



(12) 发明专利申请

(10) 申请公布号 CN 112395417 A

(43) 申请公布日 2021.02.23

(21) 申请号 202011292823.8

(22) 申请日 2020.11.18

(71) 申请人 长沙学院

地址 410000 湖南省长沙市开福区洪山路
98号

(72) 发明人 庄穆妮 李勇 谭旭 吕欣 管慧

(51) Int. Cl.

G06F 16/35 (2019.01)

G06F 16/36 (2019.01)

G06K 9/62 (2006.01)

G06Q 50/00 (2012.01)

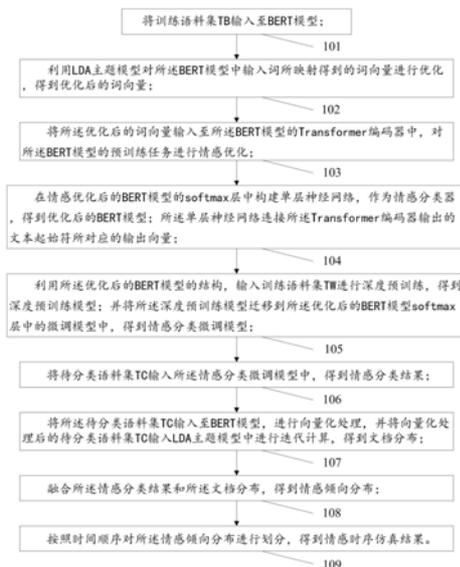
权利要求书2页 说明书7页 附图2页

(54) 发明名称

基于深度学习的网络舆情演化仿真方法及系统

(57) 摘要

本发明涉及基于深度学习的网络舆情演化仿真方法及系统,涉及自然语言处理中的文本情感分析技术领域。对BERT模型的预训练任务进行改进,并在此基础上叠加深度预训练任务,还借助于与LDA主题模型深度融合以实现主题视角下细粒度的舆情仿真分析。将待分类语料集TC输入情感分类微调模型中,得到情感分类结果,再将待分类语料集TC输入至BERT模型,进行向量化处理,并将向量化处理后的待分类语料集TC输入LDA主题模型中进行迭代计算,得到文档分布,然后融合情感分类结果和文档分布,得到情感倾向分布,并按照时间顺序对情感倾向分布进行划分,得到情感时序仿真结果,从而得到更加细粒度和精准的文本主题聚类与舆情演化仿真结果。



CN 112395417 A

1. 一种基于深度学习的网络舆情演化仿真方法,其特征在于,所述仿真方法包括如下步骤:

将训练语料集TB输入至BERT模型;

利用LDA主题模型对所述BERT模型中输入词所映射得到的词向量进行优化,得到优化后的词向量;

将所述优化后的词向量输入至所述BERT模型的Transformer编码器中,对所述BERT模型的预训练任务进行情感优化;

在情感优化后的BERT模型的softmax层中构建单层神经网络,作为情感分类器,得到优化后的BERT模型;所述单层神经网络连接所述Transformer编码器输出的文本起始符所对应的输出向量;

利用所述优化后的BERT模型的结构,输入训练语料集TW进行深度预训练,得到深度预训练模型;并将所述深度预训练模型迁移到所述优化后的BERT模型softmax层中的微调模型中,得到情感分类微调模型;

将待分类语料集TC输入所述情感分类微调模型中,得到情感分类结果;

将所述待分类语料集TC输入至BERT模型,进行向量化处理,并将向量化处理后的待分类语料集TC输入LDA主题模型中进行迭代计算,得到文档分布;

融合所述情感分类结果和所述文档分布,得到情感倾向分布;

按照时间顺序对所述情感倾向分布进行划分,得到情感时序仿真结果。

2. 如权利要求1所述的一种基于深度学习的网络舆情演化仿真方法,其特征在于,所述训练语料集TB包括中文维基百科语料以及新浪微博和百度贴吧的情感语料。

3. 如权利要求1所述的一种基于深度学习的网络舆情演化仿真方法,其特征在于,在将训练语料集TB输入至BERT模型之前,所述仿真方法还包括对训练语料集TB进行分词、去停用词的数据预处理。

4. 如权利要求1所述的一种基于深度学习的网络舆情演化仿真方法,其特征在于,所述利用LDA主题模型对所述BERT模型中输入词所映射得到的词向量进行优化,得到优化后的词向量具体包括:

将训练语料集TB输入至BERT模型,进行向量化处理,所述训练语料集TB中的每个词均被映射成一个包含词义向量、文本向量和位置向量的词向量;

将向量化处理后的训练语料集TB输入至LDA主题模型,得到优化后的主题向量,将所述优化后的主题向量嵌入所述BERT模型,对所述BERT模型中输入词所映射得到的词向量进行优化,得到优化后的词向量;所述优化后的词向量包括词义向量、文本向量、位置向量和主题向量。

5. 如权利要求1所述的一种基于深度学习的网络舆情演化仿真方法,其特征在于,所述情感分类器用于执行SC任务。

6. 如权利要求5所述的一种基于深度学习的网络舆情演化仿真方法,其特征在于,所述BERT模型的softmax层中还包括一个连接在Transformer编码器输出的文本起始符所对应的输出向量的单层神经网络,以执行NSP任务。

7. 如权利要求6所述的一种基于深度学习的网络舆情演化仿真方法,其特征在于,所述BERT模型的softmax层中还包括一个连接在Transformer编码器输出的遮盖字所对应的输

出向量的单层神经网络,以执行MLM任务。

8. 如权利要求7所述的一种基于深度学习的网络舆情演化仿真方法,其特征在于,所述利用所述优化后的BERT模型的结构,输入训练语料集TW进行深度预训练,得到深度预训练模型具体包括:

设置初始学习率、批量大小、dropout率和Epoch训练次数;

将所述MLM任务、所述NSP任务和所述SC任务的损失和作为深度预训练的损失函数;

输入训练语料集TW进行深度预训练,在训练过程中,采用Adam优化算法确定最佳学习率,使用AUC作为检验模型训练效果的指标,并采用F1分数确定正、负情感分类的最佳阈值,得到深度预训练模型。

9. 如权利要求1所述的一种基于深度学习的网络舆情演化仿真方法,其特征在于,所述Transformer编码器将多头机制和前馈层通过残差网络结构连接。

10. 一种基于深度学习的网络舆情演化仿真系统,其特征在于,所述仿真系统包括:

输入模块,用于将训练语料集TB输入至BERT模型;

词向量优化模块,用于利用LDA主题模型对所述BERT模型中输入词所映射得到的词向量进行优化,得到优化后的词向量;

情感优化模块,用于将所述优化后的词向量输入至所述BERT模型的Transformer编码器中,对所述BERT模型的预训练任务进行情感优化;

优化BERT模型获取模块,用于在情感优化后的BERT模型的softmax层中构建单层神经网络,作为情感分类器,得到优化后的BERT模型;所述单层神经网络连接所述Transformer编码器输出的文本起始符所对应的输出向量;

情感分类微调模型获取模块,用于利用所述优化后的BERT模型的结构,输入训练语料集TW进行深度预训练,得到深度预训练模型;并将所述深度预训练模型迁移到所述优化后的BERT模型softmax层中的微调模型中,得到情感分类微调模型;

情感分类结果获取模块,用于将待分类语料集TC输入所述情感分类微调模型中,得到情感分类结果;

文档分布获取模块,用于将所述待分类语料集TC输入至BERT模型,进行向量化处理,并将向量化处理后的待分类语料集TC输入LDA主题模型中进行迭代计算,得到文档分布;

情感倾向分布获取模块,用于融合所述情感分类结果和所述文档分布,得到情感倾向分布;

情感时序仿真结果获取模块,用于按照时间顺序对所述情感倾向分布进行划分,得到情感时序仿真结果。

基于深度学习的网络舆情演化仿真方法及系统

技术领域

[0001] 本发明涉及自然语言处理中的文本情感分析技术领域,特别是涉及一种基于深度学习的重大突发事件网络舆情演化仿真方法及系统。

背景技术

[0002] 传统的舆情情感演化仿真实常用方法为情感词典法和有监督的机器学习方法。近年来,深度学习逐渐取代了传统的情感仿真,长短期记忆神经网络(Long Short-Term Memory,LSTM)是深度学习领域中被广泛运用的特殊递归神经网络,通过内嵌遗忘门和输入门机制,一定程度上缓解了一般循环神经网络(Recurrent Neural Network,RNN)无法解决的长距离依赖问题。然而,对于特别长期的依赖,LSTM依然无能为力。因此,Vaswani等人构建了Transformer模型,通过自注意力机制仿真建模取代传统的Word2Vec词向量嵌入,有效解决了自然语言处理(Natural Language Processing,NLP)中棘手的长期依赖问题。随后,Devlin等人在Transformer模型的结构基础上构建了BERT预训练模型,一举打破GLUE benchmark 11项NLP测试任务的最高纪录。

