(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 111444726 A
(43)申请公布日 2020.07.24

(21)申请号 202010228609.X

(22)申请日 2020.03.27

(71)申请人 河海大学常州校区
地址 213022 江苏省常州市晋陵北路200号

(72)发明人 徐宁 于佳卉 刘小峰 姚潇 蒋爱民

(74)专利代理机构 南京纵横知识产权代理有限公司 32224
代理人 张倩倩

(51)Int.Cl.
G06F 40/30(2020.01)
G06F 40/295(2020.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

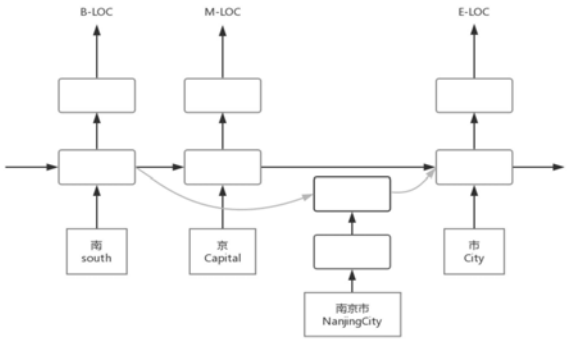
权利要求书4页 说明书11页 附图4页

(54)发明名称

基于双向格子结构的长短时记忆网络的中文语义信息提取方法和装置

(57)摘要

本发明公开一种语义信息提取方法和装置，属于自然语言处理技术领域，方法包括：获取待识别的语料数据；对获取到的语料数据进行预处理，预处理包括将语料数据转换为词向量和/或字向量；将向量转换后的语料信息输入至预先训练的语义信息提取模型，得到命名实体识别结果；所述语义信息提取模型包括双向长短时记忆网络和CRF层网络，其训练样本为已标注字符标签和实体标签的语料数据的向量形式；双向长短时记忆网络的输出为待识别语句中各词中字符映射到标签的概率矩阵，CRF层网络根据双向长短时记忆网络的输出，确定待识别语句的标签序列并输出。本发明通过将格子结构的长短时记忆网络改进为双向，使其能够更好的获知文章中一个句子前后文的信息，从而更准确地判断这个句子的语义。



CN 111444726 A

1.一种语义信息提取方法,其特征是,包括:

获取待识别的语料数据;

对获取到的语料数据进行预处理,预处理包括将语料数据转换为词向量和/或字向量;

将向量转换后的语料信息输入至预先训练的语义信息提取模型,得到命名实体识别结果;所述语义信息提取模型包括双向长短时记忆网络和CRF层网络,其训练样本为已标注字符标签和实体标签的语料数据的向量形式;双向长短时记忆网络的输出为待识别语句中各词中字符映射到标签的概率矩阵,CRF层网络根据双向长短时记忆网络的输出确定待识别语句的标签序列并输出。

2.根据权利要求1所述的方法,其特征是,所述待识别的语料数据为中文语句文本。

3.根据权利要求1所述的方法,其特征是,对获取到的语料信息进行预处理还包括数据清洗。

4.根据权利要求1所述的方法,其特征是,对获取到的语料数据进行预处理时,将待识别语料与预设的单词查找树进行匹配,得到相应的单词集合,进而采用嵌入层Embedding将语料数据转换为词向量和字向量。

5.根据权利要求1所述的方法,其特征是,语义信息提取模型的训练包括:

样本语料标注:对多个样本语句进行标注处理,标注出各样本语句中的字符标签;

对标注后的样本语料进行预处理,抽取得到训练样本语句,及其对应的标签序列和单词集合;

利用训练样本对双向格子结构的长短时记忆网络进行训练,以调整其网络参数;

基于训练样本利用双向格子结构的长短时记忆网络的输出对CRF层网络进行训练,以调整其网络参数;

得到训练完成的语义信息提取模型。

6.根据权利要求5所述的方法,其特征是,语义信息提取模型训练还包括根据训练样本及训练过程中的识别结果计算准确率P和召回率R,并根据以下公式计算评价分数F1:

$$F1=\frac{2*P*R}{P+R}$$

响应于评价分数值大于预设值,则停止模型训练。

7.根据权利要求5所述的方法,其特征是,对标注后的样本语料进行预处理包括:

(2.1)统计已标注样本语料的字符,得到字符集合,然后对每个字符编号,得到字符集合对应的字符编号集合;统计已标注样本语料的标签,得到标签集合,然后对每个标签编号,得到标签集合对应的标签编号集合;

(2.2)基于汉语词典建立单词查找树,将各语句与单词查找树进行匹配,保留匹配成功的单词,得到样本语料对应的单词集合;

(2.3)对单词集合中的单词去重处理并编号,得到新的单词集合及其对应的词编号集合;

(2.4)将样本语料中的字符和标签分别根据字符编号集合和标签编号集合转换为对应编号;同时将各语句中的各单词根据词编号集合转换为对应的编号;

模型训练时,将转换编号后的样本语料随机排列,采用随机无放回的方式从样本语料中抽取若干语句及其对应的标签和对应的单词集合,进行向量转换后,作为双向格子结构

的长短时记忆网络的输入。

8.根据权利要求1或5所述的方法，其特征是，可选的，双向格子结构的长短时记忆网络的隐藏层包括前向网络层和反向网络层，前向网络层和反向网络层分别设置字处理网络单元和词处理网络单元；字处理网络单元包括输入门、输出门和遗忘门，词处理网络单元包括输入门和遗忘门；

字处理网络单元的输入量包括当前字符的字符向量、前一个字符的细胞状态和字处理网络单元的输出，以及以当前字符为末位字符的单词从词处理网络单元输出的细胞状态；词处理网络单元的输入包括当前单词向量，以及当前单词的首位字符在字处理网络单元的输出和细胞状态；

双向长短时记忆网络的输出为待识别语句中各词中字符映射到标签的非归一化概率矩阵，CRF层网络采用维特比算法根据双向长短时记忆网络的输出，确定待识别语句的标签序列。。

9.根据权利要求8所述的方法，其特征是，定义待识别中文文本的字符序列为S＝{c_1，c_2，c_3……，c_m}，序列S与单词查找树匹配得到的单词集合为$w_{b,e}^{d}$，表示从B到E结束的单词子序列；

嵌入层对于字符和单词分别按照以下公式进行向量转换：

$$x_i^c = e^c\left(c_i\right), i \in \{1,2,3...m\}$$

$$x_{b,e}^{w} = e^{w}\left(w_{b,e}^{d}\right)$$

其中，x_i^c为字符c_i经嵌入层转换后得到的字符向量，$x_{b,e}^{w}$为单词集合$w_{b,e}^{d}$的词向量，e^c、e^w代表嵌入层的权重；

用i_i^c、o_i^c、f_i^c分别表示字符向量处理网络单元的输入门、输出门和遗忘门的控制，σ、tanh表示激活函数sigmoid和tanh，字符向量处理网络单元对于输入的字符向量x_i^c按照下式进行处理：

$$\begin{bmatrix} i_i^c \\ o_i^c \\ f_i^c \\ \tilde{c}_i^c \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ tanh \end{bmatrix}\left(W^{c^T}\begin{bmatrix} x_i^c \\ h_{i-1}^c \end{bmatrix} + b^c\right)$$

$$c_i^c = f_i^c \odot c_{i-1}^c + i_i^c \odot \tilde{c}_i^c$$

$$\vec{h}_i^c = \overleftarrow{h}_i^c = o_i^c \odot tanh\left(c_i^c\right)$$

$$h_i^c = \left[\vec{h}_i^c;\ \ \overleftarrow{h}_i^c\right]$$

$$\tilde{c}_i^c = tanh\left(W_C \cdot \left[h_{i-1}{}^c, x_i^c\right] + b_C\right)$$

式中，$\tilde{c}_i^c$表示经tanh函数处理后的细胞状态，为中间信息状态；h_{i-1}^c表示前一个字符的字符向量处理网络单元输出；$\vec{h}_i^c$和$\overleftarrow{h}_i^c$分别表示前向和反向两个方向的输出，h_i^c为结合两个

