



(12) 发明专利

(10) 授权公告号 CN 114638195 B

(45) 授权公告日 2022. 11. 18

(21) 申请号 202210069337.2

(22) 申请日 2022.01.21

(65) 同一申请的已公布的文献号
申请公布号 CN 114638195 A

(43) 申请公布日 2022.06.17

(73) 专利权人 哈尔滨工业大学(深圳)(哈尔滨
工业大学深圳科技创新研究院)

地址 518055 广东省深圳市南山区桃源街
道深圳大学城哈尔滨工业大学校区

专利权人 东莞理工学院 湖南大学

(72) 发明人 廖清 柴合言 丁烨 李肯立
高翠芸 方滨兴

(74) 专利代理机构 广州三环专利商标代理有限
公司 44202

专利代理师 郭浩辉

(51) Int. Cl.

G06F 40/126 (2020.01)

G06F 40/211 (2020.01)

G06F 40/284 (2020.01)

G06F 16/35 (2019.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

(56) 对比文件

CN 113094596 A, 2021.07.09

CN 113312478 A, 2021.08.27

CN 112257066 A, 2021.01.22

CN 113537593 A, 2021.10.22

WO 2019200806 A1, 2019.10.24

审查员 黄长霞

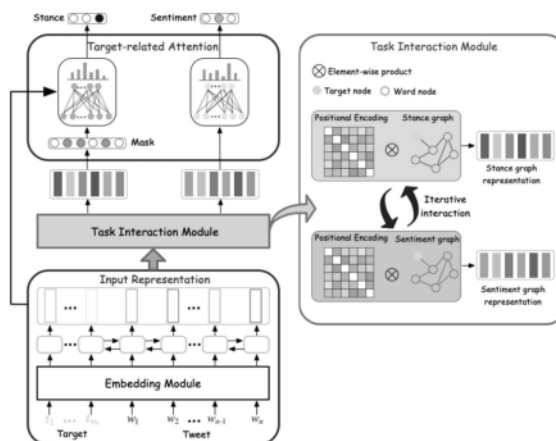
权利要求书2页 说明书9页 附图1页

(54) 发明名称

一种基于多任务学习的立场检测方法

(57) 摘要

本发明公开了一种基于多任务学习的立场检测方法。该方法通过将输入文本输入至多任务图网络模型,得到输入文本的立场检测极性和情感分类极性;所述多任务图网络模型为图卷积神经网络模型,多任务图网络模型包括文本输入处理模块、任务交互模块和任务相关注意力模块;文本输入处理模块用于将输入文本处理成多个词向量;任务交互模块用于构建立场相关图和情感相关图,并采用迭代交互异质图更新方法对立场相关图和情感相关图进行更新;任务相关注意力模块用于根据立场相关图的立场特征表示和情感相关图的情感特征表示计算输入文本的检测立场的极性和分类情感的极性。本发明技术方案提高了针对推文文本进行立场检测的准确性。



1. 一种基于多任务学习的立场检测方法,其特征在于,

将输入文本输入至多任务图网络模型,得到所述输入文本的立场检测极性和情感分类极性;所述多任务图网络模型为图卷积神经网络模型,所述多任务图网络模型包括文本输入处理模块、任务交互模块和任务相关注意力模块;所述输入文本包括推文文本和目标文本;

所述文本输入处理模块用于将输入文本处理成多个词向量,以供所述任务交互模块和任务相关注意力模块使用;

所述任务交互模块用于构建立场相关图和情感相关图,并采用迭代交互异质图更新方法对所述立场相关图和情感相关图进行更新;

所述任务相关注意力模块用于根据所述立场相关图的立场特征表示和情感相关图的情感特征表示计算输入文本的检测立场的极性和分类情感的极性;

所述构建立场相关图和情感相关图,包括以下步骤:

根据所述推文文本的句法结构构建第一句法依赖树,并获取所述第一句法依赖树的根词集合;

根据所述目标文本和所述根词集合之间的单词连接关系,将所述目标文本中的单词添加至所述第一句法依赖树中,得到输入文本的第二句法依赖树;

计算所述推文文本的第一语用权重和第二语用权重,以及所述目标文本的第三语用权重,所述第一语用权重为所述推文文本的各个单词在立场检测任务中的语用权重,所述第二语用权重为所述推文文本的各个单词在情感分类任务中的语用权重;

根据所述第一语用权重、第三语用权重和第二句法依赖树构建所述立场相关图,根据所述第二语用权重和第二句法依赖树构建所述情感相关图。

2. 根据权利要求1所述的基于多任务学习的立场检测方法,其特征在于,所述采用迭代交互异质图更新方法对所述立场相关图和情感相关图进行更新,具体为根据以下公式更新所述立场相关图和情感相关图:

$$\mathbf{g}_l^{st} = \sigma(\hat{E}_{se}^{\frac{1}{2}}(\mathcal{A}^{se} + \mathbf{I})\hat{E}_{se}^{-\frac{1}{2}}\mathbf{g}_{l-1}^{se}\mathbf{W}_{st}^l),$$

$$\mathbf{g}_l^{se} = \sigma(\hat{E}_{st}^{\frac{1}{2}}(\mathcal{A}^{st} + \mathbf{I})\hat{E}_{st}^{-\frac{1}{2}}\mathbf{g}_{l-1}^{st}\mathbf{W}_{se}^l),$$

