

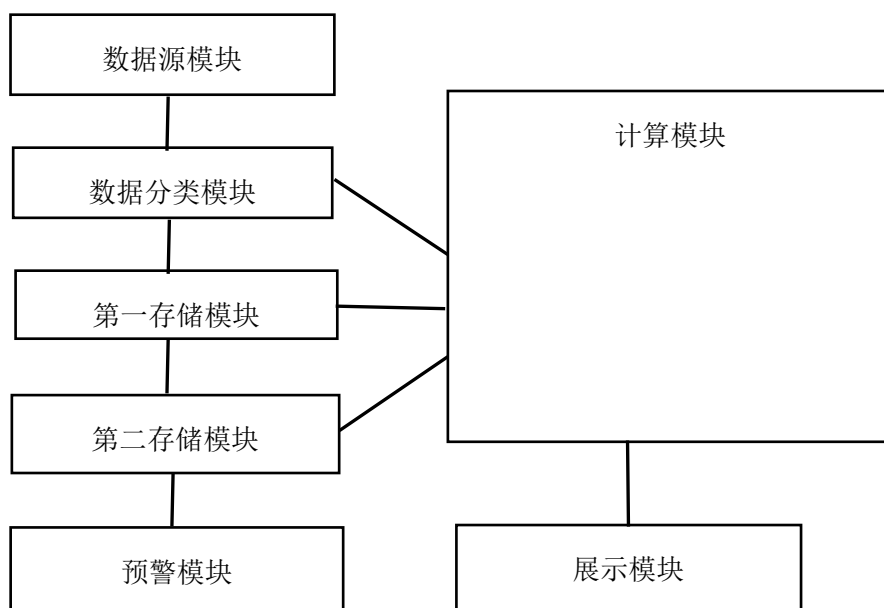
## 说明书摘要

---

一种基于城市大数据存储的检索系统，包括：数据源模块，用于获取城市数据；数据分类模块，用于对所述城市海量数据进行分类；计算模块，用于对所述数据分类模块传递的结构化数据进行计算，记经所述计算模块计算后的结构化数据为第一计算数据；第一数据存储模块，用于对所述数据分类模块传递的半结构化和非结构化数据进行存储，存储后的数据记为第二计算数据；第二数据存储模块，用于对所述第一数据存储模块传递的第二计算数据与所述计算模块传递的第一计算数据进行存储，记存储后的数据为第三计算数据；预警模块，用于根据所述第三计算数据生成模型并输出预警结果；展示模块，用于对所述预警结果进行展示；通过上述城市大数据存储检索系统的结构，解决了现有城市大数据平台的不适配于城市数据海量、实时性强及数据多样的问题。

## 摘要附图

---



# 权 利 要 求 书

---

1、一种基于城市大数据存储的检索系统，其特征在于，所述系统包括：  
数据源模块，用于获取城市数据；

数据分类模块，用于对所述城市海量数据进行分类；

计算模块，用于对所述数据分类模块传递的结构化数据进行计算，记经所述计算模块计算后的结构化数据为第一计算数据；

第一数据存储模块，用于对所述数据分类模块传递的半结构化和非结构化数据进行存储，存储后的数据记为第二计算数据；

第二数据存储模块，用于对所述第一数据存储模块传递的第二计算数据与所述计算模块传递的第一计算数据进行存储，记存储后的数据为第三计算数据；

预警模块，用于根据所述第三计算数据生成模型并输出预警结果；

展示模块，用于对所述预警结果进行展示；

所述数据源模块通过数据接入模块与所述数据分类模块连接，所述第一存储模块、所述第二存储模块与所述数据分类模块分别与所述计算模块连接，所述第二存储模块与所述预警模块连接，所述预警模块与所述展示模块连接。

2、根据权利要求1所述的一种基于城市大数据存储的检索系统，所述数据分类模块，用于对所述城市海量数据进行分类，其特征在于，所述数据分类模块为数据适配器。

3、根据权利要求1所述的一种基于城市大数据存储的检索系统，所述第一数据存储模块，用于对所述数据分类模块传递的半结构化和非结构化数据进行存储，存储后的数据记为第二计算数据，其特征在于，所述第一存储模块为hive数据仓库。

4、根据权利要求1所述的一种基于城市大数据存储的检索系统，第二数据存储模块，用于对所述第一数据存储模块传递的第二计算数据与所述计算模块传递的第一计算数据进行存储，记存储后的数据为第三计算数据，其特征在于，所述第二存储模块为MergeTree存储引擎。

5、根据权利要求1所述的一种基于城市大数据存储的检索系统，其特征在于，所述数据数据分类模块通过kafka集群与所述计算模块连接。

## 权 利 要 求 书

---

6、根据权利要求 1 所述的一种基于城市大数据存储的检索系统，其特征在于，所述计算模块包含第一处理单元与第二处理单元，所述第一处理单元与第二处理单元连接。

7、根据权利要求 6 所述的计算模块包含第一处理单元与第二处理单元，所述第一处理单元与第二处理单元连接，其特征在于，所述第一处理单元为 flink 单元，所述第二处理单元为 K8S 单元，所述 flink 单元与所述 K8S 单元连接。