方向的最后的输出；c_{i-1}^{c}表示从前一个字符及其相关的词传过来的细胞状态；W_c表示字处理单网络元的权重矩阵，W^{c^T}表示W_c的转置矩阵；b^c表示字处理网络单元中的常数项；⊙表示矩阵点积；

用$i_{b,e}^{w}$、$f_{b,w}^{w}$分别表示词向量处理网络单元中的输入门和遗忘门的控制，词向量处理网络单元对于输入的词向量$x_{b,e}^{w}$按照下式进行处理：

$$\begin{bmatrix} i_{b,e}^{w} \\ f_{b,w}^{w} \\ \tilde{c}_{b,e}^{w} \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ tanh \end{bmatrix} \left(W^{w^T} \begin{bmatrix} x_{b,e}^{w} \\ h_b^c \end{bmatrix} + b^w \right)$$

$$c_{b,e}^{w} = f_{b,e}^{w} \odot c_b^c + i_{b,e}^{w} \odot \tilde{c}_{b,e}^{w}$$

式中，$c_{b,e}^{w}$表示从B开始到E结束的词的细胞状态，$\tilde{c}_{b,e}^{w}$表示经tanh函数处理后的细胞状态，为中间信息状态；h_b^c表示第B个字在字处理网络单元的输出；W^{w^T}表示词处理网络单元权重矩阵的转置；b^w表示词处理网络单元的常数项；

字处理网络单元中，对应字符向量x_j^c的输出细胞状态c_j^c按照下式计算：

$$c_j^c = \sum_{b \in \left\{ b' | w_{b'}^{d}, j \in D \right\}} \alpha_{b,j}^{c} \odot c_{b,j}^{w} + \alpha_j^c \odot \tilde{c}_j^c$$

其中，$c_{b,j}^{w}$为从b到j组成的单词的细胞状态，$\alpha_{b,j}^{c}$为从b到j组成的单词的细胞状态的权重，α_j^c为第j个字的细胞状态的权重，$\tilde{c}_j^c$为对应x_i^c在字处理网络单元中经tanh函数处理后的细胞状态，$b \in \left\{ b' | w_{b'}^{d}, j \in D \right\}$中，b'代表所有可能的b集合，$w_{b'}^{d}$表示从b'到d组成的词，D表示所规定函数的定义域；

并有：

$$\alpha_{b,j}^{c} = \frac{\exp\left(i_{b,j}^{c}\right)}{\exp\left(i_j^c\right) + \sum_{b'' \in \left\{ b'' | w_{b''}^{d}, j \in D \right\}} \exp\left(i_{b'',j}^{c}\right)}$$

$$\alpha_j^c = \frac{\exp\left(i_j^c\right)}{\exp\left(i_j^c\right) + \sum_{b'' \in \left\{ b'' | w_{b''}^{d}, j \in D \right\}} \exp\left(i_{b'',j}^{c}\right)}$$

上式中，$i_{b,j}^{c}$表示表示从b到j组成的词的输入门，i_j^c表示第j个字的输入门，$w_{b''}^{d}$表示从b″到d组成的词，b″表示所有可能的b'集合。

10.一种采用权利要求1-9任一项语义信息提取方法的语义信息提取装置，其特征是，包括：

语料数据获取模块，被配置用于获取待识别的语料数据；

预处理模块，被配置用于对获取到的语料数据进行预处理，预处理包括将语料数据转换为词向量和/或字向量；

语义信息提取模块，用于将向量转换后的语料信息输入至预先训练的语义信息提取模型，得到命名实体识别结果；所述语义信息提取模型包括双向长短时记忆网络和CRF层网

络,其训练样本为已标注字符标签和实体标签的语料数据的向量形式;双向长短时记忆网络的输出为待识别语句中各词中字符映射到标签的概率矩阵,CRF层网络根据双向长短时记忆网络的输出,确定待识别语句的标签序列并输出。

基于双向格子结构的长短时记忆网络的中文语义信息提取方法和装置

技术领域

[0001] 本发明涉及自然语言处理技术领域，特别是一种基于双向格子结构的长短时记忆神经网络的中文语义信息提取方法和装置。

背景技术

[0002] 信息抽取是指从自然语言文本中抽取指定类型的实体、关系、事件等信息，并且形成结构化数据输出的文本处理技术。它是自然语言处理领域经常用到的一项技术，也是该领域研究的重点技术之一。信息抽取的任务有实体识别与抽取、实体消歧、关系抽取、事件抽取，其中实体识别是信息抽取的主要任务之一，意义十分重大。

[0003] 对于实体识别这一任务，目前已经存在一些方法，比如，基于规则的命名实体识别方法，该方法是定义一个规则，将语料和规则进行匹配，从而识别出实体；基于词典的命名实体识别方法，该方法是将尽量多的实体建立词典，经过训练使文本中的词与词典中的词相匹配，匹配上的即为该词典中对应分类的实体；基于传统机器学习的命名实体识别的方法，该方法是对文本信息提取特征，从而学习前后词的语义信息，做出相应的分类；基于长短时记忆网络-条件随机场(Long Short Term Memory-Conditional Random Field,LSTM-CRF)的命名实体识别的方法，该方法主要有两种，基于词向量和基于字符向量，主要是对前一种方法的改进，即令长短时记忆网络LSTM进行提取特征，令CRF层进行分类判断。

[0004] 上述方法存在以下缺陷：

[0005] 1、基于规则的命名实体识别方法比较死板，并且规则太多，费时费力；

[0006] 2、基于词典的命名实体识别方法十分依赖于词典库，并且不能识别未登录词；

[0007] 3、基于传统机器学习的命名实体识别方法的特征模板需要人工提取，耗时耗力，且建立的模板质量十分影响识别效果；

[0008] 4、基于LSTM-CRF的命名实体识别方法，基于词向量的方法十分依赖分词效果，即若分词错误则影响识别；而基于字符向量的方法虽优于基于词向量的方法，但是它不能充分利用单词和词序信息，也会影响识别效果。

发明内容

[0009] 本发明的目的是，提供基于双向格子结构的长短时记忆网络的中文语义信息提取方法和装置，提高语义识别准确度。

[0010] 本发明采取的技术方案如下。

[0011] 一方面，本发明提供一种语义信息提取方法，包括：

[0012] 获取待识别的语料数据；

[0013] 对获取到的语料数据进行预处理，预处理包括将语料数据转换为词向量和/或字向量；

[0014] 将向量转换后的语料信息输入至预先训练的语义信息提取模型，得到命名实体识

别结果;所述语义信息提取模型包括双向长短时记忆网络和CRF层网络,其训练样本为已标注字符标签和实体标签的语料数据的向量形式;双向长短时记忆网络的输出为待识别语句中各词中字符映射到标签的概率矩阵,CRF层网络根据双向长短时记忆网络的输出确定待识别语句的标签序列并输出。

[0015] 本发明将传统格子结构的长短时记忆网络Lattice LSTM由单向改进为双向,在训练和识别时不仅能够充分利用单词和词序信息,不会因为分词错误影响识别结果,且能够更好的联系上下文的信息,使得机器如人工智能问答系统,能够更好的理解词在语句中的具体意思,进而针对识别出的实体回答相应问题。

[0016] 可选的,所述待识别的语料数据为中文语句文本。也即本发明适用于中文语义信息的提取。在问答系统中,可首先将获取到的用户语音数据转换文中文语句文本,然后进行语义提取。语料数据源可根据自然语言识别所应用的领域来决定,如医疗领域,可通过爬虫三九健康网、寻医问药网等网站获取语料文本数据。

[0017] 可选的,对获取到的语料信息进行预处理还包括数据清洗。如过滤噪声数据,可采用现有技术。

[0018] 可选的,对获取到的语料数据进行预处理时,将待识别语料与预设的单词查找树进行匹配,得到相应的单词集合,进而采用嵌入层Embedding将语料数据转换为词向量和字向量。Embedding层可采用现有的word2vec工具实现语料数据到向量的转换。单词查找树可根据汉语词典设置,用于待识别语句与汉语词典词库之间的匹配,已查找到待识别语句所包含的实体单词。