式中 \mathbf{I} 是单位矩阵, σ 是非线性激活函数, l 表示当前迭代的网络层数, \mathcal{A}^{se} 表示情感任务图的邻接矩阵, \mathcal{A}^{st} 表示立场任务图的邻接矩阵; $\hat{E}_{st} = \sum_j (\mathcal{A}_j^{st} + \mathbf{I})$,

$\hat{E}_{se} = \sum_j (\mathcal{A}_j^{se} + \mathbf{I})$, j 表示邻接矩阵对角线填充1, \mathbf{W}_{task}^l 表示第 l 层网络的参数, $task \in \{st, se\}$; \mathcal{A}_j^{st} 表示立场任务图的邻接矩阵对角线填充1, \mathcal{A}_j^{se} 表示情感任务图的邻接矩阵对角线填充1, \mathbf{g}_l^{st} 表示第 l 层的立场相关图, \mathbf{g}_{l-1}^{st} 表示第 $l-1$ 层的立场相关图, \mathbf{g}_l^{se} 表示第 l 层的情感相关图, \mathbf{g}_{l-1}^{se} 表示第 $l-1$ 层的情感相关图。

3. 根据权利要求2所述的基于多任务学习的立场检测方法,其特征在于,所述多任务图

网络模型目标函数为立场检测与情感分析任务的损失函数的线性组合,具体为:

$$\mathcal{L}(\Theta) = \lambda_1 \sum_{d \in \mathcal{D}} y_d^{st} \log \hat{y}_d^{st} + \lambda_2 \sum_{d \in \mathcal{D}} y_d^{se} \log \hat{y}_d^{se} + \lambda_3 \|\Theta\|_2$$

式中 Θ 为所述多任务图网络模型的参数, λ_1 、 λ_2 、 λ_3 是对应损失项的系数, d 为第 d 条推文文本, \mathcal{D} 为所有推文文本的集合, \hat{y}_d^{se} 为模型所预测出来的第 d 条推文的情感任务标签, \hat{y}_d^{st} 为模型所预测出来的第 d 条推文的立场任务的标签, y_d^{se} 表示模型所预测出来的第 d 条推文的情感任务, y_d^{st} 为模型所预测出来的第 d 条推文的立场任务。

4. 根据权利要求3所述的基于多任务学习的立场检测方法,其特征在于,根据以下公式计算所述立场相关图的立场特征表示 r^{st} :

$$r^{st} = \sum_{i=1}^{m+n} \alpha_i h_i,$$

式中 α 为第 i 个词的立场特征表示的注意力权重, h_i 为第 i 个词经过BERT模型编码后的特征向量, $m+n$ 表示所述输入文本的长度。

5. 根据权利要求4所述的基于多任务学习的立场检测方法,其特征在于,根据以下公式计算所述情感相关图的情感特征表示 r^{se} :

$$r^{se} = \sum_{i=1}^{m+n} \alpha'_i g_i^{se},$$

式中, α' 为第 i 个词的情感特征表示的注意权重, g_i^{se} 为情感相关图的输出 g^{se} 中第 i 个词的特征表示, $m+n$ 表示所述输入文本的长度。

6. 根据权利要求1至5任一项所述的基于多任务学习的立场检测方法,其特征在于,根据以下公式计算输入文本的检测立场和分类情感的极性:

$$y^{\text{task}} = \text{softmax}(W^{\text{task}} r^{\text{task}} + b^{\text{task}});$$

式中 y^{task} 为所述多任务图网络预测的立场检测极性和情感分类极性, W^{task} 为全连接层的权重, b^{task} 为 W^{task} 对应的偏置, r^{task} 为立场特征表示和情感特征表示,softmax为激活函数。

一种基于多任务学习的立场检测方法

技术领域

[0001] 本发明涉及立场检测技术领域,尤其涉及一种基于多任务学习的立场检测方法。

背景技术

[0002] 立场检测是自然语言处理领域中一项基本的文本分析任务,主要内容是:识别文本作者对于某个特定的目标(如实体、事件、人物、观点、主题等)的立场或者态度(例如,支持、反对和保持中立)。立场检测任务对于机器理解自然语言来说是至关重要的。目前的新闻文本都是短文本,运用深度学习方法来识别其立场会造成欠拟合的问题,大大降低了模型的检测准确率。

[0003] 现有的立场检测方法主要分为单任务学习方法和多任务学习方法,其中单任务学习方法大致分为两类:(1)使用传统的机器学习技术,如支持向量机(SVM、决策树、随机森林等)。这类方法需要耗费大量的时间来进行特征工程,人工提取特征,然后输入到机器学习模型里面训练。这种方法代价太高,太依赖手工提取特征,不利于扩展到其他大规模的文本分类任务。(2)使用深度学习方法,构建深度学习模型,不需要手动提特征,让神经网络自动学习到有用的特征,例如,长短期记忆网络(LSTM)、卷积神经网络(CNN)、词向量模型(Bert)、注意力模型(Attention)等。但是由于深度学习模型需要较多的训练数据才能发挥其强大的表征能力,本发明是针对短新闻文本,文本较短,会带来欠拟合问题,从而带来性能的下降。

