



# (12)发明专利申请

(10)申请公布号 CN 106910013 A

(43)申请公布日 2017.06.30

(21)申请号 201710085225.5

(22)申请日 2017.02.16

(71)申请人 中国科学院自动化研究所

地址 100190 北京市海淀区中关村东路95号

(72)发明人 谭铁牛 王亮 吴书 刘强 余峰

(74)专利代理机构 中科专利商标代理有限责任公司 11021

代理人 钟文芳

(51)Int.Cl.

G06Q 10/06(2012.01)

G06Q 50/00(2012.01)

H04L 29/06(2006.01)

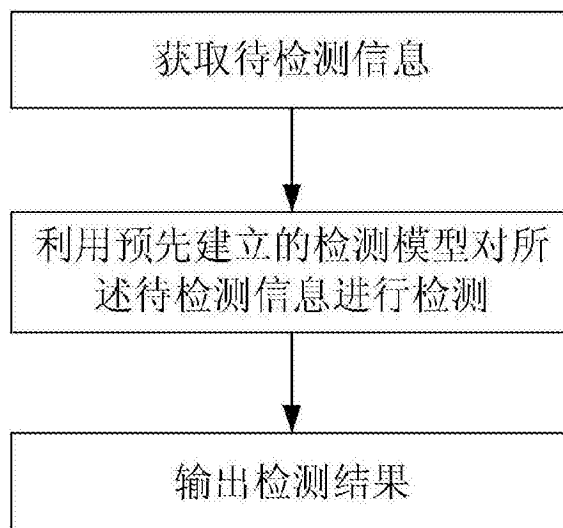
权利要求书2页 说明书6页 附图3页

## (54)发明名称

基于动态表达学习的不实信息检测方法和装置

## (57)摘要

本发明公开一种基于动态表达学习的不实信息检测方法,包括以下步骤:获取待检测信息;利用预先建立的检测模型对所述待检测信息进行检测;输出检测结果;其中,检测模型如下建立:步骤S1,首先建模联合表示用户信息和该用户行为信息的某一事件的动态行为表达式;用户信息包含用户的特征和用户可信度,行为信息包含行为类型;步骤S2,一个事件由不同信息组成,结合步骤S1中的所述动态行为表达式,最终得出事件可信度检测表达式;步骤S3,运用时间特征矩阵来取得在信息传播过程中用户动态行为特征连步骤S4,生成用户特征表达;步骤S5,利用配对学习法估算检测模型参数。



1. 一种基于动态表达学习的不实信息检测方法,包括以下步骤:

获取待检测信息;

利用预先建立的检测模型对所述待检测信息进行检测;

输出检测结果;

其中,检测模型如下建立:

步骤S1,首先建模联合表示用户信息和该用户行为信息的某一事件的动态行为表达式;用户信息包含用户的特征和用户可信度,行为信息包含行为类型;

步骤S2,一个事件由不同信息组成,结合步骤S1中的所述动态行为表达式,最终得出事件可信度检测表达式;

步骤S3,运用时间特征矩阵来取得在信息传播过程中用户动态行为特征连

步骤S4,生成用户特征表达;

步骤S5,利用配对学习法估算检测模型参数。

2. 根据权利要求1所述的方法,其特征在于,所述信息为微博信息;所述用户信息包括用户的特征和用户可信度,所述用户的特征包括用户性别、微博关注人数和被关注的人数;用户可信度的数值越大,表示用户越可信;行为信息包括原发微博或转发微博。

3. 根据权利要求1所述的方法,其特征在于,所述动态行为表达式如下表示:

$$R_j^{e_i} = T_{t_j}^{e_i} C_{c_j}^{e_i} B_{b_j}^{e_i} U_{u_j}^{e_i},$$

其中, $U_{u_j}^{e_i} \in R^d$ 代表第i个事件 $e_i$ 中的第j条微博中的用户 $u_j^{e_i}$ 的向量表示; $R^d$ 表示d维实数空间; $B_{b_j}^{e_i} \in R^{d \times d}$ 是用户行为 $b_j^{e_i}$ 的隐含矩阵表示; $C_{c_j}^{e_i} \in R^{d \times d}$ 是评论 $c_j^{e_i}$ 的矩阵表达, $T_{t_j}^{e_i} \in R^{d \times d}$ 是时间间隔 $t_j^{e_i}$ 的矩阵表达。