[0003] 尽管BERT在多项NLP任务中表现卓越,但由于BERT在预训练阶段缺乏大规模情感语料的输入,使其在执行情感分析任务时存在一定瓶颈。此外,基于BERT的情感仿真难以对隐藏在海量文本背后的语义进行主题细粒度挖掘。

发明内容

[0004] 本发明的目的是提供一种基于深度学习的网络舆情演化仿真方法及系统,对BERT模型的预训练任务进行改进,并在此基础上叠加深度预训练任务,还借助于与LDA主题模型深度融合以实现主题视角下细粒度的舆情仿真分析,从而实现更精准的舆情情感分类。

[0005] 为实现上述目的,本发明提供了如下方案:

[0006] 一种基于深度学习的网络舆情演化仿真方法,所述仿真方法包括如下步骤:

[0007] 将训练语料集TB输入至BERT模型;

[0008] 利用LDA主题模型对所述BERT模型中输入词所映射得到的词向量进行优化,得到优化后的词向量;

[0009] 将所述优化后的词向量输入至所述BERT模型的Transformer编码器中,对所述BERT模型的预训练任务进行情感优化;

[0010] 在情感优化后的BERT模型的softmax层中构建单层神经网络,作为情感分类器,得到优化后的BERT模型;所述单层神经网络连接所述Transformer编码器输出的文本起始符所对应的输出向量;

[0011] 利用所述优化后的BERT模型的结构,输入训练语料集TW进行深度预训练,得到深度预训练模型;并将所述深度预训练模型迁移到所述优化后的BERT模型softmax层中的微调模型中,得到情感分类微调模型;

[0012] 将待分类语料集TC输入所述情感分类微调模型中,得到情感分类结果;

- [0013] 将所述待分类语料集TC输入至BERT模型,进行向量化处理,并将向量化处理后的待分类语料集TC输入LDA主题模型中进行迭代计算,得到文档分布;
- [0014] 融合所述情感分类结果和所述文档分布,得到情感倾向分布;
- [0015] 按照时间顺序对所述情感倾向分布进行划分,得到情感时序仿真结果。
- [0016] 本发明还用于提供一种基于深度学习的网络舆情演化仿真系统,所述仿真系统包括:
- [0017] 输入模块,用于将训练语料集TB输入至BERT模型;
- [0018] 词向量优化模块,用于利用LDA主题模型对所述BERT模型中输入词所映射得到的词向量进行优化,得到优化后的词向量;
- [0019] 情感优化模块,用于将所述优化后的词向量输入至所述BERT模型的 Transformer编码器中,对所述BERT模型的预训练任务进行情感优化;
- [0020] 优化BERT模型获取模块,用于在情感优化后的BERT模型的softmax层中构建单层神经网络,作为情感分类器,得到优化后的BERT模型;所述单层神经网络连接所述Transformer编码器输出的文本起始符所对应的输出向量;
- [0021] 情感分类微调模型获取模块,用于利用所述优化后的BERT模型的结构,输入训练语料集TW进行深度预训练,得到深度预训练模型;并将所述深度预训练模型迁移到所述优化后的BERT模型softmax层中的微调模型中,得到情感分类微调模型;
- [0022] 情感分类结果获取模块,用于将待分类语料集TC输入所述情感分类微调模型中,得到情感分类结果;
- [0023] 文档分布获取模块,用于将所述待分类语料集TC输入至BERT模型,进行向量化处理,并将向量化处理后的待分类语料集TC输入LDA主题模型中进行迭代计算,得到文档分布;
- [0024] 情感倾向分布获取模块,用于融合所述情感分类结果和所述文档分布,得到情感倾向分布;
- [0025] 情感时序仿真结果获取模块,用于按照时间顺序对所述情感倾向分布进行划分,得到情感时序仿真结果。
- [0026] 根据本发明提供的具体实施例,本发明公开了以下技术效果:
- [0027] 本发明提供了一种基于深度学习的网络舆情演化仿真方法及系统,将训练语料集TB输入至BERT模型,并利用LDA主题模型对BERT模型中输入词所映射得到的词向量进行优化,得到优化后的词向量,既能够充分结合上下文语义信息弥补LDA词袋模型的不足,训练出更优的主题向量,又能将优化后的主题向量与BERT模型的词向量深度融合,实现主题视角下细粒度的舆情仿真分析。将优化后的词向量输入至BERT模型的Transformer编码器中,对BERT模型的预训练任务进行情感优化,使模型能够学习更多的情感信息。在情感优化后的BERT模型的softmax层中构建单层神经网络,作为情感分类器,得到优化后的BERT模型。利用优化后的BERT模型的结构,输入训练语料集TW进行深度预训练,得到深度预训练模型,并将深度预训练模型迁移到优化后的BERT模型softmax层中的微调模型中,得到情感分类微调模型,以在改进BERT模型预训练任务的基础上叠加深度预训练任务,实现更精准的舆情情感分类。
- [0028] 将待分类语料集TC输入情感分类微调模型中,得到情感分类结果,再将待分类语

