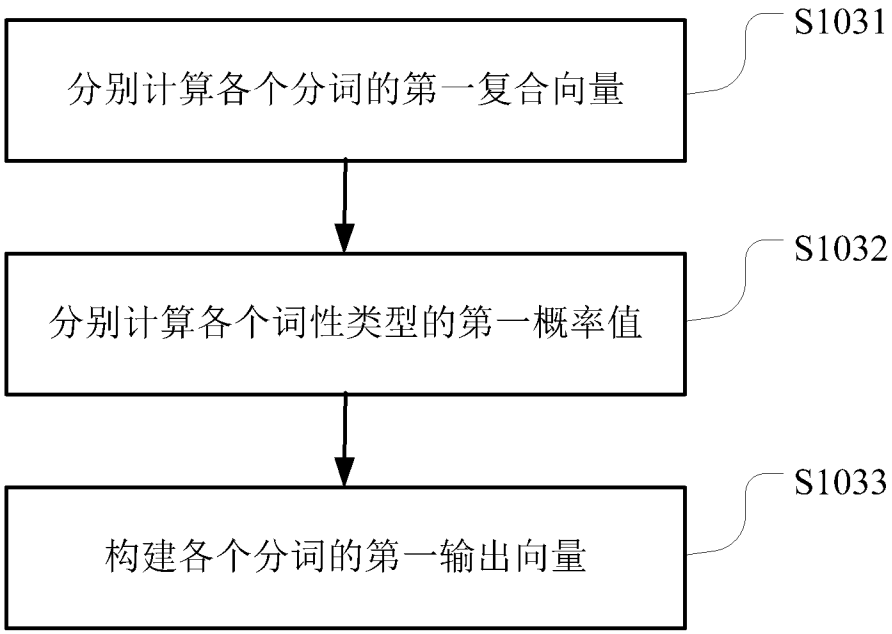

S1031
分别计算各个分词的第一复合向量
S1032
分别计算各个词性类型的第一概率值
S1033
构建各个分词的第一输出向量

图 2

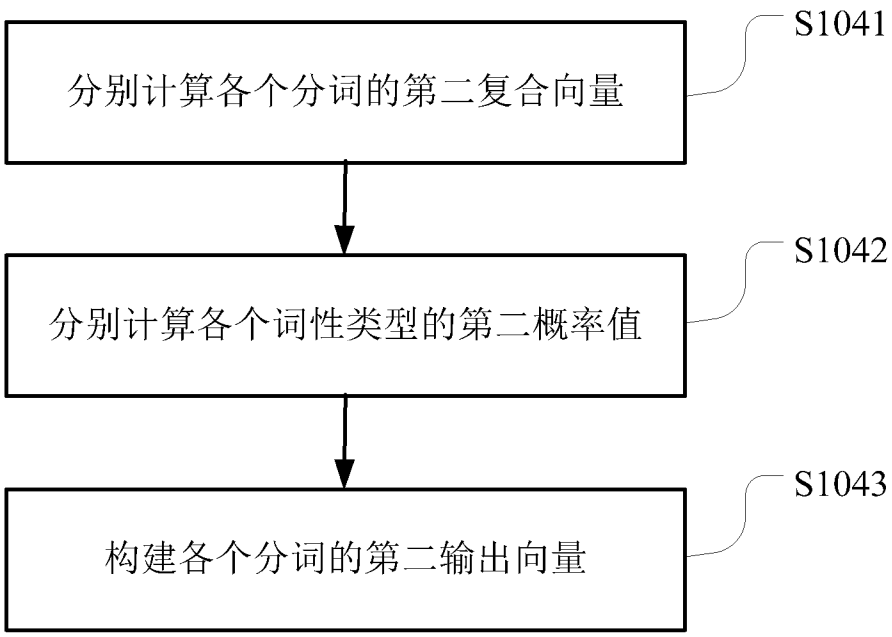

S1041
分别计算各个分词的第二复合向量
S1042
分别计算各个词性类型的第二概率值
S1043
构建各个分词的第二输出向量

图 3

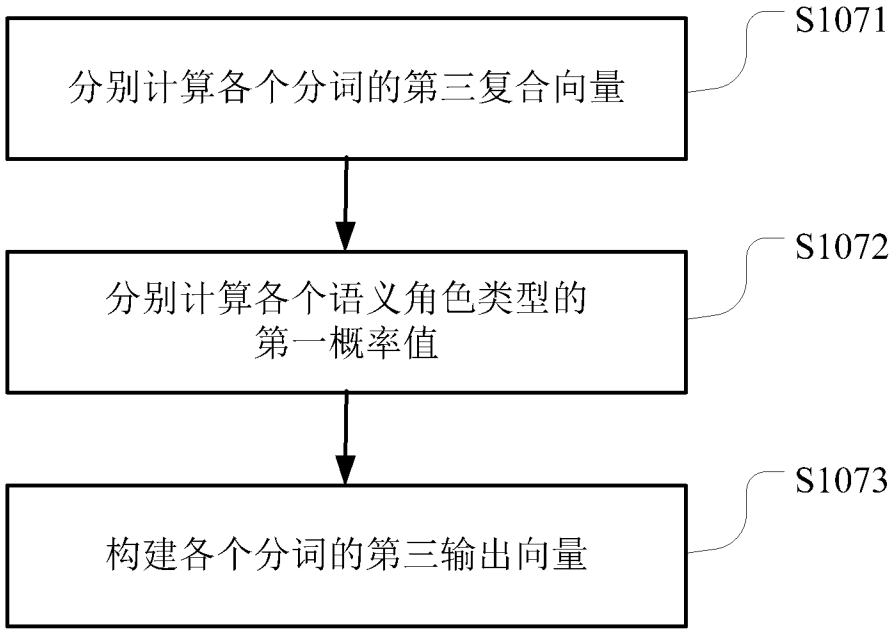

S1071
分别计算各个分词的第三复合向量
S1072
分别计算各个语义角色类型的
第一概率值
S1073
构建各个分词的第三输出向量

图 4

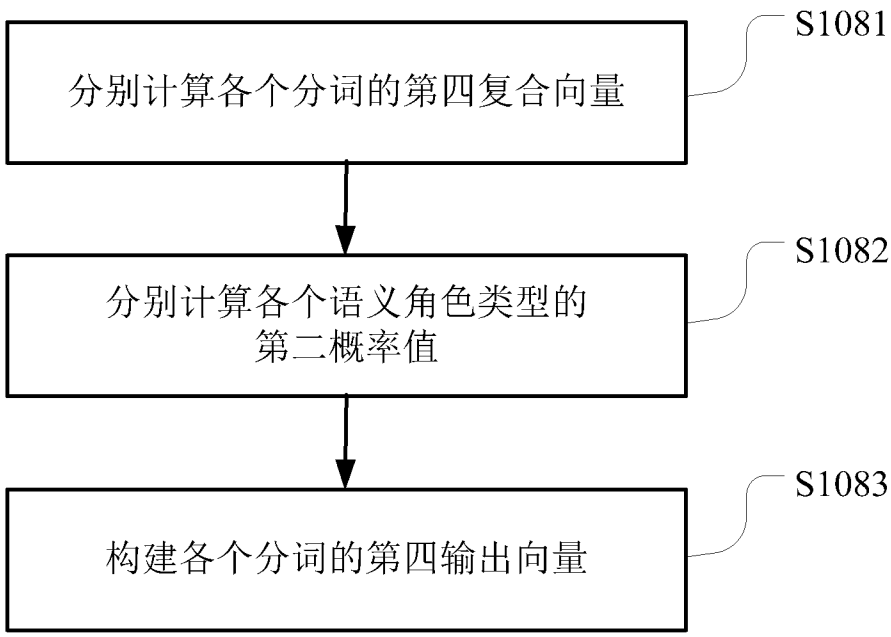

S1081
分别计算各个分词的第四复合向量
S1082
分别计算各个语义角色类型的
第二概率值
S1083
构建各个分词的第四输出向量

图 5

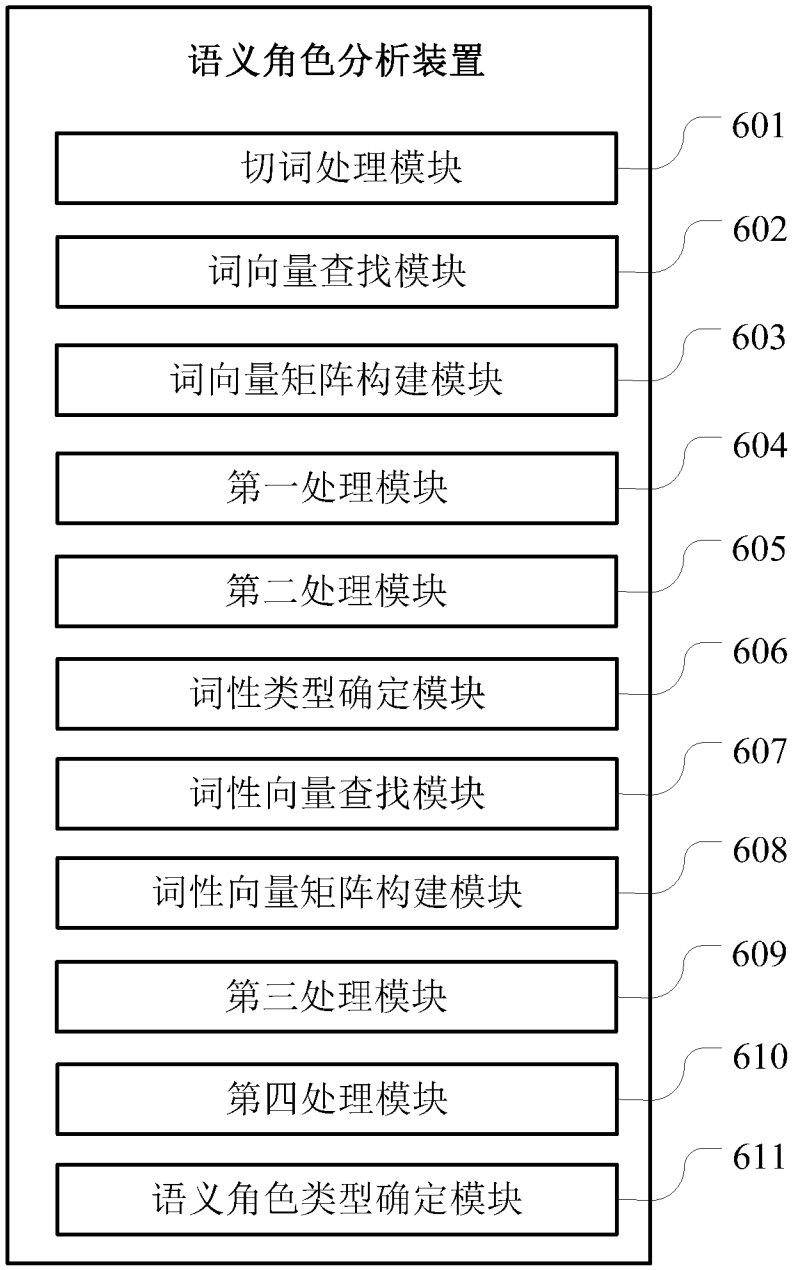

语义角色分析装置
切词处理模块
601
词向量查找模块
602
词向量矩阵构建模块
603
第一处理模块
604
第二处理模块
605
词性类型确定模块
606
词性向量查找模块
607
词性向量矩阵构建模块
608
第三处理模块
609
第四处理模块
610
语义角色类型确定模块
611

图 6

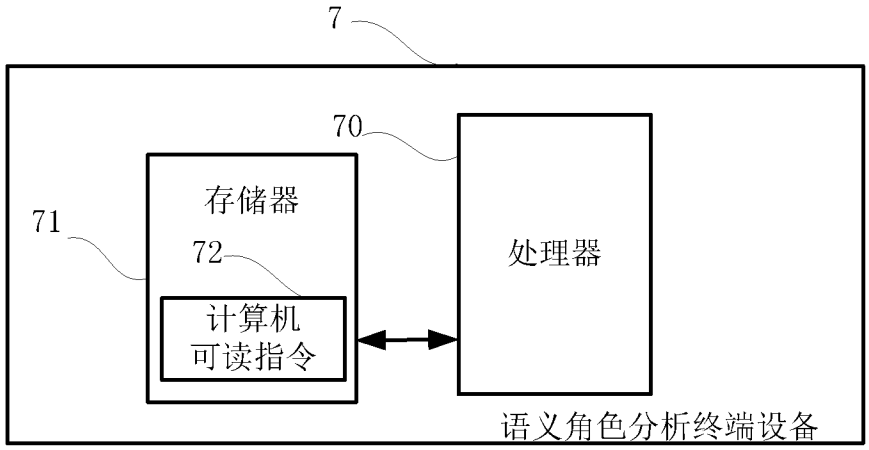

7
70
存储器
71
72
计算机
可读指令
处理器
语义角色分析终端设备

图 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/096258

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)