4. 根据权利要求1所述的方法,其特征在于,事件可信度检测表达式如下所示:

$$y^{e_i} = W^T R^{e_i}$$

其中, $y^{e_i}$ 表示事件 $e_i$ 的可信度, $W \in R^d$ 是预测函数的线性权重, $R^{e_i}$ 表示事件 $e_i$ 的表达式。

5. 根据权利要求1所述的方法,其特征在于,在步骤S3中,将连续时间段分割成不同的时间间隔,根据 $\log_2$ 来划分时间间隔,并且只学习上边界和下边界相对应的时间间隔,而对于在一个时间间隔的某一个时刻,其时间特征矩阵(一种转移矩阵)通过非线性插值法计算得出。

6. 根据权利要求1所述的方法,其特征在于,步骤S5中通过如下表达式区别真实信息和不实信息的差别:

$$p(e_n \succ e_r) = g(y^{e_n} - y^{e_r})$$

其中 $y^{e_n}$ ,  $y^{e_r}$ 分别表示真实信息 $e_n$ 和不实信息 $e_r$ 的可信度, $g(x)$ 是非线性方程, $g(x) = 1/(1+e^{-x})$ ;

检测模型的目标函数如下表示:

$$J = \sum_{\{e_n, e_r\} \in E, l_{e_n} = 1, l_{e_r} = 0} \ln(1 + e^{-W^T (R^{e_n} - R^{e_r})}) + \frac{\lambda}{2} \|\theta\|^2,$$

其中, $E$ 表示所有事件的集合, $l_{en}$ 、 $l_{er}$ 分别表示真实信息和不实信息的标签, $\Theta$ 代表所有被计算的检测模型参数, $R^{en}$ 和 $R^{er}$ 分别表示真实信息 $e_n$ 和不实信息 $e_r$ 的动态行为表达, $W \in R^d$ 是预测函数的线性权重; $\lambda$ 是控制正则化大小的参数。

7. 根据权利要求1所述的方法,其特征在于,检测模型概括了表征信息关键特性的多种特征,即用户信息、行为信息、时间信息和评论信息,并且建模了这些特征间的高阶交互表达。

8. 根据权利要求1所述的方法,其特征在于,检测模型揭示了信息数量随时间的幂律分布规律,并且依据此规律采用 $\log_2$ 将连续时间段分割成不同的时间间隔,不仅保证每一个时间间隔内有相同数目的信息数目,而且能够从整体上保证所有事件共有一个相似的时间尺度。

9. 一种基于动态表达学习的不实信息检测装置,包括:

获取模块,被配置为获取待检测信息;

检测模块,被配置为利用预先建立的检测模型对所述待检测信息进行检测;

输出模块,被配置为输出检测结果;

其中,检测模型如下建立:

首先建模联合表示用户信息和该用户行为信息的某一事件的动态行为表达式;用户信息包含用户的特征和用户可信度,行为信息包含行为类型;

一个事件由不同信息组成,结合步骤S1中的所述动态行为表达式,最终得出事件可信度检测表达式;

运用时间特征矩阵来取得在信息传播过程中用户动态行为特征连

生成用户特征表达;

利用配对学习法估算检测模型参数。

## 基于动态表达学习的不实信息检测方法和装置

### 技术领域

[0001] 本发明涉及计算机模型检测技术领域,尤其涉及一种基于动态表达学习的不实信息检测方法和装置。

### 背景技术

[0002] 社交媒体的飞速发展使网络用户体验到了前所未有的便捷。社交媒体例如 Facebook, Twitter 和新浪微博为用户提供了可以共享信息和公开发表个人能言论的平台。但与此同时,社交平台上不实信息的传播也为用户带来了极大的困扰,也危害了社会和谐和公共安全。近年来,信息可信度检测引起了学术界和工业界极大的关注。

[0003] 现今方法中考虑到的信息主要包括以下几个种类:正文信息、来源可信信息、动态信息和评论信息;评价用户行为可信度因素主要包括:时间、人物、行为、方式。事实发现法是一种无监督或半监督方法,用于在冲突数据中发现事实和检测信息可信度。基于上述信息,已有方法主要关注于人工特征,但其较为繁琐并且无法得出数据的基本特征。同时,现今方法无法模拟信息传播时不同信息种类和不同可信性因素的关联性。事实发现法主要基于来源可信信息且检测到的可信度集合于各个来源。但事实发现法仅适合于在理想情况特定的主题中,例如价格预测和航班预测,并不适用于社交媒体这种复杂环境之中。

