(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 115577112 B
(45)授权公告日 2023.04.18

(21)申请号 202211576463.3
(22)申请日 2022.12.09
(65)同一申请的已公布的文献号
申请公布号 CN 115577112 A
(43)申请公布日 2023.01.06
(73)专利权人 成都索贝数码科技股份有限公司
地址 610041 四川省成都市高新区新园南二路2号
(72)发明人 朱婷婷 杨瀚
(74)专利代理机构 成都九鼎天元知识产权代理有限公司 51214
专利代理师 黎飞
(51)Int.Cl.
G06F 16/35(2019.01)
G06F 16/33(2019.01)
G06F 18/2415(2023.01)
G06N 3/047(2023.01)
G06N 3/048(2023.01)
G06N 3/08(2023.01)
(56)对比文件
CN 115238685 A,2022.10.25
CN 113705218 A,2021.11.26
CN 112307761 A,2021.02.02
审查员 何洋

权利要求书7页 说明书10页 附图1页

(54)发明名称

一种基于类型感知门控注意力机制的事件抽取方法及系统

(57)摘要

本发明涉及信息抽取技术领域，公开了一种基于类型感知门控注意力机制的事件抽取方法及系统，该事件抽取方法，利用门控信息指导，使不同事件类别下有不同信息流向触发词，过滤与触发词无关的噪声。本发明解决了现有技术存在的事件论元抽取准确度低、角色分类效果差等问题。



CN 115577112 B

1.一种基于类型感知门控注意力机制的事件抽取方法,其特征在于,利用门控信息指导,使不同事件类别下有不同信息流向触发词,过滤与触发词无关的噪声;

包括以下步骤:

S1,文本向量化:将样本$X=\{x_1,x_2,x_3,\ldots,x_i\ \ldots,x_n\}$输入到基于语言模型的文本向量化层当中,获得文本向量化结果$E=\{e_1,e_2,e_3,\ldots e_i,\ldots,e_n\}$;其中,$x_i$表示文本中的第$i$个字,$e_i\in R^d$表示$x_i$对应的向量化结果,$E$表示文本X的向量化结果,R表示实数,d表示向量的维度,R^d表示d维实数向量;

S2,事件检测:将文本向量化结果E输入融合类型感知的门控注意力机制的事件检测模块,以完成触发词检测和事件分类两个子任务;

S3,论元抽取:对步骤S2完成触发词检测和事件分类后的结果中每种事件类型下的每一个触发词,利用融合了可学习的角色交互参数的论元抽取模块完成论元抽取和论元角色分类两个子任务;

步骤S2中融合类型感知门控注意力机制的事件检测模块包括串联的如下子模块:触发词提取层、门控注意力事件分类层;

触发词提取层的构建过程包括如下步骤:

S21,首先按照如下公式计算获得输入文本中每个字为触发词开始/结束字符的概率:

$$t_i^{ts}=sigmoid(\boldsymbol{w_{ts}}\boldsymbol{e_i}+b_{ts});$$

$$t_i^{te}=sigmoid(\boldsymbol{w_{te}}\boldsymbol{e_i}+b_{te});$$

其中,$\boldsymbol{w_{ts}}\in R^d$、$\boldsymbol{w_{te}}\in R^d$、$b_{ts}\in R$、$b_{te}\in R$为可学习的网络参数,sigmoid为激活函数,$t_i^{ts}$为输入文本中第$i$个字是触发词的开始字符的概率,$t_i^{te}$为文本中第$i$个字是触发词的结束字符的概率;

S22,根据预先设定的阈值δ_1、δ_2对S21中的结果进行过滤,从而获得位置集合Ts、Te:

$$Ts=\{i|t_i^{ts}>\delta_1\};$$

$$Te=\{i|t_i^{te}>\delta_2\};$$

其中,Ts表示触发词的开始字符位置集合,Te表示触发词的结束字符位置集合;

S23,结合步骤S22的结果,利用最近匹配原则获得触发词集合$S_{Trigger}=\{t=(x_s,x_{s+1},\ldots,x_e)|s\le e;\ s\in Ts,e\in Te\}$;

其中,t为候选触发词,s为候选触发词t的开始字符在文本X中的位置,e为集合Te中最靠近s的元素;

门控注意力事件分类层的构建过程包括如下步骤：

S24，在门控信息过滤层中，对每个事件类别c，定义事件类别语义向量$\boldsymbol{c} \in R^d$，按如下公式计算相应门控向量：

$$\boldsymbol{g}^c = sigmoid\left(\boldsymbol{W}_g \boldsymbol{c} + b_g\right);$$

其中，$\boldsymbol{g}^c \in \boldsymbol{R}^d$为事件类别$c$下的门控向量，$\boldsymbol{W}_g \in R^{d*d}$为门控单元的可学习权重参数，$b_g \in R^d$为门控单元的可学习偏置参数；

S25，结合S24中的结果，在每个事件类别下，利用元素积函数过滤上下文信息：

$$\boldsymbol{f}_i^c = \boldsymbol{e}_i \odot \boldsymbol{g}^c;$$

其中，$\boldsymbol{e}_i \in R^d$为输入文本中第i个字对应的向量，$\boldsymbol{f}_i^c \in R^d$为经过门控信息过滤层后输入文本中第$i$个字在事件类别$c$下经过信息过滤后的对应的向量；

S26，在注意力信息融合层中，利用注意力计算函数获得在事件类别c下输入文本中第i个字对于触发词$t \in S_{Trigger}$的重要性分数α_i^c；

S27，结合S26中计算所获重要性分数，利用如下公式在每个事件类别下获得与每个触发词相关的最终信息聚合结果：

$$\boldsymbol{s}_t^c = \sum_{i=1}^{n} \alpha_i^c \boldsymbol{f}_i^c;$$

其中，$\boldsymbol{s}_t^c \in R^d$为经过注意力信息融合层后事件类别$c$下与触发词$t$相关的信息聚合向量；

S28，在事件分类层中，结合步骤S27中所得的与触发词t相关的信息聚合向量，利用如下公式判定触发词t所属于的事件类型：

$$\hat{c} = sigmoid(\boldsymbol{w}^T \boldsymbol{s}_t^c + b);$$

其中，$\boldsymbol{w} \in R^d$为事件类别判定单元的可学习权重参数，$\boldsymbol{b} \in R$为事件类别判定单元的偏置参数，w^T表示w的转置；