[0004] 现有基于多任务学习的方法来进行立场检测的方法较少,主要是引入辅助任务(如情感分析任务、目标分类任务、目标文本相关任务等)来帮助主任务(立场检测任务)的性能。这些方法都忽略了联合训练的各个任务基于词级别的交互关系,没有充分利用任务之间的信息共享来提升各个任务的表现。同时也忽略辅助任务的表现,仅仅关注主任务的性能,没有发挥出多任务学习方法的优点,进而降低了模型的整体性能。因此,造成现有技术对推文文本进行立场检测时的准确率较低。

发明内容

[0005] 本发明提供一种基于多任务学习的立场检测方法,提高了针对推文文本进行立场检测的准确性。

[0006] 本发明一实施例提供一种基于多任务学习的立场检测方法,包括以下步骤:

[0007] 将输入文本输入至多任务图网络模型,得到所述输入文本的立场检测极性和情感分类极性;所述多任务图网络模型为图卷积神经网络模型,所述多任务图网络模型包括文本输入处理模块、任务交互模块和任务相关注意力模块;所述输入文本包括推文文本和目标文本;

[0008] 所述文本输入处理模块用于将输入文本处理成多个词向量,以供所述任务交互模块和任务相关注意力模块使用;

[0009] 所述任务交互模块用于构建立场相关图和情感相关图,并采用迭代交互异质图更

新方法对所述立场相关图和情感相关图进行更新；

[0010] 所述任务相关注意力模块用于根据所述立场相关图的立场特征表示和情感相关图的情感特征表示计算输入文本的检测立场的极性和分类情感的极性。

[0011] 进一步的,所述采用迭代交互异质图更新方法对所述立场相关图和情感相关图进行更新,具体为根据以下公式更新所述立场相关图和情感相关图:

$$[0012] \quad \mathbf{g}_l^{st} = \sigma(\hat{E}_{se}^{\frac{1}{2}}(\mathcal{A}^{se} + \mathbf{I})\hat{E}_{se}^{-\frac{1}{2}}\mathbf{g}_{l-1}^{se}\mathbf{W}_{st}^l),$$

$$[0013] \quad \mathbf{g}_l^{se} = \sigma(\hat{E}_{st}^{\frac{1}{2}}(\mathcal{A}^{st} + \mathbf{I})\hat{E}_{st}^{-\frac{1}{2}}\mathbf{g}_{l-1}^{st}\mathbf{W}_{se}^l),$$

[0014] 式中 \mathbf{I} 是单位矩阵, σ 是非线性激活函数, l 表示当前迭代的网络层数, \mathcal{A}^{se} 表示情感任务图的邻接矩阵, \mathcal{A}^{st} 表示立场任务图的邻接矩阵;

$$\hat{E}_{st} = \sum_j (\mathcal{A}_j^{st} + \mathbf{I}),$$

$$\hat{E}_{se} = \sum_j (\mathcal{A}_j^{se} + \mathbf{I}),$$

j 表示邻接矩阵对角线填充1, \mathbf{W}_{task}^l 表示第 l 层网络的参数, $task \in \{st, se\}$ 。

[0015] 进一步的,构建立场相关图和情感相关图,包括以下步骤:

[0016] 根据所述推文文本的句法结构构建第一句法依赖树,并获取所述第一句法依赖树的根词集合;

[0017] 根据所述目标文本和所述根词集合之间的单词连接关系,将所述目标文本中的单词添加至所述第一句法依赖树中,得到输入文本的第二句法依赖树;

[0018] 计算所述推文文本的第一语用权重和第二语用权重,以及所述目标文本的第三语用权重,所述第一语用权重为所述推文文本的各个单词在立场检测任务中的语用权重,所述第二语用权重为所述推文文本的各个单词在情感分类任务中的语用权重;

[0019] 根据所述第一语用权重、第三语用权重和第二句法依赖树构建所述立场相关图,根据所述第二语用权重和第二句法依赖树构建所述情感相关图。

[0020] 进一步的,所述多任务图网络模型目标函数为立场检测与情感分析任务的损失函数的线性组合,所述具体为:

$$[0021] \quad \mathcal{L}(\Theta) = \lambda_1 \sum_{d \in \mathcal{D}} y_d^{st} \log \hat{y}_d^{st} + \lambda_2 \sum_{d \in \mathcal{D}} y_d^{se} \log \hat{y}_d^{se} + \lambda_3 \|\Theta\|_2$$

[0022] 式中 Θ 为所述多任务图网络模型模型的参数, $\lambda_1, \lambda_2, \lambda_3$ 是对应损失项的系数。 d 为第 d 条推文文本, \mathcal{D} 为所有推文文本的集合, \hat{y}_d^{se} 为模型所预测出来的第 d 条推文的情感任务标签,

\hat{y}_d^{st} 为模型所预测出来的第 d 条推文的立场任务的标签, $y^{\text{task}}|_{\text{task}=\{st, se\}}$ 为所述多图稀疏交互网络模型预测的立场检测极性和情感分类极性。

