(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109325116 A
(43)申请公布日 2019.02.12

(21)申请号 201810963174.6

(22)申请日 2018.08.23

(71)申请人 武大吉奥信息技术有限公司
地址 430223 湖北省武汉市东湖开发区庙山小区江夏大道武大科技园

(72)发明人 吴杰 王琳 杨曦 刘奕夫 沈满 周游宇 张定祥 贺楷锴 官磊 张立 朱斌 寇晓松

(74)专利代理机构 北京双收知识产权代理有限公司 11241
代理人 曾晓芒

(51)Int.Cl.
G06F 16/35(2019.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)
G06Q 50/26(2012.01)

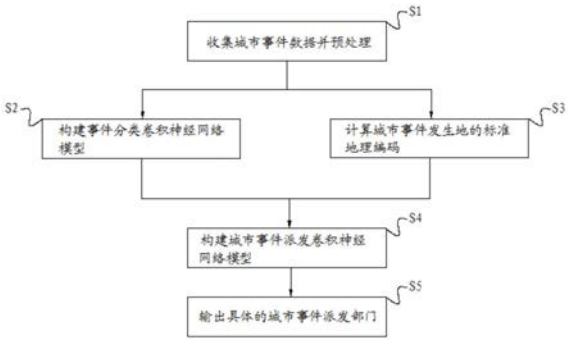
权利要求书3页 说明书8页 附图2页

(54)发明名称

一种基于深度学习的城市事件自动分类派发方法及装置

(57)摘要

本发明适用于智慧城市信息智能化技术领域，提供一种基于深度学习的城市事件自动分类派发方法及装置，包括收集城市事件数据并预处理；根据收集处理的城市事件数据构建事件分类卷积神经网络模型；计算城市事件发生地的标准地理编码；构建城市事件派发卷积神经网络模型；接收当前输入的城市事件数据，调用所述分类卷积神经网络模型输出分类类别，获取当前城市事件数据的标准地理编码，然后调用派发卷积神经网络模型，输出具体的城市事件派发部门。本发明可以提高事件的分类正确性和派送准确性；通过卷积神经网络模型派发，相比人工派发带来的各种不确定性，机器派发的准确性更高，本发明中机器根据模型一次运算得到结果，可有效提升系统运行效率。



CN 109325116 A

1.一种基于深度学习的城市事件自动分类派发方法,其特征在于,所述方法包括下述步骤:

步骤S1、收集城市事件数据并预处理;

步骤S2、根据收集处理的城市事件数据构建事件分类卷积神经网络模型;

步骤S3、计算城市事件发生地的标准地理编码;

步骤S4、构建城市事件派发卷积神经网络模型;

步骤S5,接收当前输入的城市事件数据,调用所述分类卷积神经网络模型输出分类类别,获取当前城市事件数据的标准地理编码,然后调用派发卷积神经网络模型,输出具体的城市事件派发部门。

2.如权利要求1所述基于深度学习的城市事件自动分类派发方法,其特征在于,所述步骤S1具体包括下述步骤:

步骤S1.1、收集历年城市事件文本数据,结合常用的分词字典,对收集的文本数据进行分词过滤,得到适合于城市事件的分词字典;

步骤S1.2、获取城市的标准地名地址库,每个标准地名地址都包含唯一的地理编码,并且基于标准地名地址库制作地名地址字典;

步骤S1.3、根据所述分词字典和地名地址字典对所有城市事件文本数据进行分词,并统计词频,选取词频较低的词作为停用词,建立适用于城市事件的停用词字典;

步骤S1.4、对所有城市事件的类别按照名称进行编码,每个类别对应一个类别编号,并且增加一个其他的类别作为预留类别;

步骤S1.5、根据上述步骤S1.1-S1.3得到的分词字典、地名地址字典和停用词字典,对所有的城市事件文本数据逐条进行预处理,包括分词、去停用词、去地名地址处理,同时定义一个常量K,作为每个城市事件文本预处理后保留的最终单词数量;

对单条城市事件文本数据,若预处理完毕得到单词数量为0,则视该城市事件为无效城市事件进行剔除,否则视为有效城市事件;若预处理完毕得到的单词数量超过K,则保留前面的K个单词作为最终单词;若处理完毕的单词个数介于0到K之间,则以空值UNK填充,因此每条有效城市事件文本数据预处理后单词数量均为K,然后按照步骤S1.4的编码方式提取每条有效城市事件的类别编号,将类别编号和有效城市事件的预处理结果作为训练样本,统计样本数量;

步骤S1.6、对所有训练样本中的所有单词,每个单词赋予一个唯一词编码,其中填充的空值UNK统一用一个词编码,并建立对应关系词表;

步骤S1.7、对城市事件的处理部门进行数字化编码。

3.如权利要求2所述基于深度学习的城市事件自动分类派发方法,其特征在于,所述步骤S2中分类卷积神经网络模型为内嵌词向量训练的卷积神经网络分类模型,并分为以下部分:

输入层:输入一个城市事件的训练样本,即步骤S1.5中的类别编号和预处理后的K个单词;

嵌入层:对于输入层中的K个单词进行查表操作,从对应关系词表中查出对应的词编码,并将单词转为词向量形式;

特征提取层:构建3种过滤器,每种过滤器128个,每个过滤器包含卷积层和池化层,并

在两层中间加入非线性激活函数处理,3种过滤器将同步对词向量进行卷积和池化操作;

全连接层:连接所有词向量经过过滤器池化后的特征值;

激活层:选用softmax函数归一化计算城市事件属于每个类别的概率;