根据预先设定的阈值δ_3，所有满足如下条件的事件类别c均会被判定为触发词t所属事件类别：

$$\hat{c} > \delta_3;$$

最终每个触发词t的事件类别集合为$\dot{C} = \{c | \hat{c} > \delta_3\}$。

2.根据权利要求1所述的一种基于类型感知门控注意力机制的事件抽取方法，其特征在于，步骤S26中，注意力计算函数公式如下：

$$\alpha_i^c = exp\left(\theta(t, f_i^c)\right) / \sum_{k=1}^{n} exp\left(\theta(t, f_k^c)\right);$$

其中，t为触发词$t = (x_s, x_{s+1}, \ldots, x_e)$的表征向量，通过如下公式计算获得：

$$t = \frac{e_s + e_e}{2};$$

其中，e_s表示触发词t的开始字符的表征向量，e_e表示触发词t的结束字符的表征向量；

$\theta(t, f_i^c)$的定义如下：

$$\theta(t, f_i^c) = V^T tanh(W[t; f_i^c; |t - f_i^c|; t \odot f_i^c]);$$

其中，$tanh$表示tanh激活函数，V^T表示V的转置，W表示权重，[;;]表示向量的拼接。

3.根据权利要求2所述的一种基于类型感知门控注意力机制的事件抽取方法，其特征在于，步骤S3所述融合了可学习的角色交互参数的论元抽取模块的构建过程包括如下步骤：

S31，利用如下公式计算上下文中融入触发词表征向量t：

$$z_i^c = \gamma \odot \left(\frac{f_i^c - \mu}{\sigma} + \beta \right);$$

$$\gamma = w_\gamma t + b_\gamma;$$

$$\beta = w_\beta t + b_\beta;$$

其中，μ是基于输入文本在事件类别c下经过信息过滤后的对应向量f_i^c计算所得的均值，σ是基于输入文本在事件类别c下经过信息过滤后的对应向量f_i^c计算所得的标准方差，z_i^c表示输入文本中第i个字在事件类别c下经过信息过滤并融合触发词t信息后的向量，γ、β分别表示扩展参数和平移参数，w_γ、b_γ分别表示用于计算γ的线性层的权重参数、偏置参数，w_β、b_β则分别表示用于计算β的线性层的权重参数、偏置参数；

S32，利用如下公式分别计算输入文本中每个字作为每种角色事件类别r下论元的开始字符/结束字符的概率大小：

$$t_i^{tcrs} = sigmoid(\boldsymbol{W_{rs}} z_i^c + b_{rs});$$

$$t_i^{tcre} = sigmoid(\boldsymbol{W_{re}} z_i^c + b_{re});$$

其中，t_i^{tcrs}表示输入文本第i个字是事件类别为c的触发词t的角色r下的论元的开始字符的概率值，t_i^{tcre}表示输入文本第i个字是事件类别为c的触发词t的角色r下的论元的结束字符的概率值，$\boldsymbol{W_{rs}}$、$\boldsymbol{W_{re}}$表示权重参数，b_{rs}、b_{re}表示偏置参数；

S33，定义可学习的角色交互矩阵W_{RR}，并设计如下判定函数：

$$\boldsymbol{I(c)} = sigmoid(\boldsymbol{W_{RR}}(\boldsymbol{W_R c} + \boldsymbol{b_R}) + \boldsymbol{b_{RR}});$$

其中，$\boldsymbol{I(c)}$表示事件类别c下的指示函数，$\boldsymbol{W_R}$、$\boldsymbol{b_R}$表示第一层线性层的权重参数与偏置参数，$\boldsymbol{b_{RR}}$表示第二层线性层的偏置参数；

将$\boldsymbol{I(c)}$作为权重，结合该权重修正步骤S32中的计算结果：

$$\hat{t}_i^{tcrs} = I(c)_r t_i^{tcrs};$$

$$\hat{t}_i^{tcre} = I(c)_r t_i^{tcre};$$

经过训练，判定函数不仅可以学到事件角色之间的相互关系，同时也学到了角色之间的相互关系；

其中，$\hat{t}_i^{tcrs}$为输入文本第i个字是事件类别为c的触发词t的角色r下的论元的开始字符的最终概率值，$\hat{t}_i^{tcre}$为输入文本第i个字是事件类别为c的触发词t的角色r下的论元的结束字符的最终概率值，$I(c)_r$为事件类别c下角色r的权重；

S34，根据预先设定的阈值δ_4、δ_5对S33中的结果进行过滤，从而获得位置集合As_r、Ae_r：

$$As_r = \{i | \hat{t}_i^{tcrs} > \delta_4\};$$

$$Ae_r = \{i | \hat{t}_i^{tcre} > \delta_5\};$$

其中，As_r表示角色r下论元的开始字符位置集合，Ae_r表示角色r下论元的结束字符位置集合；

S35，结合步骤S34的结果，利用最近匹配原则获得角色r下的论元集合$S_{Ar} = \{a = (x_s, x_{s+1}, \dots, x_e) | s \le e;\ s \in As_r, e \in Ae_r\}$；其中，$s$为论元$a$的开始字符在文本X中的位置，$e$为集合$Ae_r$中最靠近$s$的元素。

4.根据权利要求3所述的一种基于类型感知门控注意力机制的事件抽取方法，其特征在于，该事件抽取方法的损失函数如下：

$$-\log\left(\prod_{x\in D}\left[\prod_{(t,c,a_r)\in E_x} p\left((t,c,a_r)|x\right)\right]\right)$$

$$=-\log\left(\prod_{x\in D}\left[\prod_{t\in E_x} p(t|x)\prod_{c\in E_x|t} p(c|x,t)\prod_{a_r\in E_x|t,c} p(a_r|x,t,c)\right]\right);$$