[0004] 近年来,很多在社交媒体能自动测量信息可信度的方法得到广泛的应用。这些方法主要基于消息级别或事件级别的正文信息和来源可信信息。还有研究综合考虑了消息级别和事件级别。针对于动态信息,一些研究定义了传播过程中的时间特征或者训练具有不同时间段特征的模型。针对于评论信息,研究采取了用户反馈或标志微博方式表明可疑信息。尽管这些方法被广泛使用,但因建立在特征工程上,就显得较为繁琐并且无法得出数据的基本特征。同时,现今方法无法模拟信息传播时不同信息种类和不同可信性因素的关联性。

[0005] DBRM模型致力于根据社交媒体上用户发布和转发的内容来判断该事件是否为不实信息。模型集合用户行为因素:用户可信度、事件发生时间间隔、用户发布和转发行为和用户评论信息,用来判断一条消息的可信度。模型引入了表征学习方法(representation learning method),与传统特征工程(feature engineering)不同的是它能够抓取在传播过程中不同方面的信息。模型学习了用户,动态时间间隔,用户行为和评论态度的隐含表示(latent representation)。基于这些隐含表示,模型可生成信息的动态行为表示并在可信度检测方面提出了创新。

### 发明内容

[0006] 鉴于传统基于人工特征的方法存在技术缺陷,为了更好检测信息可信度,本发明提供一种基于动态行为特征表示的检测方法和装置。

[0007] 根据本发明一方面,提供了一种基于动态表达学习的不实信息检测方法,包括以下步骤:

- [0008] 获取待检测信息；
- [0009] 利用预先建立的检测模型对所述待检测信息进行检测；
- [0010] 输出检测结果；
- [0011] 其中,检测模型如下建立：
- [0012] 步骤S1,首先建模联合表示用户信息和该用户行为信息的某一事件的动态行为表达式；用户信息包含用户的特征和用户可信度,行为信息包含行为类型；
- [0013] 步骤S2,一个事件由不同信息组成,结合步骤S1中的所述动态行为表达式,最终得出事件可信度检测表达式；
- [0014] 步骤S3,运用时间特征矩阵来取得在信息传播过程中用户动态行为特征连
- [0015] 步骤S4,生成用户特征表达；
- [0016] 步骤S5,利用配对学习法估算检测模型参数。
- [0017] 根据本发明第二方面,提供了一种基于动态表达学习的不实信息检测装置,包括：
- [0018] 获取模块,被配置为获取待检测信息；
- [0019] 检测模块,被配置为利用预先建立的检测模型对所述待检测信息进行检测；
- [0020] 输出模块,被配置为输出检测结果；
- [0021] 其中,检测模型如下建立：
- [0022] 首先建模联合表示用户信息和该用户行为信息的某一事件的动态行为表达式；用户信息包含用户的特征和用户可信度,行为信息包含行为类型；
- [0023] 一个事件由不同信息组成,结合步骤S1中的所述动态行为表达式,最终得出事件可信度检测表达式；
- [0024] 运用时间特征矩阵来取得在信息传播过程中用户动态行为特征连
- [0025] 生成用户特征表达；
- [0026] 利用配对学习法估算检测模型参数。
- [0027] 本发明所采用的检测模型概括了表征信息关键特性的多种特征,即用户信息、行为信息、时间信息和评论信息,并且建模了这些特征间的高阶交互表达。这样建模的微博或者事件的表达能够学习到更加完整,充分,真实的向量表示,更能够适用于复杂多变的社交网络场合。检测模型揭示了信息数量随时间的幂律分布规律,并且依据此规律采用 $\log_2$ 将连续时间段分割成不同的时间间隔,不仅保证每一个时间间隔内有相同数目的信息数目,而且能够从整体上保证所有事件共有一个相似的时间尺度。模型能够更加容易地学会这些事件的表达,而且能够充分挖掘出信息分布的时间规律。本发明涉及基于动态表达学习的不实信息检测发法任务,特别针对信息数量大,时间跨度长,语义场景复杂,用户行为变化等真实复杂的社交网络场合。学习用户的动态行为表达来获得更准确的预测效果。