[0023] 进一步的,根据以下公式计算所述立场相关图的立场特征表示 \mathbf{r}^{st} :

$$[0024] \quad \mathbf{r}^{st} = \sum_{i=1}^{m+n} \alpha_i \mathbf{h}_i,$$

[0025] 式中 α 为立场特征表示的注意力权重, h_i 为第*i*个词经过BERT模型编码后的特征向量, $m+n$ 表示所述输入文本的长度。

[0026] 进一步的,根据以下公式计算所述情感相关图的情感特征表示 r^{se} :

$$[0027] \quad r^{se} = \sum_{i=1}^{m+n} \alpha_i' g_i^{se},$$

[0028] 式中 α_i' 为情感特征表示的注意权重, g_i^{se} 为情感相关图的输出 g^{se} 中第*i*个节点的特征表示, $m+n$ 表示所述输入文本的长度。

[0029] 进一步的,根据以下公式计算输入文本的检测立场和分类情感的极性:

$$[0030] \quad y^{task} = \text{softmax}(W^{task} r^{task} + b^{task});$$

[0031] 式中 $y^{task}|_{task=\{st,se\}}$ 为所述多图稀疏交互网络模型预测的立场检测极性和情感分类极性, $W^{task}|_{task=\{st,se\}}$ 为全连接层的权重, b^{task} 为 W^{task} 对应的偏置, $r^{task}|_{task=\{st,se\}}$ 为立场特征表示和情感特征表示,softmax为激活函数。

[0032] 本发明的实施例,具有如下有益效果:

[0033] 本发明提供了一种基于多任务学习的立场检测方法,该方法通过采用迭代交互异质图更新方法将情感分析任务联合立场检测任务进行模型训练以帮助任务进行信息共享,并通过构建立场相关图和情感相关图来捕捉任务细粒度的交互关系,即不同的单词对于不同的任务的重要性不同,有针对性的捕捉任务之间基于词级别的关系,从而促进了任务之间有用信息的共享,提升了所述多任务图网络模型针对立场检测任务和情感分析任务的性能,即提高了针对推文文本进行立场检测的准确性。

附图说明

[0034] 图1是本发明一实施例提供的多任务图网络模型的结构示意图。

具体实施方式

[0035] 下面将结合本发明中的附图,对本发明中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0036] 如图1所示,本发明一实施例提供了一种基于多任务学习的立场检测方法,包括以下步骤:

[0037] 将输入文本输入至多任务图网络模型,得到所述输入文本的立场检测极性和情感分类极性;所述多任务图网络模型为图卷积神经网络模型,所述多任务图网络模型包括文本输入处理模块、任务交互模块和任务相关注意力模块;所述输入文本包括推文文本和目标文本;

[0038] 所述文本输入处理模块用于将输入文本处理成多个词向量,以供所述任务交互模块和任务相关注意力模块使用;

[0039] 所述任务交互模块用于构建立场相关图和情感相关图,并采用迭代交互异质图更新方法对所述立场相关图和情感相关图进行更新;

[0040] 所述任务相关注意力模块用于根据所述立场相关图的立场特征表示和情感相关

图的情感特征表示计算输入文本的检测立场的极性和分类情感的极性。

[0041] 作为其中一种实施例,所述文本输入处理模块使用BERT来编码推文文本内容C和目标文本中的目标词T。首先将推文文本内容C和目标词T拼接起来形成输入文本S,其中 $S = \{t_1, \dots, t_m, w_1, w_2, \dots, w_n\}$ 。然后将处理成BERT模型的输入格式: $[\text{CLS}] t_1 \dots t_m w_1 w_2 \dots w_n [\text{SEP}]$ 。再将其输入到BERT网络模型以捕捉输入文本的上下文特征。其过程可由如下公式定义:

[0042] $H = \text{BERT}(S)$

[0043] 其中,H是BERT网络模型的输出, $H = \{h_1, h_2, \dots, h_{m+n}\}$ 。其中H中的每一个元素为输入文本中一个词的特征表示, $h_t \in \mathbb{R}^{2d}$ 为第t个词包含上下文信息的特征表示。所述文本输入处理模块主要用于处理输入文本,包括推文文本内容和目标文本内容,将其处理成计算机能够处理的格式,即将每个单词映射成一个词向量,然后将所有单词的词向量组成一个矩阵,就得到整个输入文本的向量矩阵,最后使用一个Bi-LSTM网络模型来训练输入文本的向量矩阵,使得每个单词的词向量能够捕捉与上下文相关的特征信息,从而使得词向量能够适应输入文本的内容。

[0044] 作为其中一种实施例,构建所述立场相关图和情感相关图,包括以下步骤:

[0045] 根据所述推文文本的句法结构构建第一句法依赖树,并获取所述第一句法依赖树的根词集合;

[0046] 根据所述目标文本和所述根词集合之间的单词连接关系,将所述目标文本中的单词添加至所述第一句法依赖树中,得到输入文本的第二句法依赖树;