CNTXT, CNABS, CNKI, SIPOABS, DWPI: 矩阵, 神经网络, 分词, 词性, 角色, 向量, 分析, 语义, matrix, neural network, devide, word, part of speech, role, vector, analys+, semantic

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 104462066 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 25 March 2015 (2015-03-25) entire document	1-20
A	CN 102662931 A (XIAMEN UNIVERSITY) 12 September 2012 (2012-09-12) 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	Date of mailing of the international search report
07 December 2018	**11 January 2019**

Name and mailing address of the ISA/CN	Authorized officer
State Intellectual Property Office of the P. R. China **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing** **100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/096258

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 104462066 A	25 March 2015	CN 104462066 B	03 October 2017
CN 102662931 A	12 September 2012	CN 102662931 B	25 March 2015

国际检索报告

国际申请号
PCT/CN2018/096258

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））

CNTXT,CNABS,CNKI,SIPOABS,DWPI:矩阵，神经网络，分词，词性，角色，向量，分析，语义,matrix, neural network,devide, word, part of speech, role, vector, analys+, semantic

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
A	CN 104462066 A (北京百度网讯科技有限公司) 2015年 3月 25日 (2015 - 03 - 25) 全文	1-20
A	CN 102662931 A (厦门大学) 2012年 9月 12日 (2012 - 09 - 12) 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	国际检索报告邮寄日期
2018年 12月 7日	2019年 1月 11日

ISA/CN的名称和邮寄地址	受权官员
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	崔丽艳
传真号 (86-10)62019451	电话号码 (86-10) 62089285

表 PCT/ISA/210 (第2页) (2015年1月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/096258

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	104462066	A	2015年 3月 25日	CN	104462066	B	2017年 10月 3日
CN	102662931	A	2012年 9月 12日	CN	102662931	B	2015年 3月 25日