8、根据权利要求 1 所述的一种基于城市大数据存储的检索系统，所述预警模块，其特征在于，所述预警模块为 GRU 模型。

# 说明书

---

## 一种基于城市大数据存储的检索系统

### 技术领域

本发明涉及大数据平台领域，尤其涉及一种基于城市大数据存储的检索系统。

### 背景技术

城市运行数据具有数据海量，数据实时性强以及数据类型多样的特点；现有的城市大数据存储系统中多采用现有的固定架构，对数据进行计算和存储，无法完全满足对城市海量数据进行快速存储和快速检索的需求；如何调整现有城市大数据平台的架构，更加适配于城市数据海量、实时性强及数据多样的特点，是亟待解决的问题。

### 发明内容

本发明的目的是提供一种基于城市大数据存储的检索系统，该系统通过重组城市数据存储系统的结构，解决当前城市数据存储系统对于城市数据海量，数据实时性强以及数据类型多样特点不适配的问题。

为实现上述目的，本发明提供了如下方案：

一种基于城市大数据存储的检索系统，系统包括：

数据源模块，用于获取城市数据；

数据分类模块，用于对所述城市海量数据进行分类；

计算模块，用于对所述数据分类模块传递的结构化数据进行计算，记经所述计算模块计算后的结构化数据为第一计算数据；

第一数据存储模块，用于对所述数据分类模块传递的半结构化和非结构化数据进行存储，存储后的数据记为第二计算数据；

第二数据存储模块，用于对所述第一数据存储模块传递的第二计算数据与所述计算模块传递的第一计算数据进行存储，记存储后的数据为第三计算数据；

预警模块，用于根据所述第三计算数据生成模型并输出预警结果；

# 说明书

---

展示模块，用于对所述预警结果进行展示；

所述数据源模块通过数据接入模块与所述数据分类模块连接，所述第一存储模块、所述第二存储模块与所述数据分类模块分别与所述计算模块连接，所述第二存储模块与所述预警模块连接，所述预警模块与所述展示模块连接。

可选的，数据分类模块，用于对所述城市海量数据进行分类，其特征在于，所述数据分类模块为数据适配器。

可选的，所述第一数据存储模块，用于对所述数据分类模块传递的半结构化和非结构化数据进行存储，存储后的数据记为第二计算数据，其特征在于，所述第一存储模块为 hive 数据仓库。

可选的，第二数据存储模块，用于对所述第一数据存储模块传递的第二计算数据与所述计算模块传递的第一计算数据进行存储，记存储后的数据为第三计算数据，其特征在于，所述第二存储模块为 MergeTree 存储引擎。

可选的，所述数据数据分类模块通过 kafka 集群与所述计算模块连接。

可选的，所述计算模块包含第一处理单元与第二处理单元，所述第一处理单元与第二处理单元连接。

可选的，所述第一处理单元与第二处理单元连接，其特征在于，所述第一处理单元为 flink 单元，所述第二处理单元为 K8S 单元，所述 flink 单元与所述 K8S 单元连接。

可选的，所述预警模块为 GRU 模型。

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：本发明提供了一种基于城市大数据存储的检索系统，通过数据分类模块，对城市数据类型进行分类，对于分类后的结果进行分别存储和融合，并通过 GRU 模型进行预警和推荐，解决了现有城市大数据平台的不适配于城市数据海量、实时性强及数据多样的问题。

## 附图说明

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性的前提下，还可以根据这些附图获得其他的附图。

# 说明书

---

图 1 为本发明系统结构示意图。

## 具体实施方式

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。

本发明的目的是提供一种方法及系统，

图 1 为本发明本发明系统结构示意图，如图 1 所示，本发明方法系统结构具体如下：

所述系统包括：

数据源模块，用于获取城市数据，数据源模块为开放数据源模块、爬虫抓取模块、传感器模块和日志采集模块中的一种或几种的线性组合；

数据分类模块，用于对所述城市海量数据进行分类；

所述数据分类模块获取待分类结构化数据；将所述待分类结构化数据输入预先训练得到的卷积神经网络模型，得到所述待分类结构化数据的分类结果，其中，所述卷积神经网络模型包括全连接层及卷积神经子网络，所述全连接层为所述卷积神经网络模型中的第一层；获取结构化数据样本集，所述结构化数据样本集中包含多个结构化数据样本；针对每个结构化数据样本，进行如下操作：将该结构化数据样本输入所述全连接层，得到第一过渡样本；将所述第一过渡样本输入所述一维卷积神经子网络，得到神经网络输出；基于所述神经网络输出和所述该结构化数据样本间的差异，对所述全连接层和所述一维卷积神经子网络进行训练，得到所述卷积神经网络模型；