## 附图说明

- [0028] 图1是本发明中基于动态表达学习的不实信息检测方法的流程图；
- [0029] 图2是本发明中动态行为表达模型DBRM的表达学习过程的示意图；
- [0030] 图3(a)和图3(b)是不同的对比方法对于谣言(a)和真实信息(b)的精确度-召回率曲线。

## 具体实施方式

[0031] 以下结合附图详细说明本发明技术方案中所涉及各个细节问题。应当指出的是,所描述的实施例仅旨在便于理解,对本发明不起任何限定作用。

[0032] 如图1所示,本发明提供了一种基于动态表达学习的不实信息检测方法,包括以下步骤:

[0033] 获取待检测信息;

[0034] 利用预先建立的检测模型对所述待检测信息进行检测;

[0035] 输出检测结果;

[0036] 其中,检测模型如下建立:

[0037] 步骤S1,首先建模联合表示用户信息和该用户行为信息的某一事件的动态行为表达式;用户信息包含用户的特征和用户可信度,行为信息包含行为类型;

[0038] 步骤S2,一个事件由不同信息组成,结合步骤S1中的所述动态行为表达式,最终得出事件可信度检测表达式;

[0039] 步骤S3,运用时间特征矩阵来取得在信息传播过程中用户动态行为特征连

[0040] 步骤S4,生成用户特征表达;

[0041] 步骤S5,利用配对学习法估算检测模型参数。

[0042] 本发明提出的动态行为表达模型(Dynamic Behavioral Representation Model,简称DBRM),用在社交媒体场景下的不实信息检测。模型可以学习动态行为表达,通过学习隐含表示可以建立含有用户可信度、动态属性、行为特征和评价观点的模型。上述方面信息的集合将生成用户行为表示,这些用户动态行为表示的集合生成描述一个事件在社交媒体上传播信息的可信度表示。在模型中,每个用户都用对应的向量表示,其中时间间隔、用户行为和用户评论分别用矩阵表示。模型更引入了成对学习法以便于最大化准确信息与不实信息的可信度差异。DBRM模型建立:1) 每一个用户由带有自身特征(如性别,关注和被关注的人数)的向量表示来表明用户信息可信度;2) 模型结合了从不实信息开始传播到发微博时间间隔的矩阵表示,以便于抓取用户行为的动态特征。用隐含操作矩阵表示用户行为(如发布和转发)可以表明不同行为特征和用户评论是否为质疑态度;3) 基于2)的表达式乘积生成了在传播过程中信息的表达;4) 在结合3)中所有动态行为表达模型之后,我们可得出事件的可信度表达式;5) 我们运用了成对学习方法来最大化准确信息与不实信息的差异以此来检测社交媒体上信息的可信度。在新浪微博数据集实验上,获得比其他现有模型的预测要更准确的效果。

[0043] 为了更好地理解DBRM模型在不实信息检测中作用,以及验证本发明的实施效果,接下来以实验为例进行说明,本示例采用新浪微博数据库。实验数据集分为60%训练集,30%测试集和10%验证集。

[0044] 实验包含四个评价指标:准确率,精确度,召回率和F1值。研究对象分别为不实信息和真实信息时分别计算了精确度和召回率来显示模型检测两种信息的能力。四种评价指标的值越大,模型的性能就越强。

[0045] 在微博数据集上具体实验步骤如下:

[0046] 步骤S1,首先建模传统用户信息和行为信息。传统用户信息包括用户的特征和用

户可信度,用户的特征例如包括用户性别、微博关注人数和被关注的人数;用户可信度的数值越大,用户越可信;行为信息包括行为类型,例如微博为原发还是转发,相对于转发,原发行为更具有原创性,对于可信度检测也更重要。具有很高的可信度的微博往往都是有高可信度用户原发,而一些不实信息往往由低可信度用户原发,高可信度用户转发。