[0019] 可选的,语义信息提取模型的训练包括:

[0020] 样本语料标注:对多个样本语句进行标注处理,标注出各样本语句中的字符标签;

[0021] 对标注后的样本语料进行预处理,抽取得到训练样本语句,及其对应的标签序列和单词集合;

[0022] 利用训练样本对双向格子结构的长短时记忆网络进行训练,以调整其网络参数;

[0023] 基于训练样本及双向格子结构的长短时记忆网络的输出对CRF层网络进行训练,以调整其网络参数;

[0024] 得到训练完成的语义信息提取模型。

[0025] 可选的,语义信息提取模型训练还包括根据训练样本及训练过程中的识别结果计算准确率P和召回率R,并根据以下公式计算评价分数F1:

[0026] $$F1=\frac{2*P*R}{P+R}$$

[0027] 响应于评价分数值大于预设值,则停止模型训练。

[0028] 可选的,样本语料标注采用BMESO(begin,middle,end,single,other)标记方法。即位于该词语最开始的字符标记为B,位于该词语中间的字符标记为M,位于该词语末尾的词语标记为E,若该词语只有一个字符则标记为S,若该词语没有带标签或者不属于实体标签则标记为O。

[0029] 可选的,对标注后的样本语料进行预处理包括:

[0030] (2.1)统计已标注样本语料的字符,得到字符集合,然后对每个字符编号,得到字符集合对应的字符编号集合;统计已标注样本语料的标签,得到标签集合,然后对每个标签

编号,得到标签集合对应的标签编号集合;

[0031] (2.2)基于汉语词典建立单词查找树,将各语句与单词查找树进行匹配,保留匹配成功的单词,得到样本语料对应的单词集合;

[0032] (2.3)对单词集合中的单词去重处理并编号,得到新的单词集合及其对应的词编号集合;

[0033] (2.4)将样本语料中的字符和标签分别根据字符编号集合和标签编号集合转换为对应编号;同时将各语句中的各单词根据词编号集合转换为对应的编号。

[0034] 模型训练时,将转换编号后的样本语料随机排列,采用随机无放回的方式从样本语料中抽取若干语句及其对应的标签和对应的单词集合,进行向量转换后,作为双向格子结构的长短时记忆网络的输入。

[0035] 可选的,双向格子结构的长短时记忆网络的隐藏层包括前向网络层和反向网络层,前向网络层和反向网络层分别设置字处理网络单元和词处理网络单元;字处理网络单元包括输入门、输出门和遗忘门,词处理网络单元包括输入门和遗忘门;

[0036] 字处理网络单元的输入量包括当前字符的字符向量、前一个字符的细胞状态和字处理网络单元的输出,以及以当前字符为末位字符的单词从词处理网络单元输出的细胞状态;词处理网络单元的输入包括当前单词向量,以及当前单词的首位字符在字处理网络单元的输出和细胞状态。

[0037] 具体的,定义待识别中文文本的字符序列为S=$\{c_1,c_2,c_3.....,c_m\}$,序列S与单词查找树匹配得到的单词集合为$w_{b,e}^{d}$,表示从B到E结束的单词子序列;

[0038] 嵌入层对于字符和单词分别按照以下公式进行向量转换:

[0039] $$x_i^c = e^c\left(c_i\right), i \in \{1,2,3...m\}$$

[0040] $$x_{b,e}^{w} = e^{w}\left(w_{b,e}^{d}\right)$$

[0041] 其中,x_i^c为字符c_i经嵌入层转换后得到的字符向量,$x_{b,e}^{w}$为单词集合$w_{b,e}^{d}$的词向量,e^c、e^w代表嵌入层的权重;

[0042] 用i_i^c、o_i^c、f_i^c分别表示字符向量处理网络单元的输入门、输出门和遗忘门的控制,σ、tanh表示激活函数sigmoid和tanh,字符向量处理网络单元对于输入的字符向量x_i^c按照下式进行处理:

[0043] $$\begin{bmatrix} i_i^c \\ o_i^c \\ f_i^c \\ \tilde{c}_i^c \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ tanh \end{bmatrix} \left(W^{c^T} \begin{bmatrix} x_i^c \\ h_{i-1}^c \end{bmatrix} + b^c \right)$$

[0044] $$c_i^c = f_i^c \odot c_{i-1}^c + i_i^c \odot \tilde{c}_i^c$$

[0045] $$\vec{h}_i^c = \overleftarrow{h}_i^c = o_i^c \odot tanh\left(c_i^c\right)$$

[0046] $$h_i^c = \left[\vec{h}_i^c;\ \overleftarrow{h}_i^c\right]$$

[0047] $$\tilde{c}_i^c = tanh\left(W_C \cdot \left[h_{i-1}^c, x_i^c\right] + b_C\right)$$

[0048] 式中，$\tilde{c}_i^c$表示经tanh函数处理后的细胞状态，为中间信息状态；h_{i-1}^c表示前一个字符的字符向量处理网络单元输出；$\vec{h}_i^c$和$\overleftarrow{h}_i^c$分别表示前向和反向两个方向的输出，h_i^c为结合两个方向的最后的输出；c_{i-1}^c表示从前一个字符及其相关的词传过来的细胞状态；W_C表示字处理单网络元的权重矩阵，W^{c^T}表示W_C的转置矩阵；b^c表示字处理网络单元中的常数项；⊙表示矩阵点积；

[0049] 用$i_{b,e}^w$、$f_{b,w}^w$分别表示词向量处理网络单元中的输入门和遗忘门的控制，词向量处理网络单元对于输入的词向量$x_{b,e}^w$按照下式进行处理：

[0050] $$\begin{bmatrix} i_{b,e}^w \\ f_{b,w}^w \\ \tilde{c}_{b,e}^w \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ tanh \end{bmatrix} \left(W^{w^T} \begin{bmatrix} x_{b,e}^w \\ h_b^c \end{bmatrix} + b^w \right)$$

[0051] $$c_{b,e}^w = f_{b,e}^w \odot c_b^c + i_{b,e}^w \odot \tilde{c}_{b,e}^w$$

[0052] 式中，$c_{b,e}^w$表示从B开始到E结束的词的细胞状态，$\tilde{c}_{b,e}^w$表示经tanh函数处理后的细胞状态，为中间信息状态；h_b^c表示第B个字在字处理网络单元的输出；W^{w^T}表示词处理网络单元权重矩阵的转置；b^w表示词处理网络单元的常数项；

[0053] 字处理网络单元中，对应字符向量x_j^c的输出细胞状态c_j^c按照下式计算：

[0054] $$c_j^c = \sum_{b \in \left\{b' | w_{b'}^d, j \in D\right\}} \alpha_{b,j}^c \odot c_{b,j}^w + \alpha_j^c \odot \tilde{c}_j^c$$

[0055] 其中，$\mathbf{c_{b,j}^w}$为从b到j组成的单词的细胞状态，$\mathbf{\alpha_{b,j}^c}$为从b到j组成的单词的细胞状态的权重，$\mathbf{\alpha_j^c}$为第j个字的细胞状态的权重，$\tilde{c}_j^c$为对应x_i^c在字处理网络单元中经tanh函数处理后的细胞状态，$b \in \left\{b' | w_{b'}^d, j \in D\right\}$中，b'代表所有可能的b集合，$w_{b'}^d$表示从b'到d组成的词，D表示所规定函数的定义域；

[0056] 并有：

[0057] $$\mathbf{\alpha_{b,j}^c = \frac{\exp\left(i_{b,j}^c\right)}{\exp\left(i_j^c\right) + \sum_{b'' \in \left\{b'' | w_{b''}^d, j \in D\right\}} \exp\left(i_{b'',j}^c\right)}}$$