其中，$c\in E_x|t$表示输入文本x中触发词t的事件类型，$a_r\in E_x|t,c$表示输入文本x中事件类型为c的触发词t的角色为r的论元，D表示所有输入样本，$p\left((t,c,a_r)|x\right)$表示样本x中触发词为t且事件类型为c且角色r的论元为a的概率，$p(t|x)$表示样本x中触发词为t的概率，$p(c|x,t)$表示样本x中触发词t的事件类型为c的概率，$p\left((t,c,a_r)|x\right)$表示在样本x中事件类别为c且触发词为$t$时角色r的论元为$a$的概率，$E_x$表示样本x中的所有事件。

5.根据权利要求1至4任一项所述的一种基于类型感知门控注意力机制的事件抽取方法，其特征在于，步骤S1中，语言模型为BERT模型。

6.一种基于类型感知门控注意力机制的事件抽取系统，其特征在于，用于实现权利要求1至5任一项所述的一种基于类型感知门控注意力机制的事件抽取方法，包括依次相连的以下模块：

文本向量化模块：用以，将样本$\mathrm{X}=\{x_1,x_2,x_3,\ldots,x_i\ldots,x_n\}$输入到基于语言模型的文本向量化层当中，获得文本向量化结果$\mathrm{E}=\{\boldsymbol{e_1},\boldsymbol{e_2},\boldsymbol{e_3},\ldots\boldsymbol{e_i},\ldots,\boldsymbol{e_n}\}$；其中，$x_i$表示文本中的第$i$个字，$\boldsymbol{e_i}\in R^d$表示$x_i$对应的向量化结果，$E$表示文本X的向量化结果，R表示实数，d表示向量的维度，R^d表示d维实数向量；

事件检测模块：用以，将文本向量化结果E输入融合类型感知的门控注意力机制的事件检测模块，以完成触发词检测和事件分类两个子任务；

论元抽取模块：用以，对事件检测模块完成触发词检测和事件分类后的结果中每种事件类型下的每一个触发词，利用融合了可学习的角色交互参数的论元抽取模块完成论元抽取和论元角色分类两个子任务；

融合类型感知门控注意力机制的事件检测模块包括串联的如下子模块：触发词提取层、门控注意力事件分类层；

触发词提取层的构建过程包括如下步骤：

S21，首先按照如下公式计算获得输入文本中每个字为触发词开始/结束字符的概率：

$$t_i^{ts}=sigmoid(\boldsymbol{w_{ts}}\boldsymbol{e_i}+b_{ts});$$

$$t_i^{te}=sigmoid(\boldsymbol{w_{te}}\boldsymbol{e_i}+b_{te});$$

其中，$\boldsymbol{w_{ts}}\in R^d$、$\boldsymbol{w_{te}}\in R^d$、$b_{ts}\in R$、$b_{te}\in R$为可学习的网络参数，sigmoid为

激活函数，t_i^{ts}为输入文本中第i个字是触发词的开始字符的概率，t_i^{te}为文本中第i个字是触发词的结束字符的概率；

S22，根据预先设定的阈值δ_1、δ_2对S21中的结果进行过滤，从而获得位置集合Ts、Te：

$$Ts = \{i | t_i^{ts} > \delta_1\};$$

$$Te = \{i | t_i^{te} > \delta_2\};$$

其中，Ts表示触发词的开始字符位置集合，Te表示触发词的结束字符位置集合；

S23，结合步骤S22的结果，利用最近匹配原则获得触发词集合$S_{Trigger} = \{t = (x_s, x_{s+1}, \dots, x_e) | s \le e;\ s \in Ts, e \in Te\}$；

其中，t为候选触发词，s为候选触发词t的开始字符在文本X中的位置，e为集合Te中最靠近s的元素；

门控注意力事件分类层的构建过程包括如下步骤：

S24，在门控信息过滤层中，对每个事件类别c，定义事件类别语义向量$\boldsymbol{c} \in R^d$，按如下公式计算相应门控向量：

$$\boldsymbol{g}^{\boldsymbol{c}} = sigmoid\left(\boldsymbol{W}_{\boldsymbol{g}}\boldsymbol{c} + b_g\right);$$

其中，$\boldsymbol{g}^{\boldsymbol{c}} \in \boldsymbol{R}^{\boldsymbol{d}}$为事件类别$c$下的门控向量，$\boldsymbol{W}_{\boldsymbol{g}} \in R^{d*d}$为门控单元的可学习权重参数，$b_g \in R^d$为门控单元的可学习偏置参数；

S25，结合S24中的结果，在每个事件类别下，利用元素积函数过滤上下文信息：

$$\boldsymbol{f}_{\boldsymbol{i}}^{\boldsymbol{c}} = \boldsymbol{e}_{\boldsymbol{i}} \odot \boldsymbol{g}^{\boldsymbol{c}};$$

其中，$\boldsymbol{e}_{\boldsymbol{i}} \in R^d$为输入文本中第$i$个字对应的向量，$\boldsymbol{f}_{\boldsymbol{i}}^{\boldsymbol{c}} \in R^d$为经过门控信息过滤层后输入文本中第$i$个字在事件类别$c$下经过信息过滤后的对应的向量；

S26，在注意力信息融合层中，利用注意力计算函数获得在事件类别c下输入文本中第i个字对于触发词$t \in S_{Trigger}$的重要性分数α_i^c；

S27，结合S26中计算所获重要性分数，利用如下公式在每个事件类别下获得与每个触发词相关的最终信息聚合结果：

$$\boldsymbol{s}_{\boldsymbol{t}}^{\boldsymbol{c}} = \sum_{i=1}^{n} \alpha_i^c \boldsymbol{f}_{\boldsymbol{i}}^{\boldsymbol{c}};$$