针对半结构化数据，将 TSGrans 特征作为表示半结构化数据的基本单元，用标签序列捕捉半结构化数据的结构信息，用 nGrans 捕捉半结构化数据的内容信息，将二者融合作为特征捕捉结构和内容间的包含关系，同时考虑内容信息中不同关键字间的相互关系，使用信息增益对 TSGrans 特征进行筛选，获取具有较强的分类能力的 TSGrans 特征构造特征空间，并根据 TSGrans 特征与类

## 说明书

---

别间的互信息构建类别分类模型；

所述数据分类模块，针对非结构化数据，获取若干非结构化数据对象，并将所述非结构化数据对象的特征抽象为属性；使用所述非结构化数据对象的所有属性对应的多维向量表示所述非结构化数据对象；将所述多维向量作为卷积神经网络输入的基本单元；通过卷积神经网络的卷积层学习所述训练数据的局部属性；通过卷积神经网络的池化层将所述局部属性进行统计操作，获得第二特征向量；将所述第二特征向量输入所述卷积神经网络的全连接层，利用分类器获得非结构化数据分类结果。

计算模块，用于对所述数据分类模块传递的结构化数据进行计算，记经所述计算模块计算后的结构化数据为第一计算数据；

第一数据存储模块，用于对所述数据分类模块传递的半结构化和非结构化数据进行存储，存储后的数据记为第二计算数据；

第二数据存储模块，用于对所述第一数据存储模块传递的第二计算数据与所述计算模块传递的第一计算数据进行存储，记存储后的数据为第三计算数据；

预警模块，用于根据所述第三计算数据生成模型并输出预警结果；

展示模块，用于对所述预警结果进行展示；

所述数据源模块通过数据接入模块与所述数据分类模块连接，所述第一存储模块、所述第二存储模块与所述数据分类模块分别与所述计算模块连接，所述第二存储模块与所述预警模块连接，所述预警模块与所述展示模块连接。所述数据分类模块，具体用于对所述城市海量数据进行分类，所述数据分类模块为数据适配器。

所述第一数据存储模块，具体用于对所述数据分类模块传递的半结构化和非结构化数据进行存储，存储后的数据记为第二计算数据，所述第一存储模块具体为 hive 数据仓库。

第二数据存储模块，具体用于对所述第一数据存储模块传递的第二计算数据与所述计算模块传递的第一计算数据进行存储，记存储后的数据为第三计算数据，其特征在于，所述第二存储模块为 MergeTree 存储引擎。

所述数据数据分类模块通过 kafka 集群与所述计算模块连接。



## 说明书

---

所述计算模块包含第一处理单元与第二处理单元,所述第一处理单元与第二处理单元连接。

所述第一处理单元为 flink 单元,所述第二处理单元为 K8S 单元,所述 flink 单元与所述 K8S 单元连接。

所述预警模块,所述预警模块具体为 GRU 模型。

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。

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。

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。

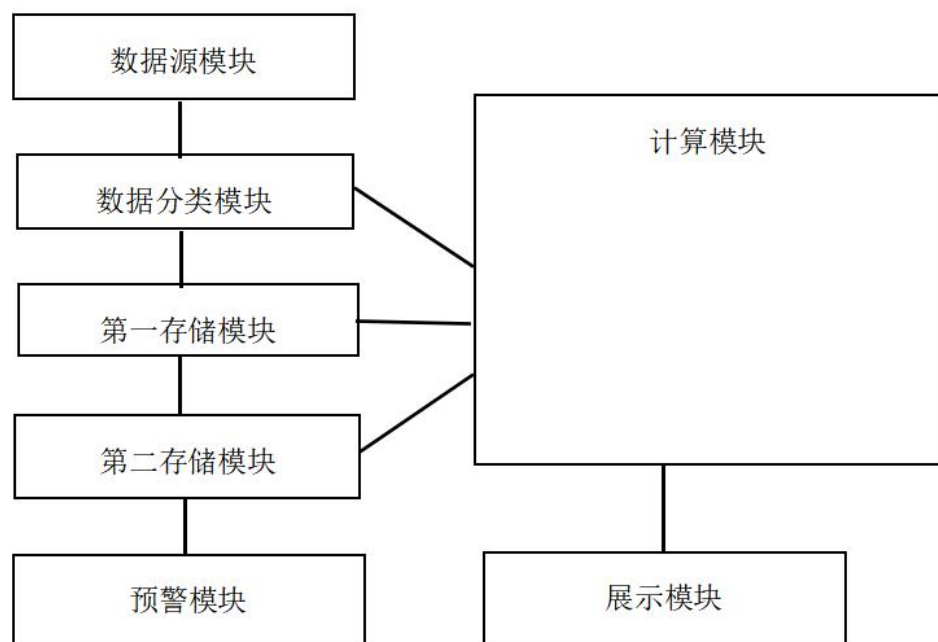


图 1