料集TC输入至BERT模型,进行向量化处理,并将向量化处理后的待分类语料集TC输入LDA主题模型中进行迭代计算,得到文档分布,然后融合情感分类结果和文档分布,得到情感倾向分布,并按照时间顺序对情感倾向分布进行划分,得到情感时序仿真结果,从而得到更加细粒度和精准的文本主题聚类 and 舆情演化仿真结果。

附图说明

[0029] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0030] 图1为本发明实施例1所提供的仿真方法的方法流程图。

[0031] 图2为本发明实施例1所提供的改进后的BERT模型的结构示意图。

具体实施方式

[0032] 下面将结合本发明实施例中的附图,对本发明实施例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。

[0033] 本发明的目的是提供一种基于深度学习的网络舆情演化仿真方法及系统,对BERT模型的预训练任务进行改进,并在此基础上叠加深度预训练任务,还借助于与LDA主题模型深度融合以实现主题视角下细粒度的舆情仿真分析,从而实现更精准的舆情情感分类。

[0034] 为使本发明的上述目的、特征和优点能够更加明显易懂,下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0035] 实施例1:

[0036] 本实施例用于提供一种基于深度学习的网络舆情演化仿真方法,如图1 所示,所述仿真方法包括如下步骤:

[0037] 步骤101:将训练语料集TB输入至BERT模型;

[0038] 其中,在将训练语料集TB输入至BERT模型之前,所述仿真方法还包括对训练语料集TB进行分词、去停用词的数据预处理。还可统计训练语料集 TB中各个词的词频信息,提取高频词并添加特殊字符组成复杂语料词典,每个词对应一个唯一索引。

[0039] 步骤102:利用LDA主题模型对所述BERT模型中输入词所映射得到的词向量进行优化,得到优化后的词向量;

[0040] BERT(Bidirectional Encoder Representations from Transformers)是Google于2018年发布的基于双向Transformer结构的语言预训练模型,通过遮蔽语言模型(Masked Language Model,MLM)和句子预测(Next Sentence Prediction,NSP)任务来增强模型的语义表示能力,并依靠Transformer强大的特征提取能力和 Fine-tuning迁移学习能力使之在多项NLP任务中脱颖而出。为了提高大规模复杂文本在情感分类任务中的精准度和细粒度,本实施例保留了BERT基础模型,同时,将LDA主题模型的主题优化结果分别嵌入到BERT模型的预训练和微调阶段,使BERT模型在执行情感分类任务时能够同时学习句法、语

义和主题等文本特征。改进后的BERT模型的结构如图2所示。其中,设定 ω 、 δ 、 ρ 分别为BERT模型获取文本的词义向量(Word Embedding)、文本向量(Segment Embedding)和位置向量(Position Embedding),Trm表示Transformer 编码器,对于文本中插入的特殊字符,本实施例进行了如表1的解析。

[0041] 表1文本分析中的特殊字符映射关系

字符	含义
[CLS]	文本起始符
[SEP]	文本间隔符和终止符
[0042] [MASK]	遮盖字
[MUN]	数字
[DEL]	已删除的字符
[PAD]	填补空白符