输出层:输出概率最大的类别作为预测分类类别。

4.如权利要求3所述基于深度学习的城市事件自动分类派发方法,其特征在于,所述步骤S3具体包括下述步骤:

判断城市事件文本数据是否带有地理坐标信息,如果带有地理坐标信息,则通过计算地理坐标信息与标准地名地址库中的行政区划的空间关系,判断城市事件属于哪个行政区划,进而取得行政区划对应的标准地理编码;如果不带有地理坐标信息,首先根据步骤S1.2中的地名地址字典提取城市事件中的详细地址,将城市事件中的详细地址输入到标准地名地址库中做文本匹配,选取匹配最高的标准地址编码。

5.如权利要求4所述基于深度学习的城市事件自动分类派发方法,其特征在于,所述步骤S4中派发卷积神经网络模型为内嵌词向量训练的卷积神经网络分类模型,具体分为以下部分:

输入层:输入步骤S2中计算得到的预测分类类别、步骤S3中提取的城市事件标准地址编码及步骤S1中处理部门的数字化编码组成的样本数据;

特征提取层:构建3种过滤器,每种过滤器128个,每个过滤器包含卷积层和池化层,并在两层中间加入非线性激活函数处理,3种过滤器将同步对样本数据进行卷积和池化操作;

全连接层:连接所有样本数据经过过滤器池化后的特征值;

激活层:选用softmax函数归一化计算样本数据属于每个处理部门的概率;

输出层:输出概率最大的处理部门作为派发部门。

6.一种基于深度学习的城市事件自动分类派发装置,其特征在于,所述装置包括:

数据处理模块:用于收集城市事件数据并预处理;

事件分类模型构建模块:用于将收集处理的城市事件数据构建事件分类卷积神经网络模型;

地理坐标信息模块:用于计算城市事件发生地的标准地理编码;

事件派发模型构建模块:用于构建城市事件派发卷积神经网络模型;

事件输出模块:用于接收当前输入的城市事件数据,调用所述分类卷积神经网络模型输出分类类别,获取当前城市事件数据的标准地理编码,然后调用派发卷积神经网络模型,输出具体的城市事件派发部门。

7.如权利要求6所述基于深度学习的城市事件自动分类派发装置,其特征在于,所述数据处理模块包括:

城市事件收集单元:用于收集历年城市事件文本数据,结合常用的分词字典,对收集的文本数据进行分词过滤,得到适合于城市事件的分词字典;

城市地址制作单元:用于获取城市的标准地名地址库,每个标准地名地址都包含唯一的地理编码,并且基于标准地名地址库制作地名地址字典;

停用词字典建立单元:用于根据所述分词字典和地名地址字典对所有城市事件文本数据进行分词,并统计词频,选取词频较低的词作为停用词,建立适用于城市事件的停用词字典;

城市事件编码单元:用于对所有的城市事件的类别按照名称进行编码,每个类别对应一个类别编号,并且增加一个其他的类别作为预留类别;

样本建立单元:用于获取最终单词数量,进而作为训练样本,统计样本数量;

对应关系词表建立单元:用于对所有训练样本中的所有单词,每个单词赋予一个唯一词编码,其中填充的空值UNK统一用一个词编码,并建立对应关系词表;

处理部门编码单元:用于对城市事件的处理部门进行数字化编码。

8.如权利要求7所述基于深度学习的城市事件自动分类派发装置,其特征在于,所述事件分类模型构建模块包括:

第一输入单元:用于输入一个城市事件的训练样本;

嵌入单元,用于对于输入层中的训练样本进行查表操作,从对应关系词表中查出对应的词编码,并将单词转为词向量形式;

第一特征提取单元:用于构建3种过滤器,每种过滤器128个,每个过滤器包含卷积层和池化层,并在两层中间加入非线性激活函数,3种过滤器将同步对词向量进行卷积和池化操作;

第一全连接单元:用于连接所有词向量经过过滤器池化后的特征值;

第一激活单元:用于选用softmax函数归一化计算城市事件属于每个类别的概率;

第一输出单元:用于输出概率最大的类别作为预测分类类别。

9.如权利要求8所述基于深度学习的城市事件自动分类派发装置,其特征在于,所述事件派发模型构建模块包括:

第二输入单元:用于输入由计算得到的预测分类类别、提取的城市事件标准地址编码及处理部门的数字化编码组成的样本数据;

第二特征提取单元:用于构建3种过滤器,每种过滤器128个,每个过滤器包含卷积层和池化层,并在两层中间加入非线性激活函数,3种过滤器将同步对样本数据进行卷积和池化操作;

第二全连接单元:用于连接所有样本数据经过过滤器池化后的特征值;

第二激活单元:用于选用softmax函数归一化计算样本数据属于每个处理部门的概率;

第二输出单元:用于将输出概率最大的处理部门作为派发部门。

一种基于深度学习的城市事件自动分类派发方法及装置

技术领域

[0001] 本发明属于智慧城市信息智能化技术领域，尤其涉及一种基于深度学习的城市事件自动分类派发方法及装置。

背景技术

[0002] 12345市长专线平台是一个接收市民反映各种问题的系统平台，其工作流程为：对于每日市民反映投诉的事件，由前台的接线人员受理，并根据事件内容和性质总结出事件的归属类别，再根据类别以及事件发生的地点将投诉事件转送至对应的办理机构或者政府部门。这个过程中要求接线人员对所有的事件性质、事件发生的详细地址和对应处理部门之间的关系都了解得非常清楚，如果派发错误，处理部门会将不属于本部门处理的事件退回至前台接线人员处，重新派发，市民投诉的事件类别非常多，涉及城市交通、城市市容市貌、市场管理、房屋土地管理、社会服务、教育、食品药品医疗卫生管理等等，大大小小的类别有几百种；同时相应的处理部门按照区域级别划分，可分为各个辖区政府部门、直属机构、事业单位等等，也有上百个。一个大中型城市中，一天的投诉事件数量少则上百，多达上万。如果完全依靠人工判断事件性质和类别，并找到正确的处理部门，这个工作量非常大，对接线办事人员的要求也很高。

