



(12)发明专利申请

(10)申请公布号 CN 108804651 A

(43)申请公布日 2018.11.13

(21)申请号 201810578405.1

(22)申请日 2018.06.07

(71)申请人 南京邮电大学

地址 210003 江苏省南京市新模范马路66号

(72)发明人 薛丽 陈志 张怡婷 岳文静
金广华 郑瑶嘉 张姝彦

(74)专利代理机构 南京瑞弘专利商标事务所
(普通合伙) 32249

代理人 沈廉

(51)Int.Cl.

G06F 17/30(2006.01)

G06F 17/27(2006.01)

权利要求书2页 说明书4页

(54)发明名称

一种基于强化贝叶斯分类的社交行为检测方法

(57)摘要

本发明公开了一种基于强化贝叶斯分类的社交行为检测方法具体为:步骤1)收集中文社交网站言论的文本训练样本集,包括攻击性言论文本及赞赏性言论文本等类别文本,建立社交网站言论的文本训练样本集;步骤2)根据停用词词库对中文社交网站言论的文本训练样本集进行中文分词得到某个中文训练集的词序列 $v_i, i \in \{1, 2, 3, \dots, f\}$, f 为训练样本总数;步骤3)通过TF-IDF中文分词算法对中文社交网站言论的文本训练样本集中不同行为类别文本的进行特征提取;步骤4)输入特征词序列,使用贝叶斯模型进行学习识别;步骤5)通过强化贝叶斯分类器对分类器进行强化。本发明方法能够更加精确和高效的对人物社交行为进行检测,具有较好的应用前景。

1. 一种基于强化贝叶斯分类的社交行为检测方法,其特征在于,该方法主要包括以下步骤:

步骤1) 收集中文社交网站言论的文本训练样本集,包括攻击性言论文本及赞赏性言论文本等类别文本,建立社交网站言论的文本训练样本集;

步骤2) 根据停用词词库对中文社交网站言论的文本训练样本集进行中文分词得到某个中文训练集的词序列 $v_i, i \in \{1, 2, 3, \dots, f\}$, f 为训练样本总数;

步骤3) 通过TF-IDF中文分词算法对中文社交网站言论的文本训练样本集中不同行为类别文本的进行特征提取;

步骤4) 输入特征词序列,使用贝叶斯模型进行学习识别;

步骤5) 通过强化贝叶斯分类器对分类器进行强化。

2. 根据权利要求1所述的一种基于强化贝叶斯分类的社交行为检测方法,其特征在于,步骤3) 所述通过TF-IDF中文分词算法对中文社交网站言论的文本训练样本集中不同行为类别文本的进行特征提取,根据提取的特征词以及特征词权值更新特征词词库,具体如下:

步骤31) 计算 v_i 词语序列中的第 j 个词语 $v_{i,j}$ 的正向词频 $tf_{i,j} = \frac{n_{i,j}}{\sum n_{i,j}}$,公式中的 $n_{i,j}$ 是词语 $v_{i,j}$ 在词序列 v_i 中的出现次数, $\sum n_{i,j}$ 是次序列 v_i 所有字词的出现次数之和;

步骤32) 计算 v_i 词语序列中的第 j 个词语 $v_{i,j}$ 的逆文件词频 $idf_{i,j} = \frac{|D|}{\text{countsum}}$, $|D|$ 是语料库中的文件总数, countsum 是包含词语 $v_{i,j}$ 的文件总数;

步骤33) 计算 v_i 词语序列中的第 j 个词语 $v_{i,j}$ 的词频逆文件词频值; $tf_{i,j} \times idf_{i,j}$,保留 v_i 序列中留 $tf_{i,j} \times idf_{i,j}$ 值超过阈值 t 的词语作为特征词。

3. 根据权利要求1所述的一种基于强化贝叶斯分类的社交行为检测方法,其特征在于,所述步骤4) 输入特征词序列,使用贝叶斯模型进行学习识别的过程为:

步骤41) 特征词序列 $x = \{a_k\}, k \in \{1, 2, 3, \dots, m\}$ 为一个待分类项, a_k 为特征词序列的第 k 个特征词,总共有 m 个特征词;