[0047] 计算所述推文文本的第一语用权重和第二语用权重,以及所述目标文本的第三语用权重,所述第一语用权重为所述推文文本的各个单词在立场检测任务中的语用权重,即利用单词与立场标签(“支持”和“反对”)在整个语料中的相对共现频率和词频计算得到词关于立场任务的语用权重;所述第二语用权重为所述推文文本的各个单词在情感分类任务中的语用权重,即利用单词与情感标签在整个语料中的相对共现频率和词频计算得到词关于立场任务的语用权重。所述语用权重是指推文中的词对于特定的目标的依赖关系(或影响)。

[0048] 根据所述第一语用权重、第三语用权重和第二句法依赖树构建所述立场相关图,根据所述第二语用权重和第二句法依赖树构建所述情感相关图。

[0049] 具体地,使用句法分析工具对每条推文文本进行句法分析以构建句法依赖树;并通过将单词表示为图的节点(即图的节点为推文文本中的每个词的词向量),将句法依赖树中单词之间的关系表示为节点之间的连接线,以构建目标文本T和推文文本C的基图 \mathcal{G} ,包括以下步骤:

[0050] 根据所述推文文本C的句法结构构建第一句法依赖树 $\mathcal{T}^c \in \mathbb{R}^{n \times n}$,并通过句法解析器获取所述第一句法依赖树的根词集合 w^r 。

[0051] 由于目标文本T不是一个完整的句子,而是一个短语或者单词,无法被建模成一棵句法依赖树,因此,本发明实施例根据所述目标文本和所述根词集合之间的单词连接关系,将所述目标文本中的单词添加至所述第一句法依赖树中,得到所述输入文本S的第二句法依赖树 \mathcal{T} ,所述第二句法依赖树 \mathcal{T} 的计算公式如下:

$$[0052] \quad \mathcal{T}(w_i, w_j) = \begin{cases} 1 & \text{if } w_i \in T \text{ and } w_j \in w^r \\ \mathcal{T}^c(w_i, w_j) & \text{otherwise} \end{cases};$$

[0053] 式中, w^r 表示第一句法依赖树 \mathcal{T}^c 的根词, $\mathcal{T} \in \mathbb{R}^{(m+n) \times (m+n)}$ 表示输入文本S的第二句法依赖树 \mathcal{T} , w_i 和 w_j 表示推文文本和目标文本构成的输入文本S中的任意两个不相同的词。

[0054] 为了捕捉输入文本中单词的重要性和单词之间的交互特征,需要计算不同任务的单词的语用权重和单词的词频。计算所述输入文本中每个单词在整个语料库中出现的频率

$$p(w_i) = \frac{N(w_i)}{N}, \text{ 式中 } N(w_i) \text{ 为单词 } w_i \text{ 出现在语料库中的次数, } N \text{ 为语料库中所有的单词的数量。}$$

本发明实施例针对不同的任务计算单词在不同任务中的语用权重,包括第一语用权重 $\phi^{\text{task}1}(w_i) |_{\text{task}=\text{stance}}$ 和第二语用权重 $\phi^{\text{task}}(w_i) |_{\text{task}=\text{sentiment}}$,在计算第一语用权重和第二语用权重时只考虑了有实际意义的类别来计算,忽略了中立的态度和中性的情感这两个不包含有用信息的类别,其具体计算过程如下式所示:

$$[0055] \quad \rho^{\text{task}}(w_i) = \left| \frac{N(w_i, \text{label}_+)}{N(\text{label}_+)} - \frac{N(w_i, \text{label}_-)}{N(\text{label}_-)} \right|;$$

$$[0056] \quad \phi^{\text{task}}(w_i) = \frac{\rho^{\text{task}}(w_i) - \mu(\rho^{\text{task}})}{\delta(\rho^{\text{task}})}, w_i \in C;$$

[0057] 式中 $N(w_i, \text{label}_+)$ 和 $N(w_i, \text{label}_-)$ 分别表示词 w_i 出现在立场任务标签“支持”和“反对”中的数量,或者分别表示词 w_i 出现在情感任务标签“积极”和“消极”中的数量; $N(\text{label}_+)$ 和 $N(\text{label}_-)$ 分别表示立场任务标签“支持”和“反对”的总数量,或者分别表示情感任务标签“积极”和“消极”的总数量, μ 为均值, δ 为标准差。

[0058] 根据公式 $\phi^{\text{task}}(s_j) = 1, s_j \in T \text{ and } s_j \notin C$,计算所述目标文本的第三语用权重。所述目标文本的第三语用权重用于建立目标文本和推文文本之间的联系,即构造图中目标文本和推文文本之间的边。

[0059] 根据所述第一语用权重、第三语用权重和第二句法依赖树计算所述立场相关图的第一邻接矩阵,进而根据所述第一邻接矩阵得到所述立场相关图。

[0060] 根据所述第二语用权重和第二句法依赖树计算所述情感相关图的第二邻接矩阵,进而根据所述二邻接矩阵得到所情感相关图。

[0061] 具体地,根据以下公式计算所述第一邻接矩阵和第二邻接矩阵:

$$[0062] \quad \mathcal{A}_{i,j}^{\text{task}} = \begin{cases} \sum_{t \in \{i,j\}} \phi^{\text{task}}(s_t) p(s_t) & \text{if } \mathcal{T}(s_i, s_j) = 1, s_t \in S; \\ 0 & \text{otherwise} \end{cases};$$