[0058] $$\mathbf{\alpha_j^c = \frac{\exp\left(i_j^c\right)}{\exp\left(i_j^c\right) + \sum_{b'' \in \left\{b'' | w_{b''}^d, j \in D\right\}} \exp\left(i_{b'',j}^c\right)}}$$

[0059] 上式中，$\mathbf{i_{b,j}^c}$表示表示从b到j组成的词的输入门，$\mathbf{i_j^c}$表示第j个字的输入门，$\mathbf{w_{b''}^d}$表示从b″到d组成的词，b″表示所有可能的b'集合。

[0060] 可选的，双向长短时记忆网络的输出为待识别语句中各词中字符映射到标签的非归一化概率矩阵，CRF层网络采用维特比算法根据双向长短时记忆网络的输出，确定待识别

语句的标签序列。可确保更准确快速的得到最优结果。

[0061] 第二方面，本发明提供一种语义信息提取装置，包括：

[0062] 语料数据获取模块，被配置用于获取待识别的语料数据；

[0063] 预处理模块，被配置用于对获取到的语料数据进行预处理，预处理包括将语料数据转换为词向量和/或字向量；

[0064] 语义信息提取模块，用于将向量转换后的语料信息输入至预先训练的语义信息提取模型，得到命名实体识别结果；所述语义信息提取模型包括双向长短时记忆网络和CRF层网络，其训练样本为已标注字符标签和实体标签的语料数据的向量形式；双向长短时记忆网络的输出为待识别语句中各词中字符映射到标签的概率矩阵，CRF层网络根据双向长短时记忆网络的输出，确定待识别语句的标签序列并输出。

[0065] 有益效果

[0066] 与现有技术相比，本发明具有以下优点和进步：

[0067] 1:采用双向格子结构的长短时记忆网络Lattice LSTM进行实体标签预测，相比于传统基于字符嵌入的方法，格子结构的长短时记忆网络Lattice LSTM能够充分利用单词和词序信息，相比于传统基于词嵌入的方法，Lattice LSTM不会因为分词错误影响识别结果；

[0068] 2:将格子结构的长短时记忆网络由单向改为双向，能够更好的联系上下文的信息，得到词在文章中的具体意思；

[0069] 3:条件随机场CRF层使用维特比算法确定待识别语句的标签序列，能够更准确快速的得到最优结果。

附图说明

[0070] 图1所示为本发明的方法原理及流程示意图；

[0071] 图2所示为本发明语义信息提取模型构建过程示意图；

[0072] 图3所示为现有Lattice LSTM网络示意图；

[0073] 图4所示为本发明双向Lattice LSTM网络示意图；

[0074] 图5所示为本发明双向Lattice LSTM网络中字符向量处理原理示意图；

[0075] 图6所示为本发明双向Lattice LSTM网络中词向量处理原理示意图；

[0076] 图7所示为应用本发明方法的一种应用例模型的评价结果输出。

具体实施方式

[0077] 以下结合附图和具体实施例进一步描述。

[0078] 实施例1

[0079] 本实施例为一种语义信息提取方法，如图1所示，包括：

[0080] 获取待识别的语料数据；

[0081] 对获取到的语料数据进行预处理，预处理包括将语料数据转换为词向量和/或字向量；

[0082] 将向量转换后的语料信息输入至预先训练的语义信息提取模型，得到命名实体识别结果；所述语义信息提取模型包括双向长短时记忆网络和CRF层网络，其训练样本为已标注字符标签和实体标签的语料数据的向量形式；双向长短时记忆网络的输出为待识别语句

中各词中字符映射到标签的概率矩阵,CRF层网络根据双向长短时记忆网络的输出确定待识别语句的标签序列并输出。

[0083] 本发明旨在通过将传统格子结构的长短时记忆网络Lattice LSTM由单向改进为双向,在训练和识别时不仅能够充分利用单词和词序信息,不会因为分词错误影响识别结果,且能够更好的联系上下文的信息,使得机器如人工智能问答系统,能够更好的理解词在语句中的具体意思,进而针对识别出的实体回答相应问题。

[0084] 实施例1-1

[0085] 基于实施例1,本实施例具体介绍语义信息提取方法的实现,主要包括以下内容。

[0086] 一、待识别语料样本及其预处理

[0087] 本实施例用于识别的自然语言为中文语句文本,可为问答系统从用户处获取的询问语句或查询命令等。

[0088] 对获取到的语料信息进行预处理包括数据清洗。如过滤文本噪声数据,可采用现有技术。

[0089] 语料信息数据清洗后,将待识别语料与预设的单词查找树进行匹配,得到相应的单词集合,进而采用嵌入层Embedding将语料数据转换为词向量和字向量。Embedding层可采用现有的word2vec工具实现语料数据到向量的转换。单词查找树可根据汉语词典设置,用于待识别语句与汉语词典词库之间的匹配,已查找到待识别语句所包含的实体单词。

[0090] 然后即可采用嵌入层Embedding将语料数据转换为词向量和/或字向量。Embedding层可采用现有的word2vec工具实现语料数据到向量的转换。

[0091] 语义信息的提取是通过预先搭建好的语义信息提取模型,对输入的字向量、词向量进行理解,识别出文本中的命名实体,如用户在问答系统中给出问题:华为在北京有公司吗?经语义信息提取模型可识别问题语句中的字符标签:华B-ORG为E-ORG在O北B-LOC京E-LOC有O公O司O吗O,即识别出其中的命名实体北京和华为,之后即可根据识别出的实体进行答案检索,进而回答出问题。

[0092] 二、语义信息提取模型构建及其训练

[0093] 语义信息提取模型包括双向格子结构的长短时记忆网络和CRF层网络,其中双向格子结构的长短时记忆网络由现有的图3所示的单向Lattice LSTM改进得到,参考图4所示,双向格子结构的长短时记忆网络LSTM在单向格子结构的LSTM的基础上,添加了一层反方向的网络层layer,原layer为前向网络Forward Layer,新添加的为反向网络Backward Layer,这样可以使语料信息能够两个方向同时流动,能够更好的提取出句子中各字、词的信息,从而能够更好的识别出语料在整个句子中的语义。

[0094] 在改进为双向的基础上,双向格子结构的长短时记忆网络设置字处理网络单元和词处理网络单元,以能够同时分别处理字向量和词向量。

[0095] 参考图5、图6所示,字处理网络单元包括输入门、输出门和遗忘门,词处理网络单元包括输入门和遗忘门;

[0096] 字处理网络单元的输入量包括当前字符的字符向量、上一个字符的细胞状态,以及以当前字符为末位字符的单词从词处理网络单元输出的细胞状态;词处理网络单元的输入包括当前单词向量,以及当前单词的首位字符在字处理网络单元的输出和细胞状态。

[0097] 双向长短时记忆网络的输出为待识别语句中各词中字符映射到标签的非归一化

概率矩阵,CRF层网络采用维特比算法根据双向长短时记忆网络的输出,确定待识别语句的标签序列,可确保更准确快速的得到最优结果。

[0098] 语义信息提取模型搭建完成后,对于特定领域的自然语言识别可利用已有的相关领域语料数据进行模型训练,如医疗领域,可以通过爬虫三九健康网、寻医问药网等网站的语料数据。

[0099] 语义信息提取模型的训练包括以下内容:

[0100] (1)对文本资料进行标注处理,生成训练集、测试集和验证集;

[0101] (2)对已标注语料进行预处理;

[0102] (3)双向格子结构的长短时记忆网络Lattice LSTM训练;

[0103] (4)条件随机场CRF层训练;

[0104] (5)根据CRF层所得结果对模型预测结果进行评分。

[0105] 步骤(1)、语料标注处理

[0106] (1.1)对已有语料数据进行标注处理,具体方式为采用BMESO(begin,middle,end,single,other)的标记方式对训练语料数据进行标注,即位于该词语最开始的字符标记为B,位于该词语中间的字符标记为M,位于该词语末尾的词语标记为E,若该词语只有一个字符则标记为S,若该词语没有带标签或者不属于实体标签则标记为O。例如有语句为“小明今年在北京上学,明年准备去华为工作。”,则其标注结果为:小B-NAME、明E-NAME、今O、年O、在O、北B-LOC、京E-LOC、上O、学O、明O、年O、准O、备O、去O、华B-ORG、为E-ORG、工O、作O。