[0047] 对于第*i*个事件 $e_i$ 的涉及的第*j*条微博 $m_j^{e_i}$ ,可以联合表示出用户和其行为的表达式 $R_j^{e_i} = B_{b_j^{e_i}} U_{u_j^{e_i}}$ ,  $U_{u_j^{e_i}} \in R^d$ 代表第*i*个事件中的第*j*条微博中的用户 $u_j^{e_i}$ 的向量表示。 $R^d$ 表示*d*维实数空间。 $B_{b_j^{e_i}} \in R^{d \times d}$ 是用户行为 $b_j^{e_i}$ 的隐含矩阵表示,其中的各个元素在训练的过程中不断学习更新,*d*代表矩阵维度。这些表达可以得出用户在特定行为下的特征。

[0048] 除此之外,用户的评论也对信息可信度检测起着至关重要的作用。用户可以根据生活常识与经验对信息做出评价。根据下述表达式,不实信息往往收到更多的质疑评论。结合微博 $m_j^{e_i}$ 的所有评论 $c_j^{e_i}$ 可以得出: $R_j^{e_i} = C_{c_j^{e_i}} B_{b_j^{e_i}} U_{u_j^{e_i}}$ ,其中 $C_{c_j^{e_i}} \in R^{d \times d}$ 是评论 $c_j^{e_i}$ 的矩阵表达,表达式中加入了作者评论态度。

[0049] 相应地,自谣言开始传播到发特定微博的时间间隔和用户行为的联合可以更好的检测可信度。将事件 $e_i$ 从开始到相应微博开始传播的时间间隔 $t_j^{e_i}$ 加入表达式中,可以得到微博 $m_j^{e_i}$ 的动态行为表达: $R_j^{e_i} = T_{t_j^{e_i}} C_{c_j^{e_i}} B_{b_j^{e_i}} U_{u_j^{e_i}}$ ,其中 $T_{t_j^{e_i}} \in R^{d \times d}$ 是时间间隔 $t_j^{e_i}$ 的矩阵表达。 $R_j^{e_i}$ 可用来表示四种不同因素对微博 $m_j^{e_i}$ 的联合影响。

[0050] 步骤S2,一个事件由不同微博组成,结合S1中的微博动态行为表达式,才能最终得出事件可信度检测表达式。设事件 $e_i$ 包含 $n_{e_i}$ 个微博,所有微博构成集合 $M^{e_i}$ ,根据平均值计算,得出事件 $e_i$ 的表达式为: $R^{e_i} = \frac{1}{n_{e_i}} \sum_{m_j^{e_i} \in M^{e_i}} R_j^{e_i}$ 。预测事件 $e_i$ 是否为不实信息,可采用表达式:

$y^{e_i} = W^T R^{e_i}$ ,其中 $W \in R^d$ 是预测函数的线性权重。其中 $y^{e_i}$ 表示事件 $e_i$ 的可信度, $y^{e_i}$ 的值越大,事件 $e_i$ 的可信度越高。

[0051] 步骤S3,在模型中,我们运用时间特征矩阵来取得在信息传播过程中用户动态行为特征。为了降低因为在连续时间段内学习不同矩阵而导致的数据稀疏的问题,我们将连续时间段分割成不同的时间间隔。根据动态行为的幂律分布图,将时间等间隔分割并不合理。本模型根据 $\log_2$ (以2为底的对数)来划分时间间隔,只学习矩阵上边界和下边界相对应的时间间隔。对于在一个时间间隔的某一个时刻,他们的转移矩阵可以非线性插值法计算得出。对于某一时刻*t*的时间特征矩阵 $T_t$ ,

$$[0052] \quad T_t = \frac{(U(\log_2^t) - \log_2^t) T_{\frac{1}{2} \log_2^t} + (\log_2^t - L(\log_2^t)) T_{\frac{3}{2} \log_2^t}}{U(\log_2^t) - L(\log_2^t)}$$

[0053] 其中 $U(\log_2^t)$ 和 $L(\log_2^t)$ 分别代表 $\log_2^t$ 的上边界和下边界。

[0054] 步骤S4,用户表示生成。对于模型中的用户表示,我们可以通过学习不同潜在向量表达来取得用户的特征和可信度。平均每个用户只有两种行为,我们无法学习每个用户的潜在表达。但是我们可以学习用户的特征表达。

[0055] 用户的特征例如可以包括性别、微博关注人数、微博被关注人数、微博数量以及用户是否被认证等信息。对于用户*u*,特征向量 $F_u \in R^f$ ,  $F_u = [F_u^{gender}, F_u^{followers}, F_u^{followees}, F_u^{microblog}, F_u^{verified}]^T$ 。