其中，$\boldsymbol{s}_{\boldsymbol{t}}^{\boldsymbol{c}} \in R^d$为经过注意力信息融合层后事件类别$c$下与触发词t相关的信息聚合向量；

S28，在事件分类层中，结合步骤S27中所得的与触发词t相关的信息聚合向量，利用如

下公式判定触发词t所属于的事件类型：

$$\hat{c} = sigmoid(\boldsymbol{w}^T \boldsymbol{s}_t^c + b)$$；

其中，$\boldsymbol{w} \in R^d$为事件类别判定单元的可学习权重参数，$\boldsymbol{b} \in R$为事件类别判定单元的偏置参数，w^T表示w的转置；

根据预先设定的阈值δ_3，所有满足如下条件的事件类别c均会被判定为触发词t所属事件类别：

$$\hat{c} > \delta_3$$；

最终每个触发词t的事件类别集合为$\dot{C} = \{c|\hat{c} > \delta_3\}$。

一种基于类型感知门控注意力机制的事件抽取方法及系统

技术领域

[0001] 本发明涉及信息抽取技术领域，具体是一种基于类型感知门控注意力机制的事件抽取方法及系统。

背景技术

[0002] 事件抽取是信息抽取领域中既基础又极具挑战的任务。事件抽取通常包括两个任务，即事件检测与事件论元抽取。更具体地，事件检测任务又包括触发词检测和事件分类两个子任务，事件论元抽取又包括论元检测和角色分类两个子任务。近年来，随着深度学习的不断发展，基于深度学习的事件抽取方法取得了一定程度上的提升，但是事件抽取的难点依然还未被完全解决。

[0003] 现阶段，大多数事件抽取方法都集中在解决重叠论元场景，却忽视了重叠触发词场景和触发词歧义问题。换句话说，不止论元可能在不同/同一事件中扮演不同的角色，触发词也可能有多种事件类型。

[0004] 此外，相比于事件检测，事件论元抽取更加困难。许多方法尝试利用角色信息提升事件论元抽取效果，比如角色出现频率(重要性)、角色相关性如层次概念关系、角色语法关系等。角色出现频率忽略了角色之间的相互关系，其他角色相关性则需要基于人的经验进行总结和归纳，并且在某些数据上并不适用，所以对事件论元抽取效果的提升不大。

发明内容

[0005] 为克服现有技术的不足，本发明提供了一种基于类型感知门控注意力机制的事件抽取方法及系统，解决现有技术存在的事件论元抽取准确度低、角色分类效果差等问题。

[0006] 本发明解决上述问题所采用的技术方案是：

[0007] 一种基于类型感知门控注意力机制的事件抽取方法，利用门控信息指导，使不同事件类别下有不同信息流向触发词，过滤与触发词无关的噪声。

[0008] 作为一种优选的技术方案，包括以下步骤：

[0009] S1，文本向量化：将样本$\mathrm{X}=\{x_1,x_2,x_3,\ldots,x_i\ \ldots,x_n\}$输入到基于语言模型的文本向量化层当中，获得文本向量化结果$\mathrm{E}=\{\boldsymbol{e_1},\boldsymbol{e_2},\boldsymbol{e_3},\ldots\boldsymbol{e_i},\ldots,\boldsymbol{e_n}\}$；其中，$x_i$表示文本中的第$i$个字，$\boldsymbol{e_i}\in R^d$表示$x_i$对应的向量化结果，$E$表示文本X的向量化结果，R表示实数，d表示向量的维度，R^d表示d维实数向量；

[0010] S2，事件检测：将文本向量化结果E输入融合类型感知的门控注意力机制的事件检测模块，以完成触发词检测和事件分类两个子任务；

[0011] S3，论元抽取：对步骤S2完成触发词检测和事件分类后的结果中每种事件类型下的每一个触发词，利用融合了可学习的角色交互参数的论元抽取模块完成论元抽取和论元角色分类两个子任务。

[0012] 作为一种优选的技术方案，步骤S2中融合类型感知门控注意力机制的事件检测模块包括串联的如下子模块：触发词提取层、门控注意力事件分类层。

[0013] 作为一种优选的技术方案，触发词提取层的构建过程包括如下步骤：

[0014] S21，首先按照如下公式计算获得输入文本中每个字为触发词开始/结束字符的概率：

$$t_i^{ts} = sigmoid(\boldsymbol{w_{ts}} \boldsymbol{e_i} + b_{ts});$$

[0015]

$$t_i^{te} = sigmoid(\boldsymbol{w_{te}} \boldsymbol{e_i} + b_{te});$$

[0016] 其中，$\boldsymbol{w_{ts}} \in R^d$、$\boldsymbol{w_{te}} \in R^d$、$b_{ts} \in R$、$b_{te} \in R$为可学习的网络参数，sigmoid为激活函数，$t_i^{ts}$为输入文本中第$i$个字是触发词的开始字符的概率，$t_i^{te}$为文本中第$i$个字是触发词的结束字符的概率；

[0017] S22，根据预先设定的阈值δ_1、δ_2对S21中的结果进行过滤，从而获得位置集合Ts、Te：

[0018] $$Ts = \{i | t_i^{ts} > \delta_1\};$$

[0019] $$Te = \{i | t_i^{te} > \delta_2\};$$

[0020] 其中，Ts表示触发词的开始字符位置集合，Te表示触发词的结束字符位置集合；

[0021] S23，结合步骤S22的结果，利用最近匹配原则获得触发词集合$S_{Trigger} = \{t = (x_s, x_{s+1}, \ldots, x_e) | s \le e;\ s \in Ts, e \in Te\}$；