[0003] 目前，处理投诉事件的方法是：针对几百种事件性质类别，根据其含义先分别建立正则规则；当市民投诉一个事件，按照正则规则判断其事件性质类别，结合事件的发生地址，把事件派送到处理部门。即：当市民来电投诉一个事件，系统根据投诉事件内容在所有的事件类型的正则规则进行逐一查找过滤，直至找到匹配的规则，作为该事件的类别，同时根据市民描述的地址获取其精确地址，再根据类别和地址把事件派发至处理部门。这个方法要求每个类别的正则关键词和表达式必须是精确的，因为当正则规则中的关键词不能覆盖事件中的词语时，事件将不能被派送。出现这种情况时，及时根据当前事件的核心内容在对应类型的正则表达式中添加规则，再重新派送。

[0004] 上述方法中，补充关键词、正则表达式等步骤属于人工处理过程，而且这个过程是需要持续开展的，此方法人工成本非常高。此外，每条事件都要经过N个正则表达式计算后才能得到结果，导致系统运行效率较低。

发明内容

[0005] 鉴于上述问题，本发明的目的在于提供一种基于深度学习的城市事件自动分类派发方法及装置，旨在解决现有处理方法出现错误率高。人工成本巨大，系统计算资源消耗较大等技术问题。

[0006] 本发明采用如下技术方案：

[0007] 所述基于深度学习的城市事件自动分类派发方法包括如下步骤：

[0008] 步骤S1、收集城市事件数据并预处理；

[0009] 步骤S2、根据收集处理的城市事件数据构建事件分类卷积神经网络模型；

[0010] 步骤S3、计算城市事件发生地的标准地理编码；

[0011] 步骤S4、构建城市事件派发卷积神经网络模型；

[0012] 步骤S5，接收当前输入的城市事件数据，调用所述分类卷积神经网络模型输出分类类别，获取当前城市事件数据的标准地理编码，然后调用派发卷积神经网络模型，输出具体的城市事件派发部门。

[0013] 进一步的，所述步骤S1具体包括下述步骤：

[0014] 步骤S1.1、收集历年城市事件文本数据，结合常用的分词字典，对收集的文本数据进行分词过滤，得到适合于城市事件的分词字典；

[0015] 步骤S1.2、获取城市的标准地名地址库，每个标准地名地址都包含唯一的地理编码，并且基于标准地名地址库制作地名地址字典；

[0016] 步骤S1.3、根据所述分词字典和地名地址字典对所有城市事件文本数据进行分词，并统计词频，选取词频较低的词作为停用词，建立适用于城市事件的停用词字典；

[0017] 步骤S1.4、对所有城市事件的类别按照名称进行编码，每个类别对应一个类别编号，并且增加一个其他的类别作为预留类别；

[0018] 步骤S1.5、根据上述步骤S1.1-S1.3得到的分词字典、地名地址字典和停用词字典，对所有的城市事件文本数据逐条进行预处理，包括分词、去停用词、去地名地址处理，同时定义一个常量K，作为每个城市事件文本预处理后保留的最终单词数量；

[0019] 对单条城市事件文本数据，若预处理完毕得到单词数量为0，则视该城市事件为无效城市事件进行剔除，否则视为有效城市事件；若预处理完毕得到的单词数量超过K，则保留前面的K个单词作为最终单词；若处理完毕的单词个数介于0到K之间，则以空值UNK填充，因此每条有效城市事件文本数据预处理后单词数量均为K，然后按照步骤S1.4的编码方式提取每条有效城市事件的类别编号，将类别编号和有效城市事件的预处理结果作为训练样本，统计样本数量；

[0020] 步骤S1.6、对所有训练样本中的所有单词，每个单词赋予一个唯一词编码，其中填充的空值UNK统一用一个词编码，并建立对应关系词表；

[0021] 步骤S1.7、对城市事件的处理部门进行数字化编码。

[0022] 进一步的，所述步骤S2中分类卷积神经网络模型为内嵌词向量训练的卷积神经网络分类模型，并分为以下部分：

[0023] 输入层：输入一个城市事件的训练样本，即步骤S1.5中的类别编号和预处理后的K个单词；

[0024] 嵌入层：对于输入层中的K个单词进行查表操作，从对应关系词表中查出对应的词编码，并将单词转为词向量形式；

[0025] 特征提取层：构建3种过滤器，每种过滤器128个，每个过滤器包含卷积层和池化层，并在两层中间加入非线性激活函数处理，3种过滤器将同步对词向量进行卷积和池化操作；

[0026] 全连接层：连接所有词向量经过过滤器池化后的特征值；

[0027] 激活层：选用softmax函数归一化计算城市事件属于每个类别的概率；

[0028] 输出层：输出概率最大的类别作为预测分类类别。

[0029] 进一步的，所述步骤S3具体包括下述步骤：

[0030] 判断城市事件文本数据是否带有地理坐标信息，如果带有地理坐标信息，则通过计算地理坐标信息与标准地名地址库中的行政区划的空间关系，判断城市事件属于哪个行政区划，进而取得行政区划对应的标准地理编码；如果不带有地理坐标信息，首先根据步骤S1.2中的地名地址字典提取城市事件中的详细地址，将城市事件中的详细地址输入到标准地名地址库中做文本匹配，选取匹配最高的标准地址编码。