[0063] 式中 s_j 和 s_i 为输入文本中的单词。

[0064] 作为其中一种实施例,采用迭代交互异质图更新方法对所述立场相关图和情感相关图进行更新;具体地,本发明实施例采用一种迭代交互异质图更新方法来训练立场相关图和情感相关图,以进行任务之间基于词级别的交互,从而提升所述多任务图网络模型对于各个任务的表现力。所述迭代交互异质图更新方法具体为:首先将图的节点初始化为经过BERT编码的词向量,经过一轮迭代后,分别获得立场图的节点表示和情感图的节点表示,

然后将立场图的节点特征 \mathbf{g}_{l-1}^{st} 输入到情感图中进行立场图和情感图交换, 将情感图的节点特征 \mathbf{g}_{l-1}^{se} 输入到立场图中进行两个任务之间的交互, 以捕捉任务之间的关系。其中通过卷积神经网络对第1层的图的结点特征来行更新, 具体过程如下:

[0065] 根据以下公式更新所述立场相关图和情感相关图(即采用迭代交互异质图更新方法进行更新):

$$[0066] \quad \mathbf{g}_l^{st} = \sigma(E_{se}^{\frac{1}{2}}(\mathcal{A}^{se} + \mathbf{I})E_{se}^{-\frac{1}{2}}\mathbf{g}_{l-1}^{se}\mathbf{W}_{st}^l),$$

$$[0067] \quad \mathbf{g}_l^{se} = \sigma(E_{st}^{\frac{1}{2}}(\mathcal{A}^{st} + \mathbf{I})E_{st}^{-\frac{1}{2}}\mathbf{g}_{l-1}^{st}\mathbf{W}_{se}^l),$$

[0068] 式中 \mathbf{I} 是单位矩阵, σ 是非线性激活函数, $\hat{E}_{st} = \sum_j (\mathcal{A}_j^{st} + \mathbf{I})$,

$\hat{E}_{se} = \sum_j (\mathcal{A}_j^{se} + \mathbf{I})$, j 表示邻接矩阵对角线填充1, \hat{E}_{st} 和 \hat{E}_{se} 为图卷积网络卷积核的特征向量和常规图卷积神经网络更新的参数; 1 表示图的网络层数, \mathcal{A}^{se} 为第二邻接矩阵, \mathcal{A}^{st} 为第一邻接矩阵。

[0069] 本发明实施例的任务交互模块用于构建立场相关图和情感相关图, 将输入文本中的每个单词表示为图的节点, 单词的任务相关的语用权重为边, 分别构造立场相关图和情感相关图; 同时对所述立场相关图和情感相关图进行基于词级别的迭代交互更新(即采用迭代交互异质图更新方法进行更新), 使得各个任务能够吸收来自于其他任务的对任务训练有帮助的信息, 从而能够获得包含更丰富特征的任务表达, 帮助所述多任务图网络模型提高针对各个任务的预测准确率。

[0070] 作为其中一种实施例, 对所述立场相关图和情感相关图进行迭代更新后, 得到立场相关图和情感相关图的输出, 即立场相关图节点和情感相关图节点更新后的表征, 然后将图的输出输入到任务相关的注意力模块中, 获得与目标词更相关的更丰富的立场任务表征和情感任务的表征, 再根据得到的立场任务表征和情感任务表征来进行立场检测和情感分类。具体地, 在所述任务相关注意力模块获取立场相关图的立场特征表示 \mathbf{g}^{st} 和情感相关图的情感特征表示 \mathbf{g}^{se} 。对于立场检测任务来说, 为了获得目标相关的立场特征表示, 需要采用掩码机制过滤掉非目标词, 具体地, 通过设计 mask 矩阵, 将目标词对应位置置1, 非目标词对应位置置0, 从而得到 mask 后的立场相关图的特征表示 $\mathbf{g}_{mask}^{st} = \text{mask} \odot \mathbf{g}^{st}$ 。再使用基于检索的注意力机制来获取与目标词相关的更丰富的立场特征表示, 所述立场特征表示的注意力权重 α 根据下式计算得出:

$$[0071] \quad \beta_t = \sum_{i=1}^{m+n} \mathbf{h}_t^\top \mathbf{g}_{mask}^{st,i} = \sum_{i=1}^m \mathbf{h}_t^\top \mathbf{g}_{mask}^{st,i};$$

$$[0072] \quad \alpha_t = \frac{\exp(\beta_t)}{\sum_{i=1}^n \exp(\beta_i)};$$

[0073] 式中, h 为所述BERT网络模型的输出, h_t^T 表示第 t 个词经过BERT网络模型编码后的特征表示; $m+n$ 表示输入文本中词的长度(目标文本的长度为 m , 推文文本的长度 n), $g_{mask}^{st,i}$ 表示立场相关图经过mask矩阵转换后的输出 g_{mask}^{st} 中第 i 个节点(即第 i 个词向量)的特征表示, β_t 表示第 t 个词向量的注意力权重, β_i 为第 i 个词向量的情感表征的注意力权重, α_t 表示第 t 个词向量的注意力权重经过归一化的注意力权重。所述立场特征表示的注意力权重表示推文中所有词对立场表征的注意力。