[0107] (1.2)然后可将数据集按照1:1:8的比例分为dev(验证集)、test(测试集)、train(训练集)三类数据集,以备后续对模型进行训练验证。

[0108] 步骤二、对标注语料进行预处理

[0109] (2.1)统计标注语料的字符,得到字符集合,然后将每个字符进行编号,得到字符集合相对应的字符编号集合;统计标注语料的标签,得到标签集合,然后将每个标签也进行编号,得到标签集合对应的标签编号集合;

[0110] (2.2)基于汉语词典建立一棵单词查找树,将标注语料中的每一条语句与单词查找树进行匹配,匹配成功的词保留,从而得到单词集合;

[0111] 比如一句话是“南京市长江小学”,匹配的过程如下所示:首先匹配“南”作为首字符的词,然后逐一查找单词树中是否有“南京市长江小学”、“南京市长江小”,“南京市长江”,“南京市长”,“南京市”,“南京”,最后可以得到以“南”字为首字符的词的一个列表[“南京市”,“南京”],然后再依次查找以‘京’,‘市’,‘长’,‘江’,‘小’,‘学’作为首字符的词,将匹配到的词保存到单词集合中;

[0112] (2.3)对单词集合中的词进行去重得到新的单词集合,并对新单词集合中的词进行编号,得到新对应的词编号集合;

[0113] (2.4)将标注语料中的字符和标签分别根据字符编号集合和标签编号集合转换为对应编号,并让标注语料中的每条语句与单词查找树进行匹配,保存每一句话匹配到的词,并将这些词根据词编号集合转换为对应的编号;

[0114] (2.5)将步骤2.4中转换成编号后的语料随机排列顺序,并采用随机无放回的方式从标注语料中抽取若干语句,以及其对应的标签和对应的单词集合。

[0115] 如对于语料“南京市长江小学”,则字符编号集合为:1南2京3市4长5江6小7学,标

签编号集合为:1.O 2.B-LOC 3.M-LOC 4.E-LOC 5.B-ORG 6.E-ORG,匹配后,获得的词编号集合为:1.南京2.南京市3.市长4.长江5.长江小学。

[0116] 然后将所有集合随机排列,训练模型时,抽取BatchSize句语句进行测试,对于语句“南京市长江小学”,即同时抽取其标签B-LOC、M-LOC、E-LOC、B-ORG、M-ORG、E-ORG和单词集合南京、南京市、市长、长江、长江小学。

[0117] 步骤三、双向格子结构的长短时记忆网络Lattice LSTM训练

[0118] 双向格子结构的长短时记忆网络Lattice LSTM模型是基于长短时记忆网络LSTM模型搭建的,并在单向长短时记忆网络结构LSTM的基础上改造为双向,从而更好的联系上下文信息。并通过设置字处理网络单元和词处理网络单元能够同时处理字符向量和词向量。

[0119] 通过模型训练对于双向Lattice LSTM模型中的相关参数进行调整,字处理网络单元和词处理网络单元分别处理字符向量和词向量时的内部结构图如图5和图6所示。

[0120] 假设需要处理一个字符序列S=c1,c2,c3.....,cm,首先需要利用嵌入层Embedding将语料数据转换为词向量和/或字向量,对于字符通过 $x_i^c = e^c(c_i), i \in \{1,2,3..m\}$ 可以得到每个字符的字符向量 x_i^c ,即字处理网络单元的输入向量。

[0121] 用 i_i^c 、 o_i^c 、 f_i^c 分别表示字符向量处理网络单元的输入门、输出门和遗忘门的控制,σ、tanh表示激活函数sigmoid和tanh,字符向量处理网络单元对于输入的字符向量 x_i^c 按照下式进行处理:

[0122] $$\begin{bmatrix} i_i^c \\ o_i^c \\ f_i^c \\ \tilde{c}_i^c \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{bmatrix} \left(W^{c^T} \begin{bmatrix} x_i^c \\ h_{i-1}^c \end{bmatrix} + b^c \right)$$

[0123] $$c_i^c = f_i^c \odot c_{i-1}^c + i_i^c \odot \tilde{c}_i^c$$

[0124] $$\overrightarrow{h_i^c} = \overleftarrow{h_i^c} = o_i^c \odot \tanh(c_i^c)$$

[0125] $$h_i^c = [\overrightarrow{h_i^c};\ \overleftarrow{h_i^c}]$$

[0126] $$\tilde{c}_i^c = tanh\left(W_C \cdot \left[h_{i-1}{}^c, x_i^c\right] + b_C\right)$$

[0127] 式中, $\tilde{c}_i^c$ 表示经tanh函数处理后的细胞状态,为中间信息状态; h_{i-1}^c 表示前一个字符的字符向量处理网络单元输出; $\overrightarrow{h_i^c}$ 和 $\overleftarrow{h_i^c}$ 分别表示前向和反向两个方向的输出, h_i^c 为结合两个方向的最后的输出; c_{i-1}^c 表示从前一个字符及其相关的词传过来的细胞状态; W_C 表示字处理单网络元的权重矩阵, W^{c^T} 表示 W_C 的转置矩阵; b^c 表示字处理网络单元中的常数项; ⊙表示矩阵点积;

[0128] 假定序列S和单词查找树进行匹配,得到这个序列的词集合表示为 $w_{b,e}^d$,从b开始到e结束的词的子序列;其向量形式为:

[0129] $$x_{b,e}^w = e^w(w_{b,e}^d)$$

[0130] 用$i_{b,e}^{w}$、$f_{b,w}^{w}$分别表示词向量处理网络单元中的输入门和遗忘门的控制，词向量处理网络单元对于输入的词向量$x_{b,e}^{w}$按照下式进行处理：

[0131] $$\begin{bmatrix} i_{b,e}^{w} \\ f_{b,w}^{w} \\ \tilde{c}_{b,e}^{w} \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \tanh \end{bmatrix} \left(W^{w^{T}} \begin{bmatrix} x_{b,e}^{w} \\ h_{b}^{c} \end{bmatrix} + b^{w} \right)$$

[0132] $$c_{b,e}^{w} = f_{b,e}^{w} \odot c_{b}^{c} + i_{b,e}^{w} \odot \tilde{c}_{b,e}^{w}$$

[0133] 式中，$c_{b,e}^{w}$表示从b开始到e结束的词的细胞状态，$\tilde{c}_{b,e}^{w}$表示经tanh函数处理后的细胞状态，为中间信息状态；h_{b}^{c}表示第b个字在字处理网络单元的输出；$W^{w^{T}}$表示词处理网络单元权重矩阵的转置；b^{w}表示词处理网络单元的常数项；

[0134] 从图6中可以看到词LSTM单元中没有输出门，是因为词LSTM单元的细胞状态都传给这个词最后一个字的字LSTM单元。除此之外，字符LSTM单元的输入不仅来自上一个字符的状态和字符向量，并且还包括前面多个词的LSTM单元输出的细胞状态$c_{b,e}^{w}$。因此字处理网络单元中，当前字的细胞状态输出计算公式如下：

[0135] $$c_{j}^{c} = \sum_{b \in \left\{ b' \middle| w_{b'}^{d}, \in D \right\}} \alpha_{b,j}^{c} \odot c_{b,j}^{w} + \alpha_{j}^{c} \odot \tilde{c}_{j}^{c}$$

[0136] 其中，$c_{b,j}^{w}$为从b到j组成的单词的细胞状态，$\alpha_{b,j}^{c}$为从b到j组成的单词的细胞状态的权重，α_{j}^{c}为第j个字的细胞状态的权重，$\tilde{c}_{j}^{c}$为对应x_{i}^{c}在字处理网络单元中经tanh函数处理后的细胞状态，$b \in \left\{ b' \middle| w_{b'}^{d}, j \in D \right\}$中，b'代表所有可能的b集合，$w_{b}^{d}$表示从b'到d组成的词，D表示所规定函数的定义域；