[0031] 进一步的，所述步骤S4中派发卷积神经网络模型为内嵌词向量训练的卷积神经网络分类模型，具体分为以下部分：

[0032] 输入层：输入步骤S2中计算得到的预测分类类别、步骤S3中提取的城市事件标准地址编码及步骤S1中处理部门的数字化编码组成的样本数据；

[0033] 特征提取层：构建3种过滤器，每种过滤器128个，每个过滤器包含卷积层和池化层，并在两层中间加入非线性激活函数处理，3种过滤器将同步对样本数据进行卷积和池化操作；

[0034] 全连接层：连接所有样本数据经过过滤器池化后的特征值；

[0035] 激活层：选用softmax函数归一化计算样本数据属于每个处理部门的概率；

[0036] 输出层：输出概率最大的处理部门作为派发部门。

[0037] 另一方面，所述基于深度学习的城市事件自动分类派发装置，包括如下模块：

[0038] 数据处理模块：用于收集城市事件数据并预处理；

[0039] 事件分类模型构建模块：用于将收集处理的城市事件数据构建事件分类卷积神经网络模型；

[0040] 地理坐标信息模块：用于计算城市事件发生地的标准地理编码；

[0041] 事件派发模型构建模块：用于构建城市事件派发卷积神经网络模型；

[0042] 事件输出模块：用于接收当前输入的城市事件数据，调用所述分类卷积神经网络模型输出分类类别，获取当前城市事件数据的标准地理编码，然后调用派发卷积神经网络模型，输出具体的城市事件派发部门。

[0043] 进一步的，所述数据处理模块包括：

[0044] 城市事件收集单元：用于收集历年城市事件文本数据，结合常用的分词字典，对收集的文本数据进行分词过滤，得到适合于城市事件的分词字典；

[0045] 城市地址制作单元：用于获取城市的标准地名地址库，每个标准地名地址都包含唯一的地理编码，并且基于标准地名地址库制作地名地址字典；

[0046] 停用词字典建立单元：用于根据所述分词字典和地名地址字典对所有城市事件文本数据进行分词，并统计词频，选取词频较低的词作为停用词，建立适用于城市事件的停用词字典；

[0047] 城市事件编码单元：用于对所有的城市事件的类别按照名称进行编码，每个类别对应一个类别编号，并且增加一个其他的类别作为预留类别；

[0048] 样本建立单元：用于获取最终单词数量，进而作为训练样本，统计样本数量；

[0049] 对应关系词表建立单元：用于对所有训练样本中的所有单词，每个单词赋予一个唯一词编码，其中填充的空值UNK统一用一个词编码，并建立对应关系词表；

[0050] 处理部门编码单元：用于对城市事件的处理部门进行数字化编码。

[0051] 进一步的，所述事件分类模型构建模块包括：

[0052] 第一输入单元:用于输入一个城市事件的训练样本;

[0053] 嵌入单元,用于对于输入层中的训练样本进行查表操作,从对应关系词表中查出对应的词编码,并将单词转为词向量形式;

[0054] 第一特征提取单元:用于构建3种过滤器,每种过滤器128个,每个过滤器包含卷积层和池化层,并在两层中间加入非线性激活函数,3种过滤器将同步对词向量进行卷积和池化操作;

[0055] 第一全连接单元:用于连接所有词向量经过过滤器池化后的特征值;

[0056] 第一激活单元:用于选用softmax函数归一化计算城市事件属于每个类别的概率;

[0057] 第一输出单元:用于输出概率最大的类别作为预测分类类别。

[0058] 进一步的,所述事件派发模型构建模块包括:

[0059] 第二输入单元:用于输入由计算得到的预测分类类别、提取的城市事件标准地址编码及处理部门的数字化编码组成的样本数据;

[0060] 第二特征提取单元:用于构建3种过滤器,每种过滤器128个,每个过滤器包含卷积层和池化层,并在两层中间加入非线性激活函数,3种过滤器将同步对样本数据进行卷积和池化操作;

[0061] 第二全连接单元:用于连接所有样本数据经过过滤器池化后的特征值;

[0062] 第二激活单元:用于选用softmax函数归一化计算样本数据属于每个处理部门的概率;

[0063] 第二输出单元:用于将输出概率最大的处理部门作为派发部门。

[0064] 本发明的有益效果是:相对于现有技术,本发明可以提高事件的分类正确性和派送准确性;通过卷积神经网络模型派发,相比人工派发带来的各种不确定性,机器派发的准确性更高,本发明中机器根据模型一次运算得到结果,可有效提升系统运行效率;采用机器学习的方法,让机器持续训练更新模型,人工成本低。

附图说明

[0065] 图1是本发明提供的基于深度学习的城市事件自动分类派发方法流程图;

[0066] 图2是本发明提供的城市事件数据构建事件分类卷积神经网络模型结构图;

[0067] 图3是本发明实施例提供的基于深度学习的城市事件自动分类派发装置结构图。

具体实施方式

[0068] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0069] 为了说明本发明所述的技术方案,下面通过具体实施例来进行说明。

[0070] 实施例一:

[0071] 如图1所示,本发明实施例提供的基于深度学习的城市事件自动分类派发方法包括如下步骤:

[0072] 步骤S1、收集城市事件数据并预处理。

[0073] 假设需要收集及预处理的是A城市的事件数据,所述城市事件是市民通过热线电

话、政府网站、移动终端等方式反馈的城市问题，比如：公积金问题、停水停电问题、拆迁问题等等，城市事件的形式可能是文本或者语音，语音形式的城市事件作为系统的输入数据，可增加语音转换文本，使其可以作为城市事件文本数据。

[0074] 所述步骤S1具体包括下述步骤：

[0075] 步骤S1.1、收集城市历年城市事件文本数据，假设该城市为A城市，结合常用的分词字典，对收集的文本数据进行分词过滤，得到适合于该城市事件的分词字典；

[0076] 步骤S1.2、获取A城市的标准地名地址库，如果没有标准地址库，可用标准地名地址服务代替，如：当地政府提供的标准地名地址服务，也可以使用百度、高德、腾讯、谷歌、bing等提供的地名地址服务，标准地名地址库应包含：标准地址信息、行政区划信息、编码信息、关联的地名信息等，每个标准地名地址都包含唯一的地理编码，在本发明中，假设共有g个地理编码，并且基于标准地名地址库制作地名地址字典；

[0077] 步骤S1.3、根据所述分词字典和地名地址字典对所有城市事件文本数据进行分词，并统计词频，所述词频指的是某个给定的词语在该城市事件文本中出现的次数，选取词频较低的词作为停用词，所述停用词是在处理自然语言数据或文本数据之前或之后会自动过滤掉某些字或词，这些字或词即被称为Stop Words(停用词)，建立适用于城市事件的停用词字典；

[0078] 步骤S1.4、对所有城市事件的类别按照名称进行编码，每个类别对应一个类别编号，并且增加一个其他的类别作为预留类别；假设有M个城市事件，并对每个城市事件从0开始编码到m-1，预留类别编号为m；

[0079] 步骤S1.5、根据上述步骤S1.1-S1.3得到的分词字典、地名地址字典和停用词字典，对所有的城市事件文本数据逐条进行预处理，包括分词、去停用词、去地名地址处理，同时定义一个常量K，作为每个城市事件文本预处理后保留的最终单词数量，这里取K＝12；

[0080] 对单条城市事件文本数据，若预处理完毕得到单词数量为0，则视该城市事件为无效城市事件进行剔除，否则视为有效城市事件；若预处理完毕得到的单词数量超过K，则保留前面的K个单词作为最终单词；若处理完毕的单词个数介于0到K之间，则以空值UNK填充，因此每条有效城市事件文本数据预处理后单词数量均为K，然后按照步骤S1.4的编码方式提取每条有效城市事件的类别编号，将类别编号和有效城市事件的预处理结果作为训练样本，统计样本数量，假设统计样本数量为n个；

[0081] 步骤S1.6、对所有训练样本中的所有单词，每个单词赋予一个唯一词编码，其中填充的空值UNK统一用一个词编码，并建立对应关系词表；

[0082] 步骤S1.7、对城市事件的处理部门进行数字化编码，假设有P个城市事件处理部门，编码是从0到p-1的自然数序列。

[0083] 步骤S2、根据收集处理的城市事件数据构建事件分类卷积神经网络模型。

[0084] 按照本技术方案的流程，城市事件在派发前要先得到其类别，那么可以利用步骤S1得到的城市事件文本数据构建一个分类器，通过分类器，输入是城市事件文本数据，输出是其预测的类别。但是城市事件是文本数据，且事件文本数据存在上下文语义关系，一般的分类器不能直接做文本数据分类，即使按照上面步骤S1.6进行编号，其中不存在语义关系。卷积神经网络(简称CNN)在图像识别、文本分类等领域应用较广，卷积神经网络是一种前馈神经网络，它的人工神经元可以响应一部分覆盖范围内的周围单元，对于大型图像处理有

出色表现。它包括卷积层(convolutional layer)和池化层(pooling layer)。但网络的输入层节点代表的都是数值,所以本方案对样本中的所有词语进行单词向量化和编码化,词向量(Word embedding),又叫Word嵌入是自然语言处理(NLP)中的一组语言建模和特征学习技术的统称,其中来自词汇表的单词或短语被映射到实数的向量。从概念上讲,它涉及从每个单词一维的空间到具有更低维度的连续向量空间的数学嵌入。在几何角度,向量可以很好的描述出两个对象之间的相似性,可以通过大量的数据不断训练得到单词的代表向量,使得词义相近的词或者有上下文关联的词向量之间比较接近。这个词向量训练的过程可以放进卷积神经网络分类模型中,称之为内嵌词向量训练的卷积神经网络分类模型。

[0085] 如图2所示,所述步骤S2中分类卷积神经网络模型为内嵌词向量训练的卷积神经网络分类模型,并分为以下部分:

[0086] 输入层:输入一个城市事件的训练样本,即步骤S1.5中的类别编号和预处理后的K个单词;

[0087] 嵌入层:对于输入层中的K个单词进行查表操作,从对应关系词表中查出对应的词编码,并将单词转为词向量形式,每个词向量长度为embedding_dim,这里设置embedding_dim=256,嵌入层共包含k*embedding_dim个数值;

[0088] 特征提取层:构建3种过滤器,过滤器的核大小分别为3*3,4*4,5*5,每种过滤器128个,每个过滤器包含卷积层和池化层,并在两层中间加入非线性激活函数处理,3种过滤器将同步对词向量进行卷积和池化操作,池化选择最大的值特征;

[0089] 全连接层:连接所有词向量经过过滤器池化后的特征值;

[0090] 激活层:选用softmax函数归一化计算城市事件属于每个类别的概率;