其中,  $F_u^{gender}$  和  $F_u^{verified}$  有两位信息,  $F_u^{gender}(1)=1$  (即两位信息中的第1位为1) 表示性别为男,  $F_u^{gender}(2)=1$  表示性别为女;  $F_u^{verified}(1)=1$  表示用户已被认证,  $F_u^{verified}(2)=1$  表示用户未被认证。用户的关注人数, 被关注人数和微博数量不容易得到每个数值的表达, 我们将人数和微博数量等数值根据  $\log_{10}$  分布分割成离散的时间区间。如果用户  $u$  有  $v_u$  个关注者, 我们可以得出相对应的特征,

$$[0056] \quad F_u^{followers}(i) = \begin{cases} U(\log_{10}^{v_u}) - \log_{10}^{v_u}, i = L(\log_{10}^{v_u}) + 1 \\ \log_{10}^{v_u} - L(\log_{10}^{v_u}), i = U(\log_{10}^{v_u}) + 1 \\ 0, i = others \end{cases}$$

[0057] 其中,  $U(\log_{10}^{v_u})$  和  $L(\log_{10}^{v_u})$  分别代表了  $\log_{10}^{v_u}$  的上边界和下边界,  $i$  表示区间的边界。同理  $F_u^{followers}$  和  $F_u^{microblog}$  表达式可以用上述方法构建出。基于特征向量  $F_u$ , 我们可以得出用户表达式  $U_u = SF_u$ , 其中  $S \in \mathbb{R}^{d \times f}$  是特征表达隐含矩阵, 在训练过程中不断学习。

[0058] 步骤S5, 配对学习法用来估算模型参数。考虑到不实信息不容易收集到来训练模型, 我们用成对学习法来扩大训练集。假设, 准确信息的可信度高于不实信息, 我们用如下式子最大化二者的差别:

$$[0059] \quad p(e_n > e_r) = g(y^{e_n} - y^{e_r})$$

[0060] 其中  $y^{e_n}$ ,  $y^{e_r}$  分别表示真实信息  $e_n$  和不实信息  $e_r$  的可信度,  $g(x)$  是非线性方程,  $g(x) = 1/(1+e^{-x})$ 。结合负数对数似然函数, 我们可以写出目标函数:

$$[0061] \quad J = \sum_{\{e_n, e_r\} \in E, l_{e_n}=1, l_{e_r}=0} \ln(1 + e^{-W^T(R^{e_n} - R^{e_r})}) + \frac{\lambda}{2} \|\Theta\|^2,$$

[0062] 其中  $E$  表示所有事件的集合,  $l_{e_n}$ ,  $l_{e_r}$  分别表示真实信息和不实信息的标签,  $\Theta = \{U, B, C, T, W\}$  代表所有被计算的参数,  $\lambda$  是控制正则化大小的参数。可以得出  $J$  关于  $W$ ,  $R^{e_n}$  和  $R^{e_r}$  的倒数如下:

$$[0063] \quad \frac{\partial J}{\partial W} = \sum_{e_n, e_r \in E, l_{e_n}=1, l_{e_r}=0} \frac{(R^{e_r} - R^{e_n})l(e_n, e_r)}{1 + l(e_n, e_r)} + \lambda W$$

$$[0064] \quad \frac{\partial J}{\partial R^{e_n}} = - \sum_{e_n \in E, l_{e_n}=1} \frac{Wl(e_n, e_r)}{1 + l(e_n, e_r)}$$

$$[0065] \quad \frac{\partial J}{\partial R^{e_r}} = \sum_{e_n \in E, l_{e_n}=1} \frac{Wl(e_n, e_r)}{1 + l(e_n, e_r)}$$

$$[0066] \quad \text{其中, } l(e_n, e_r) = e^{-W^T(R^{e_n} - R^{e_r})}.$$

[0067] 计算出事件  $R^{e_r}$  的倒数  $\partial J / \partial R^{e_r}$ , 相对应参数的梯度可表示为:

$$[0068] \quad \frac{\partial J}{\partial \mathbf{T}_j^{e_r}} = \frac{1}{n_{e_r}} \frac{\partial J}{\partial R^{e_r}} (\mathbf{C}_j^{e_r} \mathbf{B}_j^{e_r} \mathbf{U}_j^{e_r})^T$$