[0137] 并有：

[0138] $$\alpha_{b,j}^{c} = \frac{\exp\left(i_{b,j}^{c}\right)}{\exp\left(i_{j}^{c}\right) + \sum_{b' \in \left\{ b'' \middle| w_{b'',j \in D}^{d} \right\}} \exp\left(i_{b',j}^{c}\right)}$$

[0139] $$\alpha_{j}^{c} = \frac{\exp\left(i_{j}^{c}\right)}{\exp\left(i_{j}^{c}\right) + \sum_{b' \in \left\{ b'' \middle| w_{b'',j \in D}^{d} \right\}} \exp\left(i_{b',j}^{c}\right)}$$

[0140] 上式中，$i_{b,j}^{c}$表示表示从b到j组成的词的输入门，i_{j}^{c}表示第j个字的输入门，$w_{b''}^{d}$表示从b″到d组成的词，b″表示所有可能的b'集合。

[0141] 如对于句子“南京市长江小学”中的c_{7}^{c}“学”的细胞状态，输入量包含x_{7}^{c}（学）、$c_{6,7}^{c}$（小学）、$c_{4,7}^{c}$（长江小学）的信息，所以有：

[0142] $$c_{j}^{c} = \alpha_{7}^{c} \odot \tilde{c}_{j}^{c} + \alpha_{6,7}^{c} \odot c_{6,7}^{c} + \alpha_{4,7}^{c} \odot c_{4,7}^{c}$$

[0143] $$\alpha_{7}^{c} = \frac{\exp\left(i_{7}^{c}\right)}{\exp\left(i_{7}^{c}\right) + \exp\left(i_{6,7}^{c}\right) + \exp\left(i_{4,7}^{c}\right)}$$

[0144] $$\alpha_{4,7}^{c} = \frac{\exp\left(i_{4,7}^{c}\right)}{\exp\left(i_{7}^{c}\right) + \exp\left(i_{6,7}^{c}\right) + \exp\left(i_{4,7}^{c}\right)}$$

[0145] $$\alpha_{6,7}^{c}=\frac{\exp\left(i_{6,7}^{c}\right)}{\exp\left(i_{7}^{c}\right)+\exp\left(i_{6,7}^{c}\right)+\exp\left(i_{4,7}^{c}\right)}$$

[0146] 通过模型训练可不断调整网络中各层次之间的权重参数等，使得模型能够对待识别语句输出更加可靠的字符映射到标签的非归一化概率矩阵。

[0147] 步骤四、条件随机场CRF层训练，使得CRF层找出句子级别的标签信息。

[0148] 双向长短时记忆网络的输出为待识别语句中各词中字符映射到标签的非归一化概率矩阵，CRF层网络采用维特比算法根据双向长短时记忆网络的输出，确定待识别语句的标签序列。具体算法如下。

[0149] (4.1)设双向长短时记忆网络Bi-LSTM的输出矩阵为P，其中$P_{m,u}$为词w_m映射到标签的非归一化概率；我们假定存在一个转移矩阵A，则$A_{m,u}$表示标签m转移到标签u的转移概率。

[0150] 对于输入序列x对应的输出标签序列y，定义分数为：

[0151] $$\mathrm{score}(x,y)=\sum_{m=0}A_{y_m,y_{m+1}}+\sum_{m=1}^{n}P_{m,y_m}$$

[0152] (4.2)利用逻辑回归模型softmax函数，为每一个正确的标签序列y定义一个概率值：

[0153] $$p(y|x)=\frac{\exp\left(\mathrm{score}(x,y)\right)}{\sum_{y'}\exp\left(\mathrm{score}(x,y')\right)}$$

[0154] 利用对数似然，可以得到：

[0155] $\log(p(y|x))=\mathrm{score}(x,y)-\log(\sum_{y'}\mathrm{expscore}((x,y')))$

[0156] (4.3)如果存在N个样本，$\{(y^m,x^m)\}|_{m=1}^{N}$，则可以得到其损失函数为：

[0157] $$L=\sum_{M=1}^{N}\log\left(p(y^m|x^m)+\frac{\lambda}{2}\|\theta\|^2\right)$$

[0158] 其中$||\theta||^2$为L2正则项，λ为正则化参数。

[0159] 步骤五、根据CRF层所得结果对模型预测结果进行评分

[0160] 根据训练样本及训练过程中的CRF层识别结果计算准确率P和召回率R，利用以下公式计算评价分数F1：

[0161] $$F1=\frac{2*P*R}{P+R}$$

[0162] 响应于评价分数值大于预设值，则停止模型训练。

[0163] 步骤六、模型评分

[0164] (6.1)根据模型运行结果得出相应的准确率、召回率。

[0165] (6.2)根据F1分数计算标准进行计算，得出F1的总得分。

[0166] 本发明在Lattice LSTM结构的基础上，将单向LSTM变成双向LSTM改进，解决了单向结构只能影响当前位置后面序列的问题，从而使得原结构在能够影响当前位置后面序列的基础上，也能够对当前位置以及该位置前面的序列进行影响，能够更好的获知句子前后文的语义信息。参考图7所示，利用本发明的方法，在resumeNER数据集上，经过计算得分，精确率可达到0.926813，回归率可达到0.930528，f1达到0.928667，其中准确率最高，达到0.962347，有效提高了命名实体识别的准确率，可应用于其他领域进行实体识别。

[0167] 实施例2

[0168] 本实施例为一种语义信息提取装置，包括：

[0169] 语料数据获取模块，被配置用于获取待识别的语料数据；

[0170] 预处理模块，被配置用于对获取到的语料数据进行预处理，预处理包括将语料数据转换为词向量和/或字向量；

[0171] 语义信息提取模块，用于将向量转换后的语料信息输入至预先训练的语义信息提取模型，得到命名实体识别结果；所述语义信息提取模型包括双向长短时记忆网络和CRF层网络，其训练样本为已标注字符标签和实体标签的语料数据的向量形式；双向长短时记忆网络的输出为待识别语句中各词中字符映射到标签的概率矩阵，CRF层网络根据双向长短时记忆网络的输出，确定待识别语句的标签序列并输出。

[0172] 本实施例装置中各模块的具体实现，以及语义信息提取模型的构建、训练等内容，采取实施例1和实施例1-1的实施方式。

[0173] 本实施例的语义信息提取装置可实现于人工智能领域中的问答系统中，实现对用户语句的语义识别，以更加准确的执行用户指令或返回用户所需信息。

[0174] 本领域内的技术人员应明白，本申请的实施例可提供为方法、系统、或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0175] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的系统。

[0176] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令系统的制造品，该指令系统实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0177] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0178] 以上结合附图对本发明的实施例进行了描述，但是本发明并不局限于上述的具体实施方式，上述的具体实施方式仅仅是示意性的，而不是限制性的，本领域的普通技术人员在本发明的启示下，在不脱离本发明宗旨和权利要求所保护的范围情况下，还可做出很多形式，这些均属于本发明的保护之内。