[0091] 输出层:输出概率最大的类别作为预测分类类别。

[0092] 在本方案中,如图2所示构建的分类模型,输入步骤S1处理好的训练样本,并训练模型。卷积神经网络模型可以边训练边优化模型,即将训练样本分批抽样迭代训练模型,对于步骤S1中的样本数量为n,每个迭代周期抽样选择10%的数据作为验证集,其余90%的数据分批次进行训练:每次放入模型中训练数量batch_size=1024,那么通过0.9*n/batch_size次训练,完成一个迭代周期,并且每训练一次模型都根据验证结果进行自我修正,总的迭代周期设置为50000,迭代周期计算完毕,模型的准确度达到99.8%。

[0093] 在本发明中,卷积神经网络模型训练与优化的过程是以人工辅助,机器为主的模式,这种模式相比以前人工阅读数据编写正则表达式的方式,大大降低了人工成本;另外,通过卷积神经网络模型一次计算就可确定城市事件类型,相比以前系统逐条计算正则表达式判断城市事件类型的方法,在系统运行的性能方面有较大提升。此外,正则表达式逐条判断事件类型,性能存在不确定性,而卷积神经网络模型的计算性能是基本确定的。

[0094] 步骤S3、计算城市事件发生地的标准地理编码。

[0095] 所述步骤S3具体包括下述步骤:

[0096] 判断城市事件文本数据是否带有地理坐标信息,其中移动端上报的城市事件含GPS坐标信息,电话热线或网站上报的城市事件不含坐标信息,如果带有地理坐标信息,记为(x_i,y_i)(i=1,2,...,n),则通过计算地理坐标信息与标准地名地址库中的行政区划的空间关系,判断城市事件属于哪个行政区划,进而取得行政区划对应的标准地理编码g_i;如果不带有地理坐标信息,首先根据步骤S1.2中的地名地址字典提取城市事件中的详细地址,

将城市事件中的详细地址输入到标准地名地址库中做文本匹配，选取匹配最高的标准地址编码g_i。

[0097] 步骤S4、构建城市事件派发卷积神经网络模型。

[0098] 所述步骤S4中派发卷积神经网络模型为内嵌词向量训练的卷积神经网络分类模型，具体分为以下部分：

[0099] 输入层：输入步骤S2中计算得到的预测分类类别、步骤S3中提取的城市事件标准地址编码及步骤S1中处理部门的数字化编码组成的样本数据，那么样本数据为：由城市事件类别和标准地理编码$\{e_1,e_2,...e_n,g_1,g_2...,g_n\}$以及所属部门编码$\{o_1,o_2,...,o_n\}$组成的样本数据；

[0100] 特征提取层：构建3种过滤器，每种过滤器128个，每个过滤器包含卷积层和池化层，并在两层中间加入非线性激活函数处理，3种过滤器将同步对样本数据进行卷积和池化操作；

[0101] 全连接层：连接所有样本数据经过过滤器池化后的特征值；

[0102] 激活层：选用softmax函数归一化计算样本数据属于每个处理部门的概率；

[0103] 输出层：输出概率最大的处理部门作为派发部门。

[0104] 在本方案中，城市事件派发模型也是一个卷积神经网络分类模型，其结构与骤S2中分类卷积神经网络模型为内嵌词向量训练的卷积神经网络分类模型相比，少了词向量训练的嵌入层，同时输入层是步骤S2通过事件分类卷积神经网络模型计算得到的城市事件预测分类类别、步骤S3提取的事件发生地标准地理编码及步骤S1中处理部门的数字化编码组成的样本数据，输出层是处理部门。

[0105] 步骤S5，接收当前输入的城市事件数据，调用所述分类卷积神经网络模型输出分类类别，获取当前城市事件数据的标准地理编码，然后调用派发卷积神经网络模型，输出具体的城市事件派发部门。

[0106] 前述步骤S1-S4建立完成各模型后，针对每次接收的城市事件数据，调用这些模型进行处理，最终输出具体的城市事件派发部门。本发明可以提高事件的分类正确性和派送准确性；通过卷积神经网络模型派发，相比人工派发带来的各种不确定性，机器派发的准确性更高，本发明中机器根据模型一次运算得到结果，可有效提升系统运行效率；采用机器学习的方法，让机器持续训练更新模型，人工成本低。

[0107] 实施例二：

[0108] 如图3所示，本发明提供一种基于深度学习的城市事件自动分类派发装置，用于完成本发明提供的基于深度学习的城市事件自动分类派发方法，所述基于深度学习的城市事件自动分类派发装置包括：

[0109] 数据处理模块：用于收集城市事件数据并预处理；

[0110] 事件分类模型构建模块：用于将收集处理的城市事件数据构建事件分类卷积神经网络模型；

[0111] 地理坐标信息模块：用于计算城市事件发生地的标准地理编码；

[0112] 事件派发模型构建模块：用于构建城市事件派发卷积神经网络模型；

[0113] 事件输出模块：用于接收当前输入的城市事件数据，调用所述分类卷积神经网络模型输出分类类别，获取当前城市事件数据的标准地理编码，然后调用派发卷积神经网络

模型,输出具体的城市事件派发部门。

[0114] 所述数据处理模块包括:

[0115] 城市事件收集单元:用于收集历年城市事件文本数据,结合常用的分词字典,对收集的文本数据进行分词过滤,得到适合于城市事件的分词字典;

[0116] 城市地址制作单元:用于获取城市的标准地名地址库,每个标准地名地址都包含唯一的地理编码,并且基于标准地名地址库制作地名地址字典;