[0074] 再根据下式计算立场相关图的立场特征表示 r^{st} :

$$[0075] \quad r^{st} = \sum_{i=1}^{m+n} \alpha_i h_i,$$

[0076] h_i 为第 i 个词经过所述BERT模型编码后的特征向量, α_i 为第 i 个词向量的立场特征表示的注意力权重。

[0077] 同理, 根据公式(4) - (6) 计算情感相关图的情感特征表示 r^{se} :

$$[0078] \quad \beta_t' = \sum_{i=1}^{m+n} h_t^T g_i^{se} \quad (4);$$

$$[0079] \quad \alpha_t' = \frac{\exp(\beta_t')}{\sum_{i=1}^n \exp(\beta_i')} \quad (5);$$

$$[0080] \quad r^{se} = \sum_{i=1}^{m+n} \alpha_i' g_i^{se} \quad (6);$$

[0081] g_i^{se} 表示情感相关图的输出 g^{se} 中第 i 个节点的特征表示, $m+n$ 表示输入文本中词的长度(目标文本的长度为 m , 推文文本的长度 n), h_t^T 表示第 t 个词经过BERT编码后的特征表示, β_t' 为第 t 个词向量的情感表征的注意力权重, β_i' 为第 i 个词向量的情感表征的注意力权重。

[0082] 得到最终的立场特征表示 r^{st} 和情感特征表示 r^{se} 后, 使用一层全连接层来融合文本特征和丰富的上下文特征, 并得到所述输入文本的检测立场的极性和分类情感的极性:

$$[0083] \quad y^{\text{task}} = \text{softmax}(W^{\text{task}} r^{\text{task}} + b^{\text{task}})$$

[0084] 式中, $y^{\text{task}}|_{\text{task}=\{\text{st}, \text{se}\}}$ 为所述多任务图网络模型预测的立场检测极性和情感分类极性, $W^{\text{task}}|_{\text{task}=\{\text{st}, \text{se}\}}$ 为全连接层的权重, b^{task} 为对应的偏置, softmax 为激活函数。

[0085] 最后, 整个多图稀疏交互模块的目标函数为立场检测任务与情感分析任务的损失函数的线性组合:

$$[0086] \quad \mathcal{L}(\Theta) = \lambda_1 \sum_{d \in \mathcal{D}} y_d^{st} \log \hat{y}_d^{st} + \lambda_2 \sum_{d \in \mathcal{D}} y_d^{se} \log \hat{y}_d^{se} + \lambda_3 \|\Theta\|_2$$

[0087] 式中 Θ 为所述多任务图网络模型模型的参数, λ_1 、 λ_2 、 λ_3 是对应损失项的系数。 d 为第 d 条推文文本, \mathcal{D} 为所有推文文本的集合, \hat{y}_d^{se} 为模型所预测出来的第 d 条推文的情感任务标签, \hat{y}_d^{st} 为模型所预测出来的第 d 条推文的立场任务的标签。

[0088] 本发明实施例的任务相关注意力模块用于计算最终的所述立场相关图的立场特征表示和情感相关图的情感特征表示, 并根据立场特征表示和情感特征表示来进行立场检

测和情感分类。所述任务相关注意力模块首先是聚合立场相关图和情感相关图的输出,获得立场相关图的特征表示和情感相关图的特征表示,然后使用注意力机制来分别计算与立场任务的目标词相关的推文文本词的注意力权重和情感任务的推文文本词之间的注意力权重,最后加权求和得到最终的立场任务的表达和情感任务表达(即最终的立场特征表示和情感特征表示)。

[0089] 现有的多任务立场检测方法没有深入地考虑到任务之间的交互关系,也没有充分进行任务间信息共享,因此无法提高立场检测和情感分类的准确率。本发明通过构建异构的任务相关语用图来捕捉单词之间与任务相关的语用关系,本发明实施例能同时检测推文文本的立场和识别出其情感极性,并能提高其准确性。同时,通过为每个任务构建异构的任务相关语用图和设计任务交互模块,以捕获各个任务的词级交互。

[0090] 本发明实施例致力于解决目前基于多任务学习的立场检测方法忽略了任务之间细粒度的交互从而导致任务表现下降的问题,即目前的多任务立场检测方法仅仅使用硬参数共享的方式简单地联合两个任务进行训练,没有考虑到多任务细粒度的交互问题,从而带来性能瓶颈。本发明实施例通过情感分析任务联合立场检测任务进行模型训练。并通过构建任务相关的图来捕捉了任务细粒度的交互关系,即不同的单词对于不同的任务的重要性不同,有针对性的捕捉任务之间基于词级别的关系,从而促进任务之间有用信息的共享。本发明通过构建任务相关的图来捕捉任务之间基于词级别的交互关系,从而促进任务之间有用信息的共享,提升了所述多任务图网络模型针对立场检测任务和情感分析任务的性能。