步骤42) 有社交行为类别集合 $C = \{y_z\}, z \in \{1, 2, 3, \dots, n\}$,其中的 y_z 是第 z 个社交行为类别,总共有 n 社交行为类别;

步骤43) 在各个特征属性是条件独立的基础上,见算出每个类别下的各个属性的条件概率 $P(a_k | y_z) = \frac{P(y_z | a_k)}{P(y_z)}$, $P(*)$ 为贝叶斯概率公式, $*$ 为概率事件;

步骤44) 根据公式 $P(x | y_z) P(y_z) = \prod_{u=1}^m P(a_u | y_z) P(y_z)$,对每个类别的概率 $P(x | y_z) P(y_z)$ 进行计算, $P(y_z)$ 为类别 y_z 的概率。 $P(a_k | y_z)$ 是在分类属性为 a_k 条件下的的条概率, u 为迭代变量 $u \in \{1, 2, \dots, m\}$;

步骤45) 计算出 $P(x | y_z) P(y_z)$ 最大项,作为词序列 x 所属类别并得到弱分类器函数 $h(x_p)$, x_p 为词序列 x 的第 p 个属性值。

4. 根据权利要求1所述的一种基于强化贝叶斯分类的社交行为检测方法,其特征在于,所述步骤5) 加强贝叶斯分类器进行训练的方法为:

步骤51): 准备 N 个训练样本 $\{(x_p, y_q)\}, p \in \{1, 2, 3, \dots, m\}$ 组成训练集和 D , y_q 是 x_p 对应的

第 q 个社交行为类别,赋予每个样本相等的权重 $\omega = \frac{1}{N}$ 权重;

步骤52):训练样本集 D 上,利用样本权重 ω 和步骤4中学习得到弱分类器 $h(x_p)$ 计算弱分类器 $h(x_p)$ 的错误率 $\varepsilon = \sum_{i=1}^N \omega * I(h(x_p) \neq y_q)$, $I(h(x_p) \neq y_q) = \begin{cases} 1, h(x_p) \neq y_q \\ 0, \text{其他} \end{cases}$, $I(h(x_p) \neq y_q)$ 为每个分类器偏差率;

步骤53)假设 $\varepsilon > 0.5$,使用简单贝叶斯进行分类,否则计算模型迭代权值 a ,

$$a = \frac{1}{2} \ln\left(\frac{1-\varepsilon}{\varepsilon}\right);$$

步骤54)根据公式 $h^* = \arg \max_{y \in C} \sum_{i=1}^T a * I(h(x_p) \neq y_q)$,计算出强化贝叶斯分类模型 h^* 。 T 为总的行为总类数目, x 中文文本的特征词序列, a 是不同类别的加权值, h^* 是强化分类器模型, $I(h(x_p) \neq y_q)$ 为每个分类器正确率。

一种基于强化贝叶斯分类的社交行为检测方法

技术领域

[0001] 本发明涉及一种基于强化贝叶斯分类的社交行为检测方法,尤其是在在线社交网络中进行社交行为检测,通过TF-IDF中文分词算法对社交网络人物言论进行检测分类,并且提取特征词及特征值权值,之后特征词及其权值输入贝叶斯分类器中进行社交网络人物行为进行分类检测。

背景技术

[0002] 贝叶斯算法以其高效、易于实现、扩展性好的特点,已广泛的应用在行为分类的过程中此外,贝叶斯算法能够通过对中国语言样本的训练,自动学习样本内容来对社交人物行为进行检测。在现有的社交人物行为进行检测中,贝叶斯算法在行为检测中显现出了极好的应用效果。

[0003] TF-IDF是一种用于资讯检索与资讯探勘的常用加权技术。TF-IDF是一种统计方法,用以评估一字词对于一个文件集或一个语料库中的其中一份文件的重要程度。字词的重要性随着它在文件中出现的次数成正比增加,但同时会随着它在语料库中出现的频率成反比下降。