[0117] 停用词字典建立单元:用于根据所述分词字典和地名地址字典对所有城市事件文本数据进行分词,并统计词频,选取词频较低的词作为停用词,建立适用于城市事件的停用词字典;

[0118] 城市事件编码单元:用于对所有的城市事件的类别按照名称进行编码,每个类别对应一个类别编号,并且增加一个其他的类别作为预留类别;

[0119] 样本建立单元:用于获取最终单词数量,进而作为训练样本,统计样本数量;

[0120] 对应关系词表建立单元:用于对所有训练样本中的所有单词,每个单词赋予一个唯一词编码,其中填充的空值UNK统一用一个词编码,并建立对应关系词表;

[0121] 处理部门编码单元:用于对城市事件的处理部门进行数字化编码。

[0122] 所述事件分类模型构建模块包括:

[0123] 第一输入单元:用于输入一个城市事件的训练样本;

[0124] 嵌入单元,用于对于输入层中的训练样本进行查表操作,从对应关系词表中查出对应的词编码,并将单词转为词向量形式;

[0125] 第一特征提取单元:用于构建3种过滤器,每种过滤器128个,每个过滤器包含卷积层和池化层,并在两层中间加入非线性激活函数,3种过滤器将同步对词向量进行卷积和池化操作;

[0126] 第一全连接单元:用于连接所有词向量经过过滤器池化后的特征值;

[0127] 第一激活单元:用于选用softmax函数归一化计算城市事件属于每个类别的概率;

[0128] 第一输出单元:用于输出概率最大的类别作为预测分类类别。

[0129] 所述事件派发模型构建模块包括:

[0130] 第二输入单元:用于输入由计算得到的预测分类类别、提取的城市事件标准地址编码及处理部门的数字化编码组成的样本数据;

[0131] 第二特征提取单元:用于构建3种过滤器,每种过滤器128个,每个过滤器包含卷积层和池化层,并在两层中间加入非线性激活函数,3种过滤器将同步对样本数据进行卷积和池化操作;

[0132] 第二全连接单元:用于连接所有样本数据经过过滤器池化后的特征值;

[0133] 第二激活单元:用于选用softmax函数归一化计算样本数据属于每个处理部门的概率;

[0134] 第二输出单元:用于将输出概率最大的处理部门作为派发部门。

[0135] 本实施例提供的各个功能模块及单元对应实现实施例一中的步骤S1-S5,具体实现过程这里不再赘述。

[0136] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

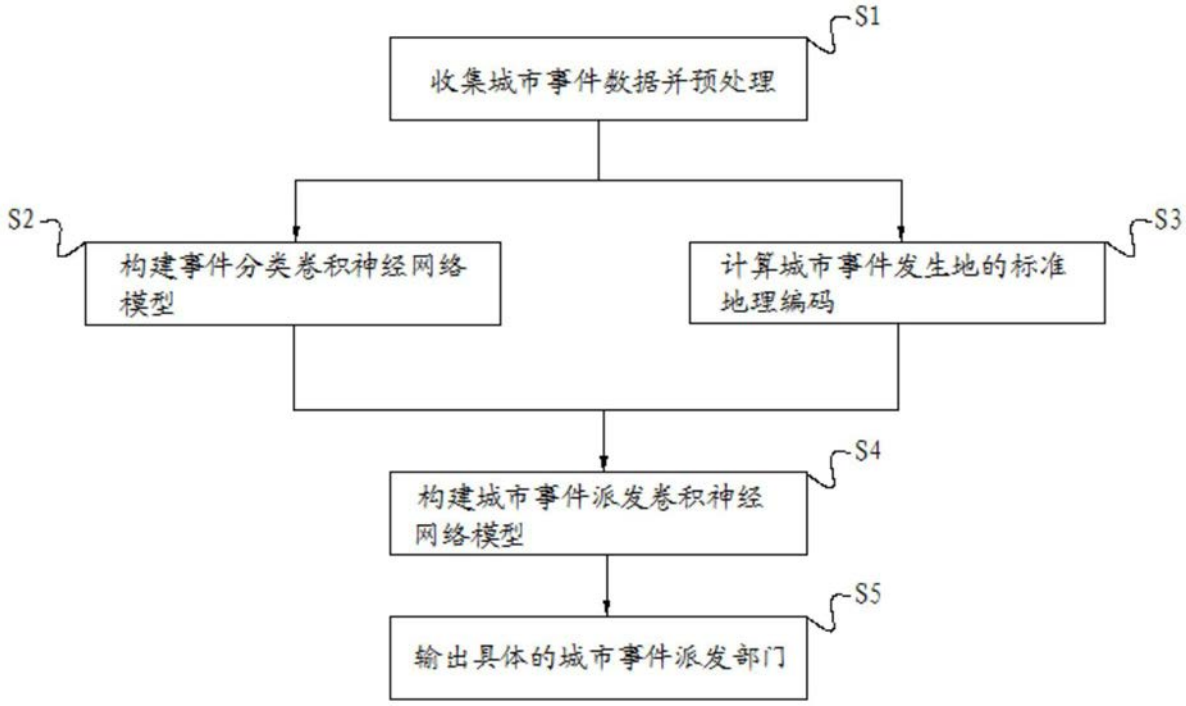

S1
收集城市事件数据并预处理
S2
构建事件分类卷积神经网络模型
S3
计算城市事件发生地的标准地理编码
S4
构建城市事件派发卷积神经网络模型
S5
输出具体的城市事件派发部门