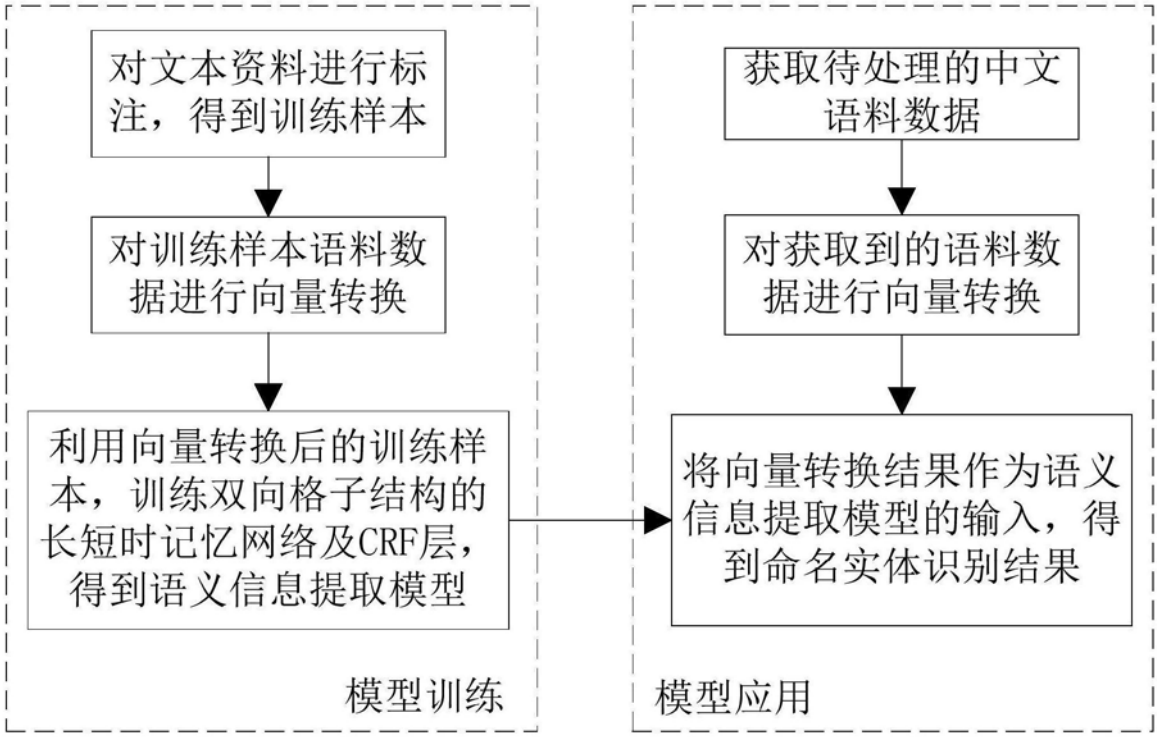

对文本资料进行标注，得到训练样本
对训练样本语料数据进行向量转换
利用向量转换后的训练样本，训练双向格子结构的长短时记忆网络及CRF层，得到语义信息提取模型
模型训练
获取待处理的中文语料数据
对获取到的语料数据进行向量转换
将向量转换结果作为语义信息提取模型的输入，得到命名实体识别结果
模型应用

图1

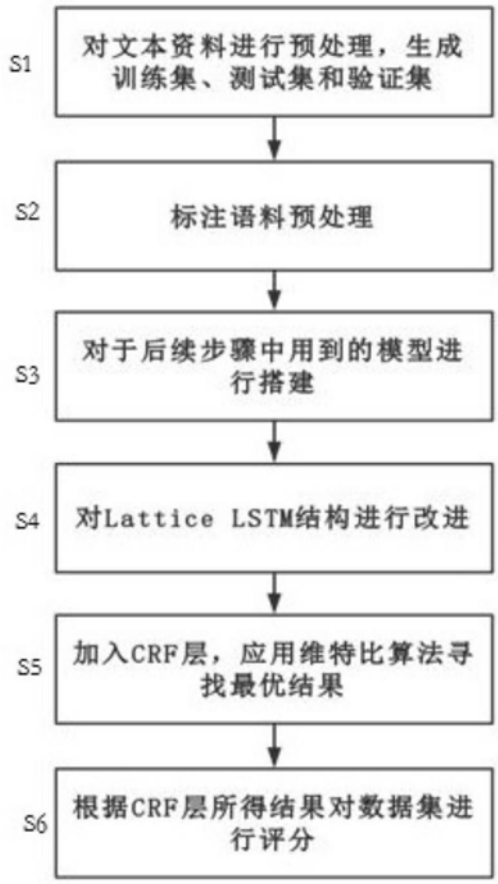

S1
对文本资料进行预处理，生成训练集、测试集和验证集
S2
标注语料预处理
S3
对于后续步骤中用到的模型进行搭建
S4
对Lattice LSTM结构进行改进
S5
加入CRF层，应用维特比算法寻找最优结果
S6
根据CRF层所得结果对数据集进行评分