[0022] 其中，t为候选触发词，s为候选触发词t的开始字符在文本X中的位置，e为集合Te中最靠近s的元素；

[0023] 门控注意力事件分类层的构建过程包括如下步骤：

[0024] S24，在门控信息过滤层中，对每个事件类别c，定义事件类别语义向量$\boldsymbol{c} \in R^d$，按如下公式计算相应门控向量：

[0025] $$\boldsymbol{g^c} = sigmoid(\boldsymbol{W_g c} + b_g);$$

[0026] 其中，$\boldsymbol{g^c} \in \boldsymbol{R^d}$为事件类别$c$下的门控向量，$\boldsymbol{W_g} \in R^{d*d}$为门控单元的可学习权重参数，$b_g \in R^d$为门控单元的可学习偏置参数；

[0027] S25，结合S24中的结果，在每个事件类别下，利用元素积函数过滤上下文信息：

[0028] $$\boldsymbol{f_i^c} = \boldsymbol{e_i} \odot \boldsymbol{g^c};$$

[0029] 其中，$\boldsymbol{e_i} \in R^d$为输入文本中第i个字对应的向量，$\boldsymbol{f_i^c} \in R^d$为经过门控信息

过滤层后输入文本中第i个字在事件类别c下经过信息过滤后的对应的向量；

[0030] S26，在注意力信息融合层中，利用注意力计算函数获得在事件类别c下输入文本中第i个字对于触发词$t \in S_{Trigger}$的重要性分数α_i^c；

[0031] S27，结合S26中计算所获重要性分数，利用如下公式在每个事件类别下获得与每个触发词相关的最终信息聚合结果：

[0032] $$s_t^c = \sum_{i=1}^{n} \alpha_i^c f_i^c;$$

[0033] 其中，$s_t^c \in R^d$为经过注意力信息融合层后事件类别c下与触发词t相关的信息聚合向量；

[0034] S28，在事件分类层中，结合步骤S27中所得的与触发词t相关的信息聚合向量，利用如下公式判定触发词t所属于的事件类型：

[0035] $$\hat{c} = sigmoid(w^T s_t^c + b);$$

[0036] 其中，$w \in R^d$为事件类别判定单元的可学习权重参数，$b \in R$为事件类别判定单元的偏置参数，w^T表示w的转置；

[0037] 根据预先设定的阈值δ_3，所有满足如下条件的事件类别c均会被判定为触发词t所属事件类别：

[0038] $$\hat{c} > \delta_3;$$

[0039] 最终每个触发词t的事件类别集合为$\dot{C} = \{c|\hat{c} > \delta_3\}$。

[0040] 作为一种优选的技术方案，步骤S26中，注意力计算函数公式如下：

[0041] $$\alpha_i^c = exp\left(\theta(t, f_i^c)\right) / \sum_{k=1}^{n} exp\left(\theta(t, f_k^c)\right);$$

[0042] 其中，t为触发词$t = (x_s, x_{s+1}, \ldots, x_e)$的表征向量，通过如下公式计算获得：

[0043] $$t = \frac{e_s + e_e}{2};$$

[0044] 其中，e_s表示触发词t的开始字符的表征向量，e_e表示触发词t的结束字符的表征向量；

[0045] $\theta(t, f_i^c)$的定义如下：

[0046] $$\theta(t, f_i^c) = V^T tanh(W[t; f_i^c; |t - f_i^c|; t \odot f_i^c]);$$

[0047] 其中，$tanh$表示tanh激活函数，V^T表示V的转置，W表示权重，[;;]表示向量的拼接。

[0048] 作为一种优选的技术方案，步骤S3所述融合了可学习的角色交互参数的论元抽取模块的构建过程包括如下步骤：

[0049] S31，利用如下公式计算上下文中融入触发词表征向量t：

$$z_i^c = \gamma \odot \left(\frac{f_i^c - \mu}{\sigma} + \beta \right);$$

[0050]

$$\gamma = w_\gamma t + b_\gamma;$$

$$\beta = w_\beta t + b_\beta;$$

[0051] 其中，μ是基于输入文本在事件类别c下经过信息过滤后的对应向量f_i^c计算所得的均值，σ是基于输入文本在事件类别c下经过信息过滤后的对应向量f_i^c计算所得的标准方差，z_i^c表示输入文本中第i个字在事件类别c下经过信息过滤并融合触发词t信息后的向量，γ、β分别表示扩展参数和平移参数，w_γ、b_γ分别表示用于计算γ的线性层的权重参数、偏置参数，w_β、b_β则分别表示用于计算β的线性层的权重参数、偏置参数；

[0052] S32，利用如下公式分别计算输入文本中每个字作为每种角色事件类别r下论元的开始字符/结束字符的概率大小：

$$t_i^{tcrs} = sigmoid(W_{rs} z_i^c + b_{rs});$$

[0053]

$$t_i^{tcre} = sigmoid(W_{re} z_i^c + b_{re});$$

[0054] 其中，t_i^{tcrs}表示输入文本第i个字是事件类别为c的触发词t的角色r下的论元的开始字符的概率值，t_i^{tcre}表示输入文本第i个字是事件类别为c的触发词t的角色r下的论元的结束字符的概率值，W_{rs}、W_{re}表示权重参数，b_{rs}、b_{re}表示偏置参数；

[0055] S33，定义可学习的角色交互矩阵W_{RR}，并设计如下判定函数：

[0056] $I(c) = sigmoid(W_{RR}(W_R c + b_R) + b_{RR})$；

[0057] 其中，$I(c)$表示事件类别c下的指示函数，W_R、b_R表示第一层线性层的权重参数与偏置参数，b_{RR}表示第二层线性层的偏置参数；

[0058] 将$I(c)$作为权重，结合该权重修正步骤S32中的计算结果：

[0059] $\hat{t}_i^{tcrs} = I(c)_r t_i^{tcrs}$；

[0060] $\hat{t}_i^{tcre} = I(c)_r t_i^{tcre}$；

[0061] 经过训练，判定函数不仅可以学到事件角色之间的相互关系，同时也学到了角色之间的相互关系；

[0062] 其中，$\hat{t}_i^{tcrs}$为输入文本第i个字是事件类别为c的触发词t的角色r下的论元的开始字符的最终概率值，$\hat{t}_i^{tcre}$为输入文本第i个字是事件类别为c的触发词t的角色r下的论元的结束字符的最终概率值，$I(c)_r$为事件类别c下角色r的权重；