[0043] 如图2所示,将分词后的文档输入模型后,每个词被映射成三个向量和表示 w_{ij} ($\omega + \delta + \rho$),将其统称为词向量。在改进过程中,为了使BERT模型能够实现主题视角下细粒度的舆情仿真分析,将词向量 w_{ij} ($\omega + \delta + \rho$)与LDA主题模型的主题表示相结合,但LDA(Latent Dirichlet Allocation)是常用的文档主题抽取优化模型,通过将文本集中每一篇文档主题以概率分布的形式给出,再根据主题分布进行主题聚类或文本分类优化。随后利用Gibbs Sampling算法进行参数估计,迭代抽样直到收敛,得到“主题-词”分布矩阵和“文本-主题”分布。在原始模型的计算过程中,主题向量 μ 由每个主题的高频词与文档的余弦距离计算得到,主题词受限于LDA词袋模型的传统结构,无法有效结合文本的语义和上下文信息,且分词技术的好坏对 μ 的计算结果影响较大。为了得到更优的主题向量,本实施例将BERT模型中的词向量与LDA模型的主题表示相融合,通过迭代计算词权重得到优化后的主题向量 u' ,以期为大规模复杂文本的情感仿真提供更为精准的主题语义信息。

[0044] 具体的,所述利用LDA主题模型对所述BERT模型中输入词所映射得到的词向量进行优化,得到优化后的词向量具体包括:将训练语料集TB输入至BERT模型,进行向量化处理,所述训练语料集TB中的每个词均被映射成一个包含词义向量、文本向量和位置向量的词向量;将向量化处理后的训练语料集TB输入至LDA主题模型,得到优化后的主题向量,将所述优化后的主题向量嵌入所述BERT模型,对所述BERT模型中输入词所映射得到的词向量进行优化,得到优化后的词向量;所述优化后的词向量包括词义向量、文本向量、位置向量和主题向量。进而得到融合更优主题向量的词向量 $w_{ij}(\omega + \delta + \rho + u')$,再将其传入双向Transformer编码器中。

[0045] 步骤103:将所述优化后的词向量输入至所述BERT模型的Transformer 编码器中,对所述BERT模型的预训练任务进行情感优化;

[0046] 舆情情感演化仿真模型的构建,旨在精准模拟大规模网络舆情的情感演化过程,掌握群体的情感演化特征与规律,辅助政府部门快速形成有效的舆情应对机制。由于BERT在预训练阶段缺少情感语料的训练,导致其执行情感分类任务时表现欠佳。因此,本实施例

为BERT模型制定了新的预训练任务,引入了改进的预训练语料集TB,训练语料集TB包括中文维基百科语料以及新浪微博和百度贴吧的情感语料,即在原有中文维基百科语料的基础上,加入新浪微博和百度贴吧情感语料,以期模型能够学习更多的情感信息,能够更好的执行情感分类任务。

[0047] 另外,为了使模型学习更多信息,Transformer编码器将多头机制 (Multi-Head Attention) 和前馈层 (Feed Forward) 通过残差网络结构连接,由多头机制对输入的向量进行多次线性变换得到不同的线性值,再对注意力权重进行计算。其计算过程如式 (1) 和式 (2) 所示:

$$[0048] \quad \text{MultiHead}(Q,K,V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h) W^0 \quad (1)$$

$$[0049] \quad \text{head}_f = \text{Attention}(QW_f^Q, KW_f^K, VW_f^V) \quad (2)$$

[0050] 式中,Q、K、V为输入词向量矩阵,将Q、K、V通过参数矩阵映射后计算Attention,重复h次后对计算结果进行拼接, $\text{head}_f = \{\text{head}_f | f \in \{1, 2, \dots, h\}\}$ 表示第f个超参数头, W^0 为权重矩阵, W_f^Q 、 W_f^K 、 W_f^V 表示第f个超参数头对应的 W^Q 、 W^K 、 W^V 权重矩阵。

[0051] 由此,Transformer编码器学习并存储了文档 d_i 的语义关系和语法结构信息,由于文档 d_i 经过BERT模型改进后融合了更优主题向量,因此,将改进后的文档 d'_i 通过特殊字符[CLS]与softmax的输出层连接以适配多任务下的迁移学习。由此,融合了主题优化信息的特征向量与BERT词向量相结合,可望提高模型在大规模复杂文本情感演化仿真的精确度和细粒度。