$$[0069] \quad \frac{\partial J}{\partial \mathbf{C}_j^{e_r}} = \frac{1}{n_{e_r}} (\mathbf{T}_j^{e_r})^T \frac{\partial J}{\partial R^{e_r}} (\mathbf{B}_j^{e_r} \mathbf{U}_j^{e_r})^T$$

$$[0070] \quad \frac{\partial J}{\partial \mathbf{B}_j^{e_r}} = \frac{1}{n_{e_r}} (\mathbf{T}_j^{e_r} \mathbf{C}_j^{e_r})^T \frac{\partial J}{\partial R^{e_r}} (\mathbf{U}_j^{e_r})^T$$

[0071] 
$$\frac{\partial J}{\partial \mathbf{U}_j^q} = \frac{1}{n_{e_j}} (\mathbf{T}_j^q \mathbf{C}_j^q \mathbf{B}_j^q)^T \frac{\partial J}{\partial \mathbf{R}^q}$$

[0072] 进而推出，

[0073] 
$$\frac{\partial J}{\partial \mathbf{T}_t} = \sum_{e_j \in E} \sum_{m_j^q \in M^q, l_j^q = t} \frac{\partial J}{\partial \mathbf{T}_j^q} + \lambda \mathbf{T}_t$$

[0074] 
$$\frac{\partial J}{\partial \mathbf{C}_c} = \sum_{e_j \in E} \sum_{m_j^q \in M^q, c_j^q = c} \frac{\partial J}{\partial \mathbf{C}_j^q} + \lambda \mathbf{C}_c$$

[0075] 
$$\frac{\partial J}{\partial \mathbf{B}_b} = \sum_{e_j \in E} \sum_{m_j^q \in M^q, b_j^q = b} \frac{\partial J}{\partial \mathbf{B}_j^q} + \lambda \mathbf{B}_b$$

[0076] 
$$\frac{\partial J}{\partial \mathbf{U}_u} = \sum_{e_j \in E} \sum_{m_j^q \in M^q, u_j^q = u} \frac{\partial J}{\partial \mathbf{U}_j^q} + \lambda \mathbf{U}_u$$

[0077] 所有梯度计算出之后，我们可以运用随机梯度下降计算出模型参数。重复以上过程直到模型收敛。

[0078] 如图3 (a) 和图3 (b) 分别表示各种方法在谣言数据集和非谣言数据集上面的PR曲线；

[0079] 下表1为数据集的统计信息；

[0080] 表1

[0081]

元素	事件	谣言	真实信息	微博	原发微博	转发微博	用户
数量	936	500	436	630363	98429	532236	321246

[0082] 下表2为本模型与当下最先进的模型的实验对比结果：

[0083] 表2

[0084]

方法	准确率	谣言			真实信息		
		精确度	召回率	F1 值	精确度	召回率	F1 值
NewsCP [1]	0.756	0.741	0.808	0.773	0.728	0.770	0.749
EP [2]	0.812	0.795	0.899	0.844	0.802	0.793	0.798
DBRM	0.887	0.831	0.990	0.903	0.946	0.805	0.870

[0085] 以上所述的具体实施示例，对本发明的目的、技术方案和效果进行了进一步详细说明。所应理解的是，以上所述仅为本发明的具体示例而已，并不用于限制本发明，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