图1

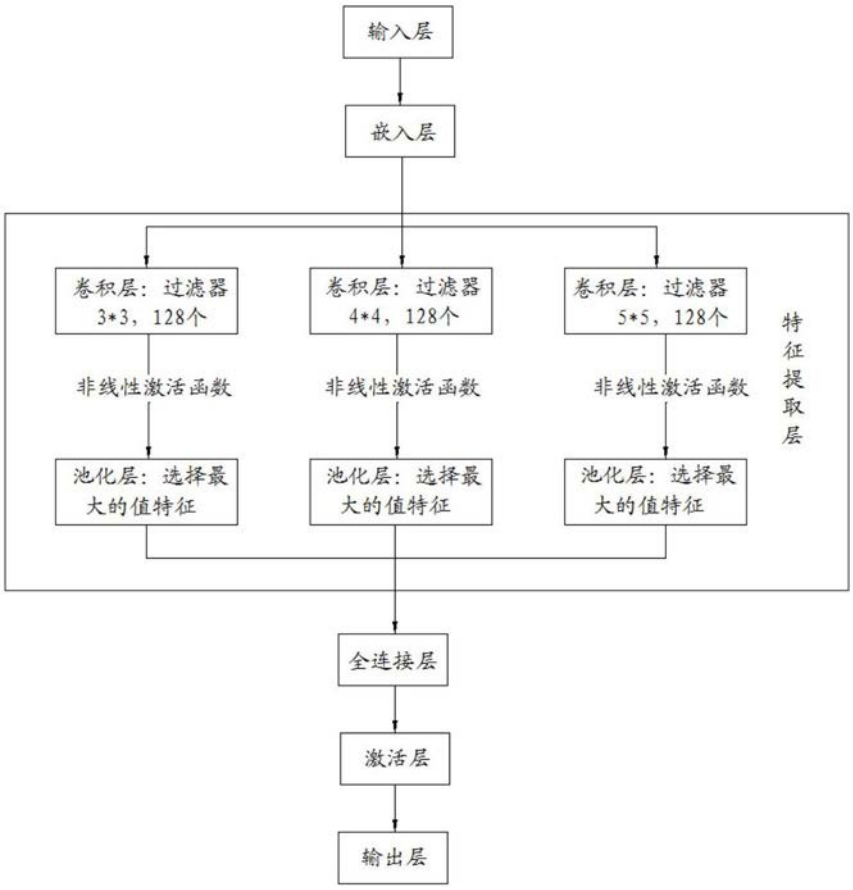

输入层
嵌入层
卷积层：过滤器 3*3，128个
卷积层：过滤器 4*4，128个
卷积层：过滤器 5*5，128个
非线性激活函数
非线性激活函数
非线性激活函数
池化层：选择最大的值特征
池化层：选择最大的值特征
池化层：选择最大的值特征
特征提取层
全连接层
激活层
输出层

图2

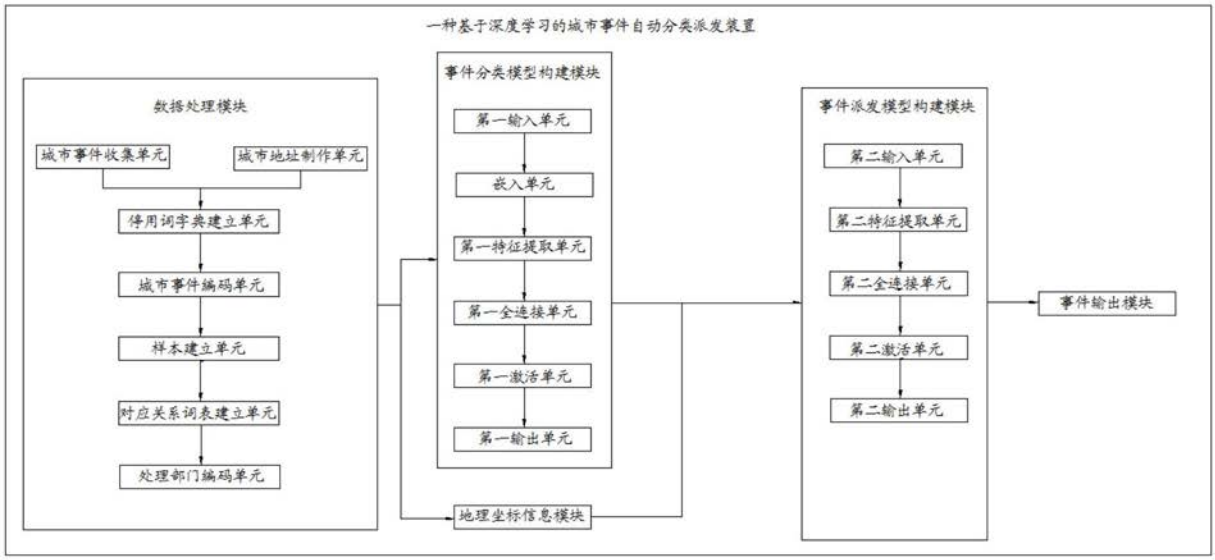

一种基于深度学习的城市事件自动分类派发装置
数据处理模块
城市事件收集单元
城市地址制作单元
停用词字典建立单元
城市事件编码单元
样本建立单元
对应关系词表建立单元
处理部门编码单元
事件分类模型构建模块
第一输入单元
嵌入单元
第一特征提取单元
第一全连接单元
第一激活单元
第一输出单元
地理坐标信息模块
事件派发模型构建模块
第二输入单元
第二特征提取单元
第二全连接单元
第二激活单元
第二输出单元
事件输出模块

图3