[0052] 步骤104:在情感优化后的BERT模型的softmax层中构建单层神经网络,作为情感分类器,得到优化后的BERT模型;所述单层神经网络连接所述 Transformer编码器输出的文本起始符[CLS]所对应的输出向量;

[0053] 所述情感分类器用于执行情感分类 (Sentiment Classification, SC) 任务,同时情感优化后的BERT模型仍然保留了MLM和NSP任务。所述BERT模型的softmax层中还包括一个连接在Transformer编码器输出的文本起始符所对应的输出向量的单层神经网络,以执行NSP任务。所述BERT模型的softmax层中还包括一个连接在Transformer编码器输出的遮盖字所对应的输出向量的单层神经网络,以执行MLM任务。

[0054] 步骤105:利用所述优化后的BERT模型的结构,输入训练语料集TW进行深度预训练,得到深度预训练模型;并将所述深度预训练模型迁移到所述优化后的BERT模型softmax层中的微调模型中,得到情感分类微调模型;

[0055] 本实施例在改进预训练任务的基础上,增加深度预训练任务,以实现更精准的舆情情感分类。引入公开的新浪微博标注集和少量特定事件的情感标注集 TW作为BERT深度预训练语料。

[0056] 具体的,所述利用所述优化后的BERT模型的结构,输入训练语料集TW进行深度预训练,得到深度预训练模型具体包括:

[0057] 设置初始学习率LR、批量大小、dropout率和Epoch训练次数;

[0058] 将所述MLM任务、所述NSP任务和所述SC任务的损失和作为深度预训练的损失函数;

[0059] 损失函数为:

[0060] $Loss = \lambda_1 Loss_{MLM} + \lambda_2 Loss_{NSP} + \lambda_3 Loss_{SC}$ (3)

[0061] 式3中, λ_1 、 λ_2 、 λ_3 分别为MLM任务、NSP任务和SC任务对应的权重分配。

[0062] 输入训练语料集TW进行深度预训练,在训练过程中,采用Adam优化算法确定最佳学习率LR,使用AUC作为检验模型训练效果的指标,并采用F1 分数确定正、负情感分类的最佳阈值,以在训练过程中避免过拟合现象,得到深度预训练模型。

[0063] 步骤106:将待分类语料集TC输入所述情感分类微调模型中,得到情感分类结果;

[0064] 其中,TC为特定事件的情感分类任务语料集。输入待分类语料集TC并计算每个文档 TC_i 的情感值 $E(TC_i)$,根据情感分类阈值将情感值转化为相应的情感极性 $E'(TC_i)$,得到粗粒度情感分类结果RTC; $RTC = (MnE_{pos}(TC_i), MnE_{neg}(TC_i), E'_{pos}(TC_i), E'_{neg}(TC_i))$;其中, $MnE_{pos}(TC_i)$ 为正向情感均值, $MnE_{neg}(TC_i)$ 为负向情感均值, $E'_{pos}(TC_i)$ 为正向情感文本数, $E'_{neg}(TC_i)$ 为负向情感文本数。

[0065] 然后,将RTC按照时间顺序整合到时间序列文本集 TC_time , $TC_time = \{(TC_1, time_1), (TC_2, time_2), \dots, (TC_M, time_q)\}$;其中, $time$ 含 q 个时间片,得到粗粒度的“文本_时间”情感时序演化结果 $time_RTC$; $time_RTC = \{time_RTC_1, time_RTC_2, \dots, time_RTC_q\}$ 。

[0066] 步骤107:将所述待分类语料集TC输入至BERT模型,进行向量化处理,并将向量化处理后的待分类语料集TC输入LDA主题模型中进行迭代计算,得到文档分布;

[0067] 将待分类语料集TC输入至BERT模型,进行向量化处理,所述待分类语料集TC中的每个词均被映射成一个包含词义向量、文本向量和位置向量的词向量;将向量化处理后的待分类语料集TC输入至LDA主题模型,得到1个最优主题数下“主题-词”对应的文档分布TTC, $TTC = \{(TC_1, l_1), (TC_2, l_2), \dots, (TC_M, l_1)\}$ 。

[0068] 步骤108:融合所述情感分类结果和所述文档分布,得到情感倾向分布;