[0004] 加强学习是一种迭代的算法,会对同一个训练集使用不同的分类器训练,之后,再把这些分类器集合起来,构建一个最终的最强的分类器。其算法本身是通过改变一个权重D的分布来实现的,该权重D初始化一致,然后改变之后交给下一次分类器。使用加强分类器能够过滤掉一些不必要的训练数据特征,然后放在关键的训练数据上面。

发明内容

[0005] 技术问题:本发明公开了一种基于强化贝叶斯分类的社交行为检测方法。该方法首先将人物的言论量化为行为向量,行为向量的每个维度的值相当于人物对该行为的取值。使用朴素贝叶斯分类模型实现对人物行为向量进行简单分类。

[0006] 技术方案:本发明的一种基于强化贝叶斯分类的社交行为检测方法主要包括以下步骤:

[0007] 步骤1)收集中文社交网站言论的文本训练样本集,包括攻击性言论文本及赞赏性言论文本等类别文本,建立社交网站言论的文本训练样本集;

[0008] 步骤2)根据停用词词库对中文社交网站言论的文本训练样本集进行中文分词得到某个中文训练集的词序列 $V_i, i \in \{1, 2, 3, \dots, f\}$, f 为训练样本总数;

[0009] 步骤3)通过TF-IDF中文分词算法对中文社交网站言论的文本训练样本集中不同行为类别文本的进行特征提取;

[0010] 步骤4)输入特征词序列,使用贝叶斯模型进行学习识别;

[0011] 步骤5)通过强化贝叶斯分类器对分类器进行强化。

[0012] 其中,

[0013] 步骤3)所述通过TF-IDF中文分词算法对中文社交网站言论的文本训练样本集中

不同行为类别文本的进行特征提取,根据提取的特征词以及特征词权值更新特征词词库,具体如下:

[0014] 步骤31) 计算 v_i 词语序列中的第 j 个词语 $v_{i,j}$ 的正向词频 $tf_{i,j} = \frac{n_{i,j}}{\sum n_{i,j}}$,公式中的 $n_{i,j}$ 是词语 $v_{i,j}$ 在词序列 v_i 中的出现次数, $\sum n_{i,j}$ 是次序列 v_i 所有字词的出现次数之和;

[0015] 步骤32) 计算 v_i 词语序列中的第 j 个词语 $v_{i,j}$ 的逆文件词频 $idf_{i,j} = \frac{|D|}{\text{countsum}}$, $|D|$ 是语料库中的文件总数,countsum是包含词语 $v_{i,j}$ 的文件总数;

[0016] 步骤33) 计算 v_i 词语序列中的第 j 个词语 $v_{i,j}$ 的词频逆文件词频值; $tf_{i,j} \cdot idf_{i,j} = tf_{i,j} \times idf_{i,j}$,保留 v_i 序列中留 $tf_{i,j} \cdot idf_{i,j}$ 值超过阈值 t 的词语作为特征词。

[0017] 所述步骤4) 输入特征词序列,使用贝叶斯模型进行学习识别的过程为:

[0018] 步骤41)、特征词序列 $x = \{a_k\}, k \in \{1, 2, 3, \dots, m\}$ 为一个待分类项, a_k 为特征词序列的第 k 个特征词,总共有 m 个特征词;

[0019] 步骤42) 有社交行为类别集合 $C = \{y_z\}, z \in \{1, 2, 3, \dots, n\}$,其中的 y_z 是第 z 个社交行为类别,总共有 n 社交行为类别;

[0020] 步骤43) 在各个特征属性是条件独立的基础上,见算出每个类别下的各个属性的条件概率 $P(a_k | y_z) = \frac{P(y_z | a_k)}{P(y_z)}$, $P(*)$ 为贝叶斯概率公式, $*$ 为概率事件;

[0021] 步骤44) 根据公式 $P(x | y_z) P(y_z) = \prod_{u=1}^m P(a_k | y_z) P(y_z)$,对每个类别的概率 $P(x | y_z) P(y_z)$ 进行计算, $P(y_z)$ 为类别 y_z 的概率。 $P(a_k | y_z)$ 是在分类属性为 a_k 条件下的条概率, u 为迭代变量 $u \in \{1, 2, \dots, m\}$;