[0063] S34，根据预先设定的阈值δ_4、δ_5对S33中的结果进行过滤，从而获得位置集合As_r、Ae_r：

[0064] $As_r = \{i | \hat{t}_i^{tcrs} > \delta_4\}$；

[0065] $Ae_r = \{i | \hat{t}_i^{tcre} > \delta_5\}$；

[0066] 其中，As_r表示角色r下论元的开始字符位置集合，Ae_r表示角色r下论元的结束字符位置集合；

[0067] S35，结合步骤S34的结果，利用最近匹配原则获得角色r下的论元集合$S_{Ar} = \{a = (x_s, x_{s+1}, \dots, x_e) | s \le e;\ s \in As_r, e \in Ae_r\}$；其中，$s$为论元$a$的开始字符在文本X中的位置，$e$为集合$Ae_r$中最靠近$s$的元素。

[0068] 作为一种优选的技术方案，该事件抽取方法的损失函数如下：

[0069]
$$-\log\left(\prod_{x\in D}\left[\prod_{(t,c,a_r)\in E_x} p((t,c,a_r)|x)\right]\right)$$
$$= -\log\left(\prod_{x\in D}\left[\prod_{t\in E_x} p(t|x) \prod_{c\in E_x|t} p(c|x,t) \prod_{a_r\in E_x|t,c} p(a_r|x,t,c)\right]\right)$$

[0070] 其中，$c \in E_x|t$表示输入文本x中触发词t的事件类型，$a_r \in E_x|t,c$表示输入文本x中事件类型为c的触发词t的角色为r的论元，D表示所有输入样本，$p((t,c,a_r)|x)$表示样本x中触发词为t且事件类型为c且角色r的论元为a的概率，$p(t|x)$表示样本x中触发词为t的概率，$p(c|x,t)$表示样本x中触发词t的事件类型为c的概率，$p(a_r|x,t,c)$表示在样本x中事件类别为c且触发词为t时角色r的论元为a的概率，E_x表示样本x中的所有事件。

[0071] 作为一种优选的技术方案，步骤S1中，语言模型为BERT模型。

[0072] 文本向量化模块：用以，将样本$\mathrm{X} = \{x_1, x_2, x_3, \dots, x_i \dots, x_n\}$输入到基于语言模型的文本向量化层当中，获得文本向量化结果$\mathrm{E} = \{e_1, e_2, e_3, \dots e_i, \dots, e_n\}$；其中，$x_i$表示文本中的第$i$个字，$e_i \in R^d$表示$x_i$对应的向量化结果，$E$表示文本X的向量化结果，R表示实数，d表示向量的维度，R^d表示d维实数向量；

[0073] 事件检测模块:用以,将文本向量化结果E输入融合类型感知的门控注意力机制的事件检测模块,以完成触发词检测和事件分类两个子任务;

[0074] 论元抽取模块:用以,对事件检测模块完成触发词检测和事件分类后的结果中每种事件类型下的每一个触发词,利用融合了可学习的角色交互参数的论元抽取模块完成论元抽取和论元角色分类两个子任务。

[0075] 本发明相比于现有技术,具有以下有益效果:

[0076] 本发明利用门控信息指导,使得不同事件类别下有不同信息流向触发词,有效地过滤与触发词无关的噪声并有效地整合其他相关信息,从而更好地消除含重叠触发词的事件抽取场景下的触发词的歧义问题,应对重叠触发词场景,提升事件分类效果;同时,该方法考虑了被大家忽视的角色共现关系,通过引入可学习的角色交互参数建模该角色共现关系,进一步提升了论元抽取和角色分类任务的效果。

附图说明

[0077] 图1为本发明所述的一种基于类型感知门控注意力机制的事件抽取方法的步骤示意图;

[0078] 图2为本发明具体实施方式中融合类型感知门控注意力机制的事件检测模块的模型结构示意图。

具体实施方式

[0079] 下面结合实施例及附图,对本发明作进一步的详细说明,但本发明的实施方式不限于此。

[0080] 本发明提出一种基于类型感知门控注意力机制的事件抽取方法,该方法设计了类型感知的门控注意力机制,区别于纯粹的注意力机制,该方法利用门控信息指导,使得不同事件类别下有不同信息流向触发词,有效地过滤与触发词无关的噪声并有效地整合其他相关信息,从而更好地消除含重叠触发词的事件抽取场景下的触发词的歧义问题,应对重叠触发词场景,提升事件分类效果;同时,该方法考虑了被大家忽视的角色共现关系,通过引入可学习的角色交互参数建模该角色共现关系,进一步提升了论元抽取和角色分类任务的效果。

[0081] 实施例1

[0082] 如图1所示,一种基于类型感知门控注意力机制的事件抽取方法,包括步骤:

[0083] S1,将样本$\mathrm{X}=\{x_1,x_2,x_3,\ldots,x_i\ \ldots,x_n\}$输入到基于语言模型如BERT的文本向量化层当中,获得文本向量化结果$\mathrm{E}=\{\boldsymbol{e_1},\boldsymbol{e_2},\boldsymbol{e_3},\ldots\boldsymbol{e_i},\ldots,\boldsymbol{e_n}\}$。其中$x_i$表示文本中的第$i$个字,$\boldsymbol{e_i}\in R^d$为其对应的向量化结果;

[0084] S2,将文本向量化结果E输入融合类型感知的门控注意力机制的事件检测模块以完成触发词检测和事件分类两个子任务;

[0085] S3,对S2结果中属于每种事件类型的每一个触发词,利用融合了可学习的角色交互参数的论元抽取模块完成论元抽取和论元角色分类两个子任务;