[0069] 将文档分布TTC与粗粒度情感分类结果RTC相结合,得到“主题-词”的情感倾向分布TTC'。

[0070] 步骤109:按照时间顺序对所述情感倾向分布进行划分,得到情感时序仿真结果。

[0071] 将情感倾向分布TTC'整合到时间序列文本集 TC_time 中,得到细粒度的“文本-主题”情感时序仿真结果 $time_TTC'$; $time_TTC' = \{time_TTC'_1, time_TTC'_2, \dots, time_TTC'_q\}$ 。

[0072] 本实施例提供的网络舆情情感仿真方法融合LDA主题模型和BERT模型,既能够充分结合上下文语义信息、弥补LDA词袋模型的不足,训练出更优的主题向量,又能将优化后的主题向量与BERT词向量深度融合,得到更加细粒度和精准的文本主题聚类与舆情演化仿真结果。

[0073] 实施例2:

[0074] 本实施例用于提供一种基于深度学习的网络舆情演化仿真系统,采用如实施例1所述的仿真方法进行工作,所述仿真系统包括:

[0075] 输入模块,用于将训练语料集TB输入至BERT模型;

[0076] 词向量优化模块,用于利用LDA主题模型对所述BERT模型中输入词所映射得到的词向量进行优化,得到优化后的词向量;

[0077] 情感优化模块,用于将所述优化后的词向量输入至所述BERT模型的Transformer编码器中,对所述BERT模型的预训练任务进行情感优化;

[0078] 优化BERT模型获取模块,用于在情感优化后的BERT模型的softmax层中构建单层神经网络,作为情感分类器,得到优化后的BERT模型;所述单层神经网络连接所述Transformer编码器输出的文本起始符所对应的输出向量;

[0079] 情感分类微调模型获取模块,用于利用所述优化后的BERT模型的结构,输入训练语料集TW进行深度预训练,得到深度预训练模型;并将所述深度预训练模型迁移到所述优化后的BERT模型softmax层中的微调模型中,得到情感分类微调模型;

[0080] 情感分类结果获取模块,用于将待分类语料集TC输入所述情感分类微调模型中,得到情感分类结果;

[0081] 文档分布获取模块,用于将所述待分类语料集TC输入至BERT模型,进行向量化处理,并将向量化处理后的待分类语料集TC输入LDA主题模型中进行迭代计算,得到文档分布;

[0082] 情感倾向分布获取模块,用于融合所述情感分类结果和所述文档分布,得到情感倾向分布;

[0083] 情感时序仿真结果获取模块,用于按照时间顺序对所述情感倾向分布进行划分,得到情感时序仿真结果。

[0084] 本说明书中各个实施例采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似部分互相参见即可。对于实施例公开的系统而言,由于其与实施例公开的方法相对应,所以描述的比较简单,相关之处参见方法部分说明即可。

[0085] 本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处。综上所述,本说明书内容不应理解为对本发明的限制。