[0022] 步骤45) 计算出 $P(x | y_z) P(y_z)$ 最大项,作为词序列 x 所属类别并得到弱分类器函数 $h(x_p)$, x_p 为词序列 x 的第 p 个属性值。

[0023] 所述步骤5) 加强贝叶斯分类器进行训练的方法为:

[0024] 步骤51): 准备 N 个训练样本 $\{(x_p, y_q)\}, p \in \{1, 2, 3, \dots, m\}$ 组成训练集和 D , y_q 是 x_p 对应的第 q 个社交行为类别,赋予每个样本相等的权重 $\omega = \frac{1}{N}$ 权重;

[0025] 步骤52): 训练样本集 D 上,利用样本权重 ω 和步骤4中学习得到弱分类器 $h(x_p)$ 计算弱分类器 $h(x_p)$ 的错误率 $\varepsilon = \sum_{i=1}^N \omega * I(h(x_p) \neq y_q)$, $I(h(x_p) \neq y_q) = \begin{cases} 1, h(x_p) \neq y_q \\ 0, \text{其他} \end{cases}$, $I(h(x_p) \neq y_q)$ 为每个分类器偏差率;

[0026] 步骤53) 假设 $\varepsilon > 0.5$,使用简单贝叶斯进行分类,否则计算模型迭代权值 a , $a = \frac{1}{2} \ln\left(\frac{1-\varepsilon}{\varepsilon}\right)$;

[0027] 步骤54) 根据公式 $h^* = \arg \max_{y \in C} \sum_{i=1}^T a^* I(h(x_p) \neq y_q)$,计算出强化贝叶斯分类模型 h^* 。 T 为总的行为总类数目, x 中文文本的特征词序列, a 是不同类别的加权值, h^* 是强化分类器模型, $I(h(x_p) \neq y_q)$ 为每个分类器正确率。

[0028] 有益效果: 本发明采用以上技术方案与现有技术相比:

[0029] (1) 本发明算法较为简单,节省了计算成本,能够更好地提取文章关键词

[0030] (2) 本发明方法基于词典对训练集的属性进行筛选,并且朴素贝叶斯分类对数据进行分类,使用使用强化贝叶斯分类模型对模型进行连接,能够更加精确和高效的对人物社交行为进行检测,检测准确度非常高。

具体实施方式

[0031] 本发明的一种基于强化贝叶斯分类的社交行为检测方法方法主要包括以下步骤:

[0032] 步骤1) 收集中文社交网站言论的文本训练样本集,包括攻击性言论文本及赞赏性言论文本等类别文本,建立社交网站言论的文本训练样本集。

[0033] 步骤2) 根据停用词词库对中文社交网站言论的文本训练样本集进行中文分词得到某个中文训练集的词序列 $V_i, i \in \{1, 2, 3, \dots, f\}$, f 为训练样本总数, f 为500, 第 i 个文本训练集为内容为“这是一个非常优秀的人,喜欢帮助别人,热爱学习”,分词后的词序列 V_i 为{优秀,帮助,热爱,学习,非常,这,是,一个,喜欢,帮助};

[0034] 步骤3) 通过TF-IDF中文分词算法对中文社交网站言论的文本训练样本集中不同行为类别文本的进行特征提取,步骤如下:

[0035] 步骤31) 计算 v_i 词语序列中的第 j 个词语 $v_{i,j}$ 的正向词频 $tf_{i,j} = \frac{n_{i,j}}{\sum n_{i,j}}$, 公式中的 $n_{i,j}$ 是词语 $v_{i,j}$ 在词序列 v_i 中的出现次数, $\sum n_{i,j}$ 是次序列 v_i 所有字词的出现次数之和;

[0036] 步骤32) 计算 v_i 词语序列中的第 j 个词语 $v_{i,j}$ 的逆文件词频 $idf_{i,j} = \frac{|D|}{\text{countsum}}$, $|D|$ 是语料库中的文件总数, countsum 是包含词语 $v_{i,j}$ 的文件总数;