[0086] 实施例2

[0087] 在实施例1的基础上，如图2所示，步骤S2中融合类型感知门控注意力机制的事件检测模块按串联顺序包括如下子模块：触发词提取层、门控注意力事件分类层。

[0088] 触发词提取层的构建过程包括如下步骤：

[0089] S21，首先按照如下公式计算获得输入文本中每个字为触发词开始/结束字符的概率：

$$t_i^{ts} = sigmoid(\boldsymbol{w}_{ts}\boldsymbol{e}_i + b_{ts});$$

[0090]

$$t_i^{te} = sigmoid(\boldsymbol{w}_{te}\boldsymbol{e}_i + b_{te});$$

[0091] 其中，$\boldsymbol{w}_{ts} \in R^d$、$\boldsymbol{w}_{te} \in R^d$、$b_{ts} \in R$、$b_{te} \in R$为可学习的网络参数，sigmoid为激活函数。$t_i^{ts}$对应输入文本中第$i$个字是触发词的开始字符的概率，$t_i^{te}$则对应文本中第$i$个字是触发词的结束字符的概率；

[0092] S22，根据预先设定的阈值δ_1、δ_2对S21中的结果进行过滤，从而获得位置集合Ts、Te：

[0093] $$Ts = \{i | t_i^{ts} > \delta_1\}$$

[0094] $$Te = \{i | t_i^{te} > \delta_2\}$$

[0095] 这里，位置集合Ts、Te分别表示触发词的开始字符位置集合、结束字符位置集合；

[0096] S23，结合步骤S22的结果，利用最近匹配原则获得触发词集合$S_{Trigger} = \{t = (x_s, x_{s+1}, \ldots, x_e) | s \le e;\ s \in Ts, e \in Te\}$，这里$s$为候选触发词t的开始字符在文本X中的位置，$e$为集合$Te$中最靠近$s$的元素。

[0097] 门控注意力事件分类层的构建过程包括如下步骤：

[0098] S24，在门控信息过滤层中，对每个事件类别c，定义事件类别语义向量$\boldsymbol{c} \in R^d$，按如下公式计算相应门控向量：

[0099] $$\boldsymbol{g}^c = sigmoid(\boldsymbol{W}_g\boldsymbol{c} + b_g)$$

[0100] 这里$\boldsymbol{g}^c \in \boldsymbol{R}^d$为事件类别$c$下的门控向量，$\boldsymbol{W}_g \in R^{d*d}$为门控单元的可学习权重参数，$b_g \in R^d$为门控单元的可学习偏置参数；

[0101] S25，结合S24中的结果，在每个事件类别下，利用元素积函数(element-wise product)过滤上下文信息：

[0102] $$\boldsymbol{f}_i^c = \boldsymbol{e}_i \odot \boldsymbol{g}^c$$

[0103] 这里，$e_i \in R^d$为输入文本中第i个字对应向量，$f_i^c \in R^d$为经过门控信息过滤层后输入文本中第i个字在事件类别c下经过信息过滤后的对应向量；

[0104] S26，在注意力信息融合层中，经过S25步骤所述运算后，利用如下设计的注意力计算函数获得在事件类别c下输入文本中第i个字对于触发词$t \in S_{Trigger}$的重要性分数α_i^c：

[0105] $$\alpha_i^c = exp(\theta(t, f_i^c)) / \sum_{k=1}^{n} exp(\theta(t, f_k^c))$$

[0106] 这里，t为触发词$t = (x_s, x_{s+1}, \ldots, x_e)$的表征向量，通过如下公式计算获得：

[0107] $$t = \frac{e_s + e_e}{2}$$

[0108] 此外，$\theta(t, f_i^c)$的定义如下：

[0109] $$\theta(t, f_i^c) = V^T tanh(W[t; f_i^c; |t - f_i^c|; t \odot f_i^c])$$；

[0110] S27，结合S26中计算所获重要性分数，利用如下公式在每个事件类别下获得与每个触发词相关的最终信息聚合结果：

[0111] $$s_t^c = \sum_{i=1}^{n} \alpha_i^c f_i^c$$

[0112] 这里，$s_t^c \in R^d$为经过注意力信息融合层后事件类别c下与触发词t相关的信息聚合向量；

[0113] S28，在事件分类层中，结合步骤S27中所得的与触发词t相关的信息聚合向量，利用如下公式判定触发词t属于哪些事件类型：

[0114] $$\hat{c} = sigmoid(w^T s_t^c + b)$$

[0115] 这里，$w \in R^d$为事件类别判定单元的可学习权重参数，$b \in R$为事件类别判定单元的偏置参数，w^T表示w的转置；具体地，根据预先设定的阈值δ_3，所有满足如下条件的事件类别c均会被判定为触发词t所属事件类别：

[0116] $$\hat{c} > \delta_3$$

[0117] 即最终每个触发词t的事件类别集合为$\dot{C} = \{c | \hat{c} > \delta_3\}$。

[0118] 实施例3

[0119] 在实施例1、2的基础上，步骤S3所述融合了可学习的角色交互参数的论元抽取模块的构建过程包括如下步骤：

[0120] S31，为了更好地抽取事件类型为C的给定触发词t相关的所有论元，首先利用如下公式在上下文中融入触发词表征向量t：

$$z_i^c = \gamma \odot \left(\frac{f_i^c - \mu}{\sigma} + \beta \right);$$

[0121]

$$\gamma = w_\gamma t + b_\gamma;$$

$$\beta = w_\beta t + b_\beta;$$

[0122] 这里，μ、σ是基于输入文本在事件类别C下经过信息过滤后的对应向量f_i^c计算所得的均值与标准方差；

[0123] S32，利用如下公式分别计算输入文本中每个字作为每种角色事件类别r下论元的开始字符/结束字符的概率大小：

$$t_i^{tcrs} = sigmoid(W_{rs} z_i^c + b_{rs});$$

[0124]

$$t_i^{tcre} = sigmoid(W_{re} z_i^c + b_{re});$$

[0125] 这里，t_i^{tcrs}、t_i^{tcre}分别表示输入文本第i个字是事件类别为C的触发词t的角色r下的论元的开始字符/结束字符的概率值；

[0126] S33，定义可学习的角色交互矩阵W_{RR}，并设计如下判定函数：

[0127] $$I(c) = sigmoid(W_{RR}(W_R c + b_R) + b_{RR})$$

[0128] 将其计算结果作为权重，结合该权重修正步骤S32中的计算结果：

[0129] $$\hat{t}_i^{tcrs} = I(c)_r t_i^{tcrs};$$

[0130] $$\hat{t}_i^{tcre} = I(c)_r t_i^{tcre};$$

[0131] 经过训练，判定函数不仅可以学到事件角色之间的相互关系，同时也学到了角色之间的相互关系。$\hat{t}_i^{tcrs}$、$\hat{t}_i^{tcre}$为输入文本第i个字是事件类别为C的触发词t的角色r下的论元的开始字符/结束字符的最终概率值。这里$I(c)_r$为事件类别C下角色r的权重。