图2

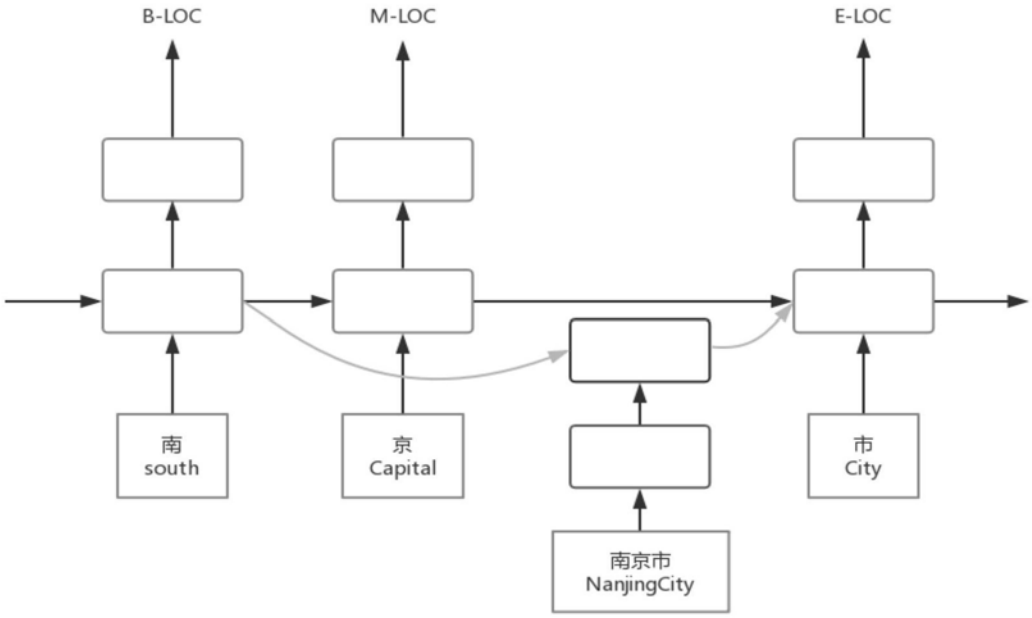

B-LOC
M-LOC
E-LOC
南
south
京
Capital
市
City
南京市
NanjingCity

图3

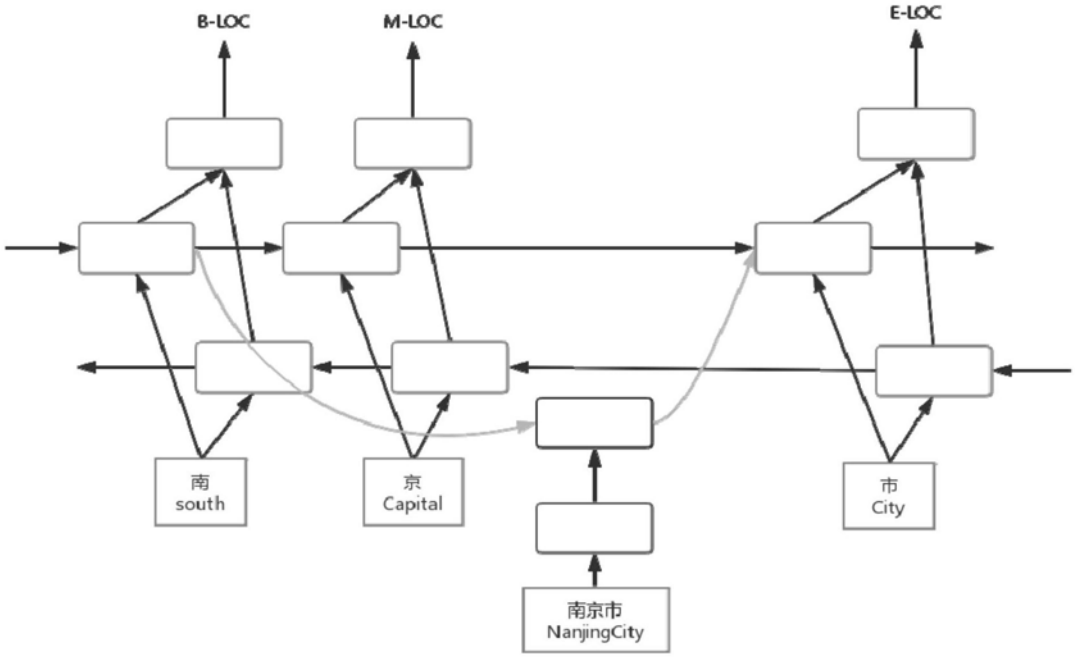

B-LOC
M-LOC
E-LOC
南
south
京
Capital
市
City
南京市
NanjingCity

图4

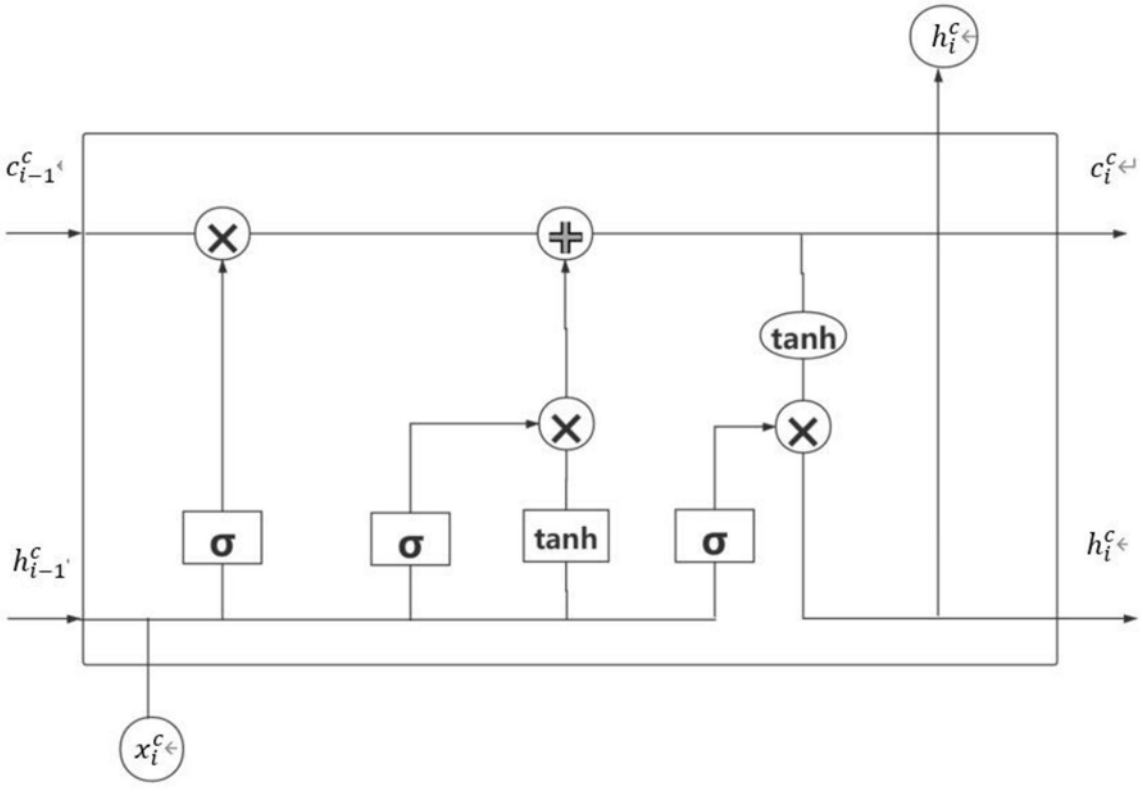

h_i^c
c_{i-1}^c
c_i^c
tanh
σ
σ
tanh
σ
h_{i-1}^c
h_i^c
x_i^c

图5

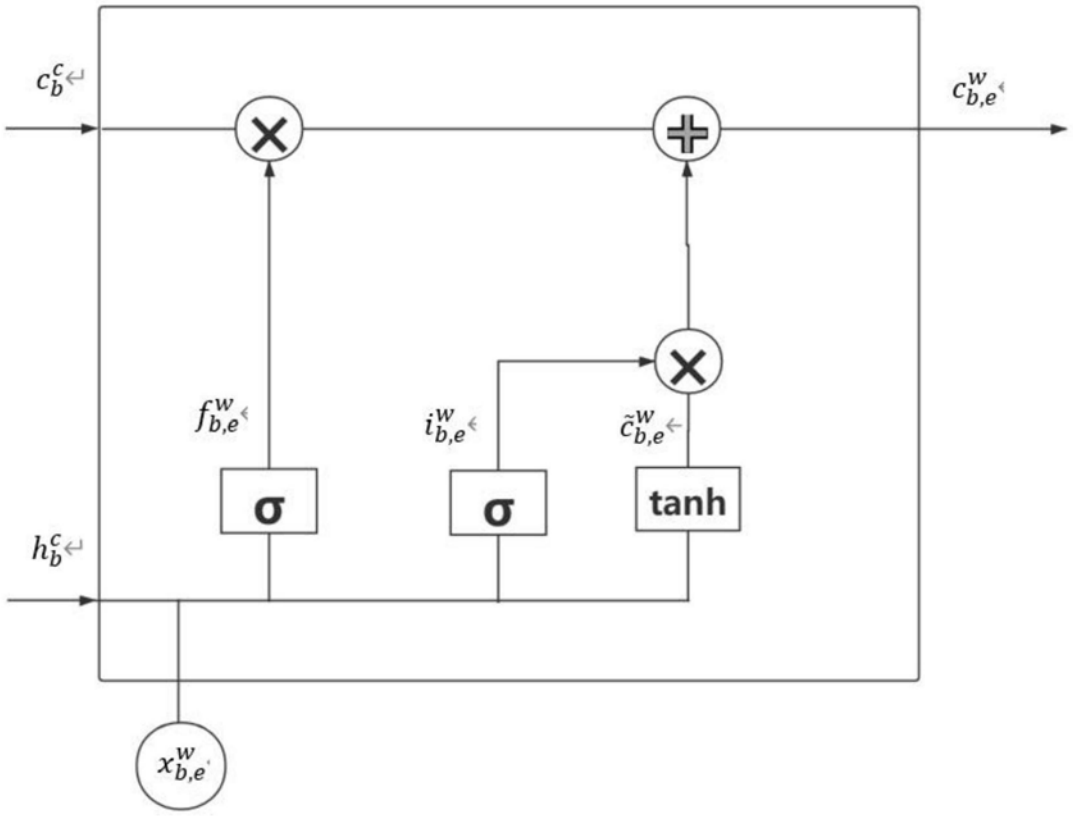


图6

```
Epoch 10/10: 100%|██████████| 38210/38210 [10:47:37<00:00,  1.02s/it, loss:0.01657]
  0%|          | 0/1 [00:00<?, ?it/s]
Test:   0%|          | 0/1 [00:00<?, ?it/s]
Evaluate data in 2.77 seconds!
Test: 100%|██████████| 1/1 [00:02<00:00,  2.77s/it]
Epoch 10/10: 100%|██████████| 38210/38210 [10:47:40<00:00,  1.02s/it, loss:0.01657]Evaluation on dev at Epoch 10/10. Step:38210/38210:
                                                                    SpanFPreRecMetric: f=0.928667, pre=0.926813, rec=0.930528

AccuracyMetric: acc=0.962347


In Epoch:10/Step:38210, got best dev performance:
SpanFPreRecMetric: f=0.928667, pre=0.926813, rec=0.930528
AccuracyMetric: acc=0.962347
Reloaded the best model.

Process finished with exit code 0
```

图7