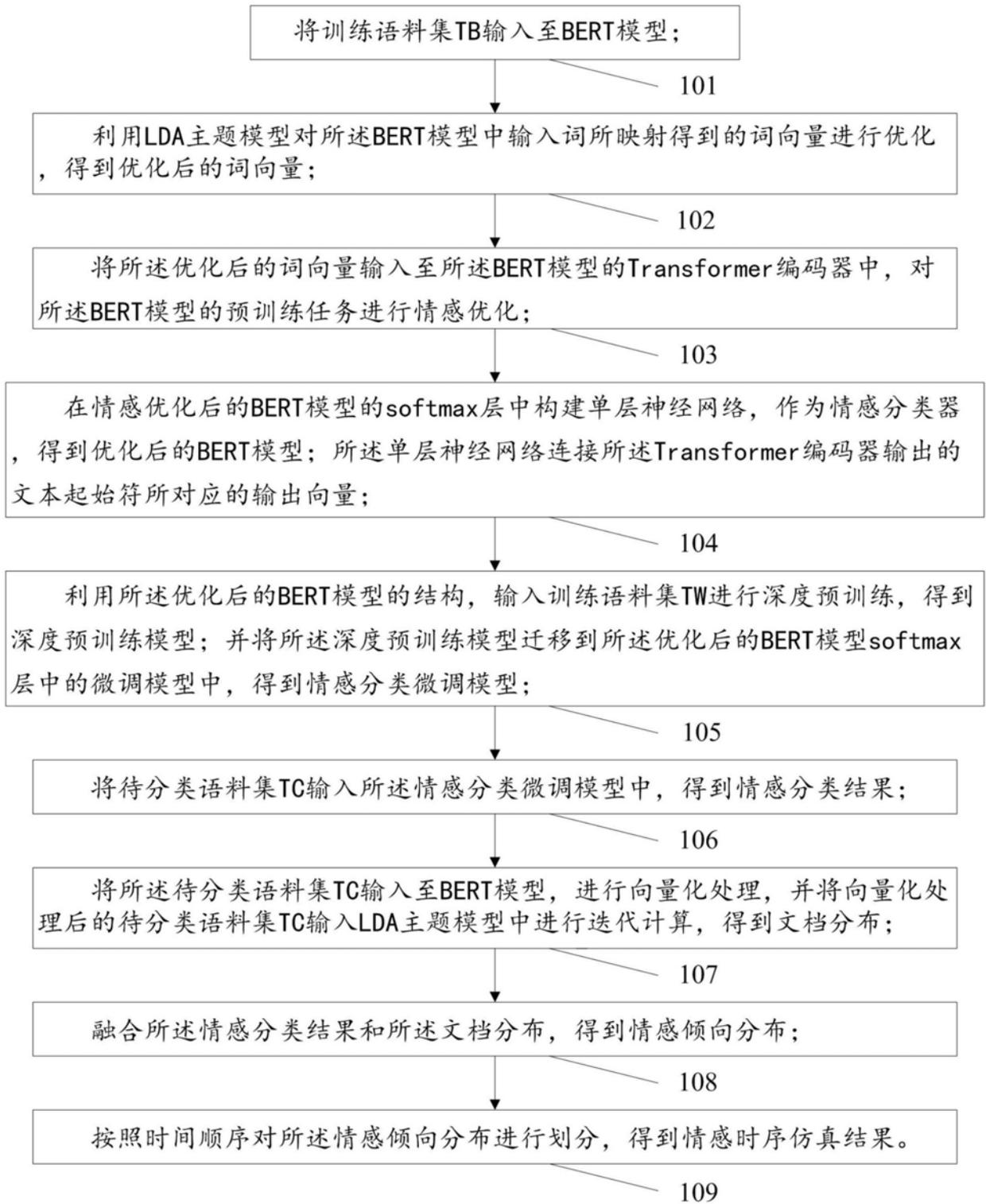


图1

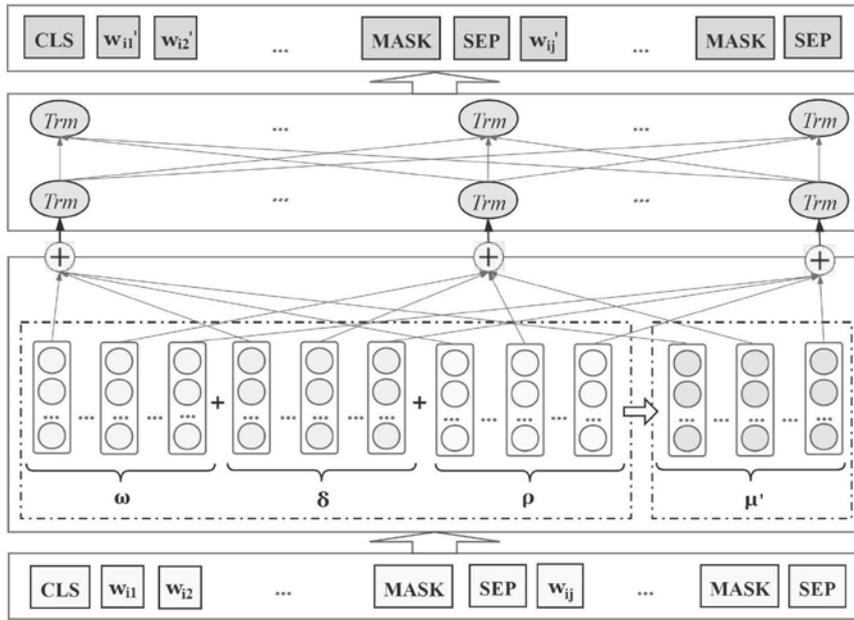


图2