[0132] S34，根据预先设定的阈值δ_4、δ_5对S33中的结果进行过滤，从而获得位置集合As_r、Ae_r：

[0133] $$As_r = \{i | \hat{t}_i^{tcrs} > \delta_4\};$$

[0134] $$Ae_r = \{i | \hat{t}_i^{tcre} > \delta_5\};$$

[0135] 这里，位置集合As_r、Ae_r分别表示角色r下论元的开始字符位置集合、结束字符位置集合；

[0136] S35，结合步骤S34的结果，利用最近匹配原则获得角色r下的论元集合$S_{Ar}=\{a=(x_s,x_{s+1},\dots,x_e)|s\le e;\ s\in As_r,e\in Ae_r\}$，这里$s$为论元$a$的开始字符在文本X中的位置，$e$为集合$Ae_r$中最靠近$s$的元素。

[0137] 实施例4

[0138] 在实施例1、2、3的基础上，该事件抽取方法的损失函数如下：

[0139]
$$-\log\left(\prod_{x\in D}\left[\prod_{(t,c,a_r)\in E_x}p((t,c,a_r)|x)\right]\right)$$
$$=-\log\left(\prod_{x\in D}\left[\prod_{t\in E_x}p(t|x)\prod_{c\in E_x|t}p(c|x,t)\prod_{a_r\in E_x|t,c}p(a_r|x,t,c)\right]\right)$$

[0140] 这里，$c\in E_x|t$表示输入文本x中触发词t的事件类型，$a_r\in E_x|t,c$表示输入文本x中事件类型为c的触发词t的角色为r的论元。

[0141] 如上所述，可较好地实现本发明。

[0142] 本说明书中所有实施例公开的所有特征，或隐含公开的所有方法或过程中的步骤，除了互相排斥的特征和/或步骤以外，均可以以任何方式组合和/或扩展、替换。

[0143] 以上所述，仅是本发明的较佳实施例而已，并非对本发明作任何形式上的限制，依据本发明的技术实质，在本发明的精神和原则之内，对以上实施例所作的任何简单的修改、等同替换与改进等，均仍属于本发明技术方案的保护范围之内。


S1，文本向量化；
S2，事件检测；
S3，论元抽取。

图1

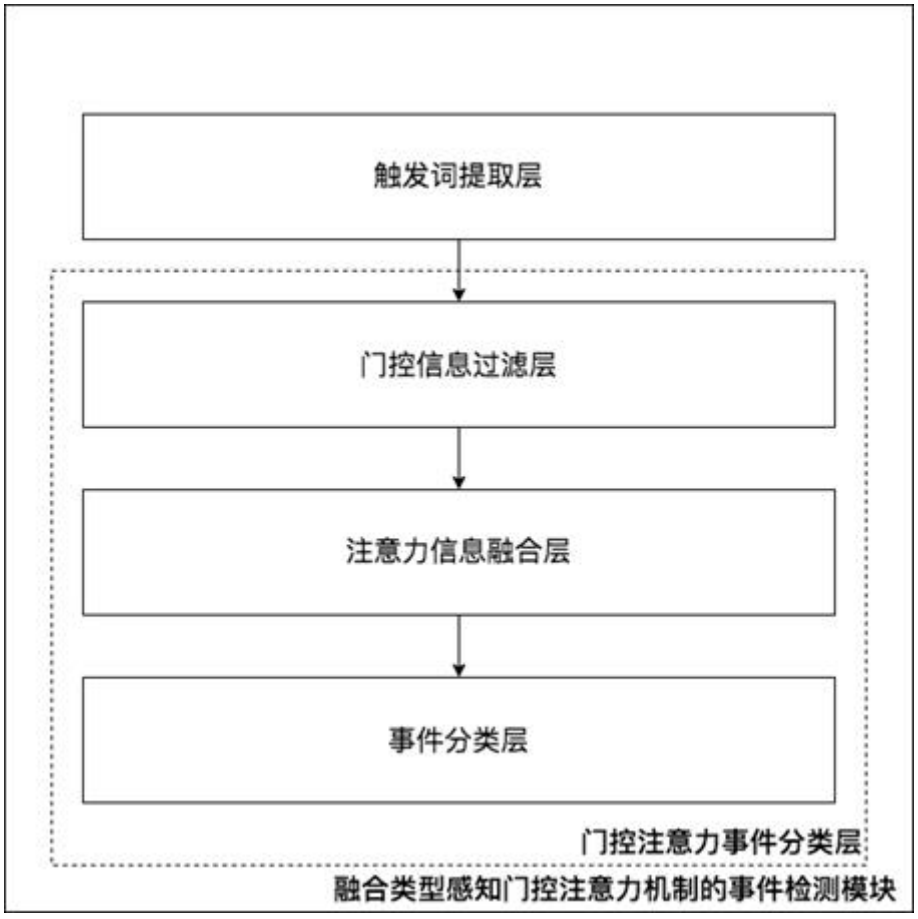

触发词提取层
门控信息过滤层
注意力信息融合层
事件分类层
门控注意力事件分类层
融合类型感知门控注意力机制的事件检测模块

图2