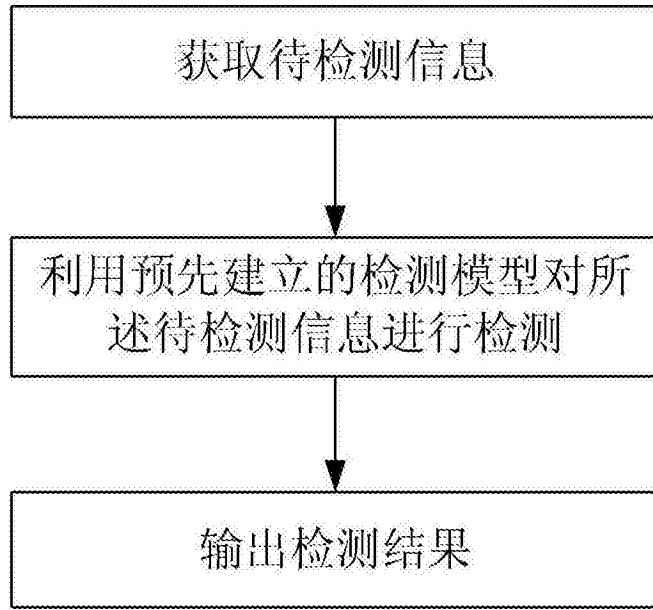


图1

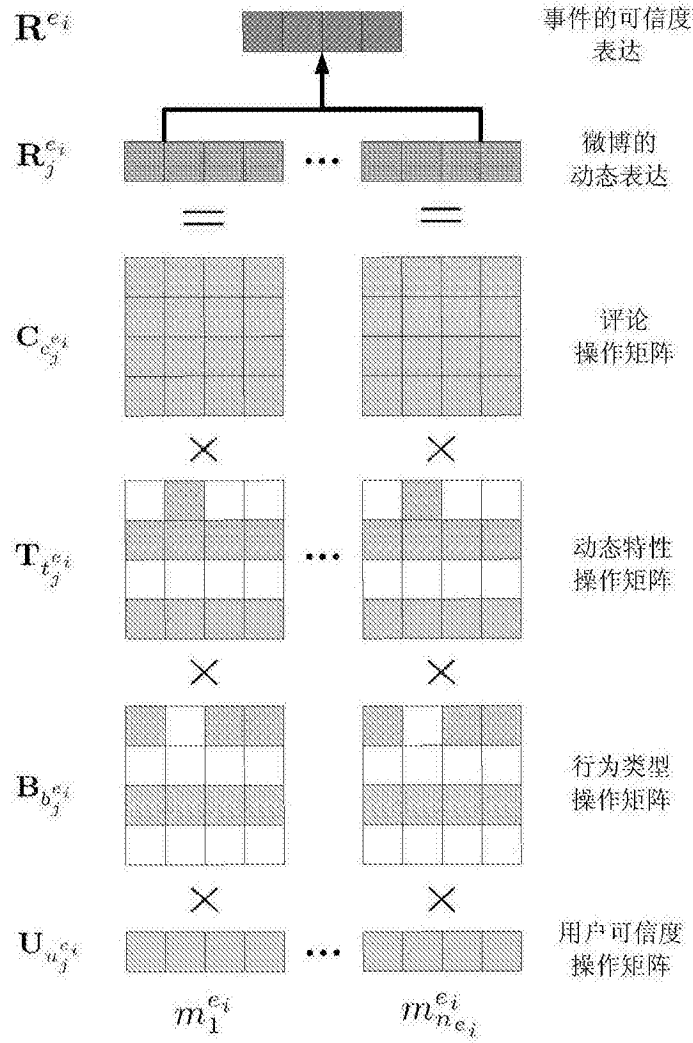


图2

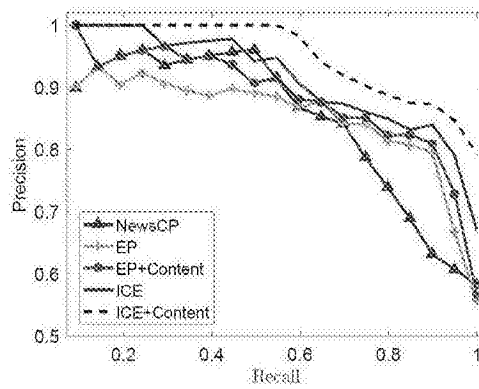


图3 (a)

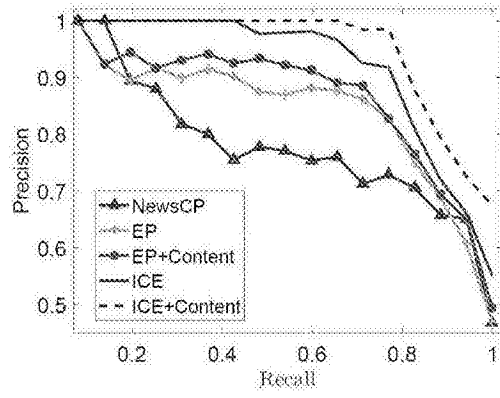


图3 (b)