[0037] 步骤33) 计算 v_i 词语序列中的第 j 个词语 $v_{i,j}$ 的词频逆文件词频值 $tf_{i,j} \cdot idf_{i,j} = tf_{i,j} \times idf_{i,j}$, 保留 v_i 序列中留 $tf_{i,j} \cdot idf_{i,j}$ 值超过阈值 t 的词语作为特征词, 阈值 t 为50。

[0038] 步骤4) 输入特征词序列,使用贝叶斯模型进行学习识别,具体步骤如下:

[0039] 步骤41) 特征词序列 $x = \{a_k\}, k \in \{1, 2, 3, \dots, m\}$ 为一个待分类项, a_k 为特征词序列的第 k 个特征词, 总共有 m 个特征词。特征词序列 x 为{帮助,喜欢,热爱}, m 的值为4。社交行为类别集合 $C = \{\text{赞赏,攻击,喜爱,痴迷,讨厌}\}, n$ 为5。

[0040] 步骤42) 有社交行为类别集合 $C = \{y_z\}, z \in \{1, 2, 3, \dots, n\}$, 其中的 y_z 是第 z 个社交行为类别, 总共有 n 社交行为类别。

[0041] 步骤43) 在各个特征属性是条件独立的基础上, 见算出每个类别下的各个属性的

条件概率 $P(a_k | y_z) = \frac{P(y_z | a_k)}{P(y_z)}$, $P(*)$ 为贝叶斯概率公式, $*$ 为概率事件。

[0042] 步骤44) 根据公式 $P(x | y_z) P(y_z) = \prod_{u=1}^m P(a_k | y_z) P(y_z)$, 对每个类别的概率 $P(x | y_z) P(y_z)$

进行计算, $P(y_z)$ 为类别 y_z 的概率。 $P(a_k | y_z)$ 是在分类属性为 a_k 条件下的的条概率, u 为迭代变量 $u \in \{1, 2, \dots, m\}$ 。

[0043] 步骤45) 计算出 $P(x | y_z) P(y_z)$ 最大项, 作为词序列 x 所属类别并得到弱分类器函数 $h(x_p)$, x_p 为词序列 x 的第 p 个属性值。

[0044] 步骤5) 通过强化贝叶斯分类器对分类器进行强化, 具体步骤如下:

[0045] 步骤51) : 准备 N 个训练样本 $\{(x_p, y_q)\}, p \in \{1, 2, 3, \dots, m\}$ 组成训练集和 D, y_q 是 x_p 对

应的第 q 个社交行为类别,赋予每个样本相等的权重 $\omega = \frac{1}{N}$ 权重, N 为500, ω 为0.002;

[0046] 步骤52):训练样本集 D 上,利用样本权重 ω 和步骤4中学习得到弱分类器 $h(x_p)$ 计算弱分类器 $h(x_p)$ 的错误率 $\varepsilon = \sum_{i=1}^N \omega * I(h(x_p) \neq y_q), I(h(x_p) \neq y_q) = \begin{cases} 1, h(x_p) \neq y_q \\ 0, \text{其他} \end{cases}, I(h(x_p) \neq y_q)$ 为每个分类器偏差率。

[0047] 步骤53) 假设 $\varepsilon = 0.6$,使用简单贝叶斯进行分类,否则计算模型迭代权值 a ,
 $a = \frac{1}{2} \ln\left(\frac{1-\varepsilon}{\varepsilon}\right)$ 。

[0048] 步骤54) 根据公式 $h^* = \arg \max_{y \in C} \sum_{i=1}^T a^* I(h(x_p) \neq y_q)$,计算出强化贝叶斯分类模型 h^* 。 T 为总的行为总类数目, x 中文文本的特征词序列, a 是不同类别的加权值, h^* 是强化分类器模型, $I(h(x_p) \neq y_q)$ 为每个分类器正确率。

[0049] 通过该方法验证得知,该文本社交数据属于攻击性言论。