[0091] 目前基于多任务学习的立场检测方法存在以下问题:(1)都是采用辅助任务形式,即不关注辅助任务的性能,辅助任务只用来帮助提升主任务的表现,这跟多任务学习的目标是不一致的(多任务学习方法能够同时提升所有任务的表现)。本发明实施例提出的基于多任务学习的立场检测方法能够同时提升立场检测的准确率和情感分类的准确率;(2)现有的多任务学习的立场检测方法在进行任务之间信息交互时,仅仅考虑简单地提取两个任务的特征然后简单拼接起来进行交互,这是基于整个任务特征层面的信息交互(即只考虑了句子层面之间的信息交互),这种粗粒度的交互关系,忽略了任务之间的关系的复杂性,认为任务之间的交互都是对称等价的,实际上,任务之间的关系是十分复杂的且在不同的维度上是不一样的。现有技术完全没有考虑到输入文本中的词对于不同任务的重要性是不同的这个问题,在设计任务交互时没有充分考虑到任务之间在词水平上的关系。从而导致了现有的模型无法保证各个任务之间的信息共享是充分的,即大部分情况下任务之间进行共享的信息对整个多任务模型没有用的,不能帮助立场检测模型提高性能,反而带来会带来很大的信息冗余,从而在整个任务共享网络中产生较大的负迁移(即任务A从任务B中吸收信息后,反而降低了任务A的表现),一旦任务共享中产生了很大的负迁移,带来了很大的任务噪声,那么所有任务的性能都会下降。针对任务之间信息交互产生较大负迁移问题,本发明实施例提出一种在词级别的多任务交互方法,让各个任务在交互时能够学到哪些词之间信息共享是有利于提升多任务性能的,哪些词之间的交互信息是对模型没有用的会产生较大负迁移和影响多任务模型的性能,通过抑制这种产生负迁移的词之间的信息共享,从而提升多任务共享网络的效率,最终使得所述多任务图网络模型的所有任务都能在任务共享中吸收到有用的信息,提高了所述多任务图网络模型的任务表现力。

[0092] 以上所述是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也视为本发明的保护范围。

[0093] 本领域普通技术人员可以理解实现上述实施例中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的程序可存储于一计算机可读取存储介质中,该程序在执行时,可包括如上述各实施例的流程。其中,所述的存储介质可为磁碟、光盘、只读存储记忆体 (Read-Only Memory, ROM) 或随机存储记忆体 (Random Access Memory, RAM) 等。

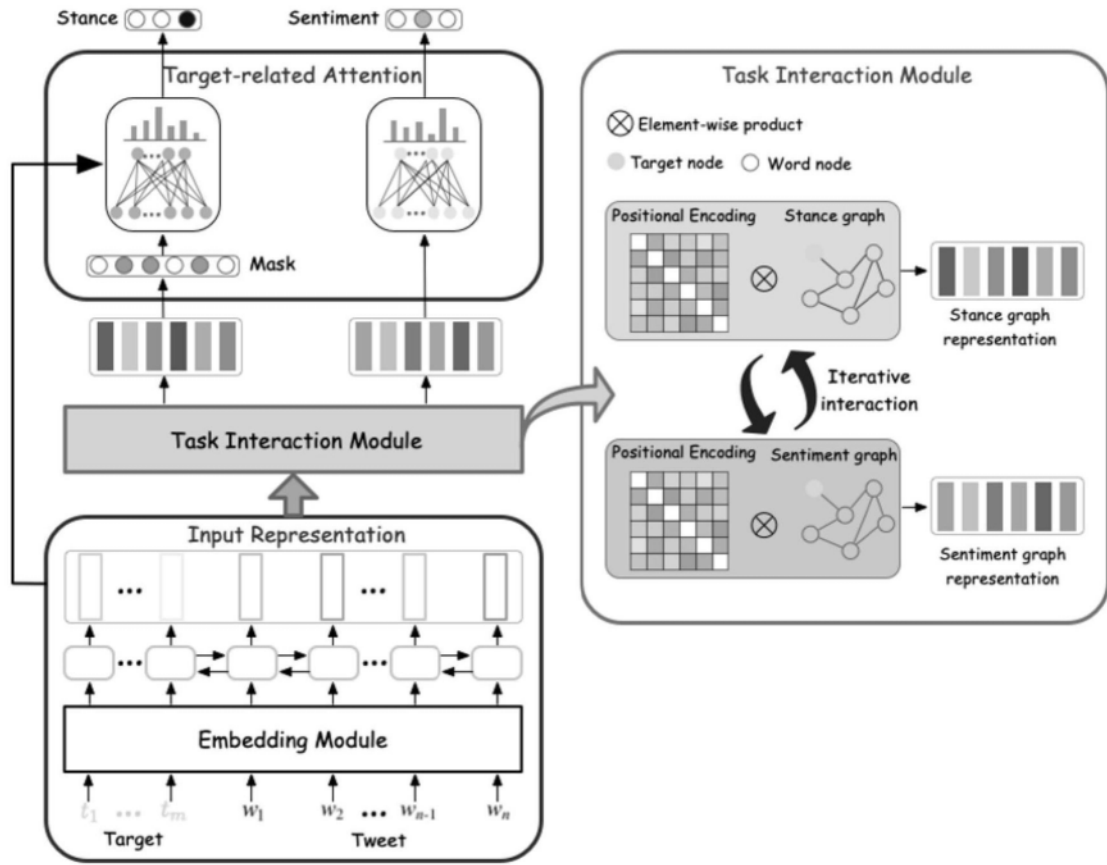


图1