一种大规模时空数据处理与可视化平台

杜一[1] 郭旦怀[1] 周园春[1] 黎建辉[1]

[1]中国科学院计算机网络信息中心 科学数据中心, 北京 中国 100190

(guodanhuai@cnic.cn)

A Data Processing and Visualization Platform for Large-scale Spatio-temporal Data

DU Yi[1], GUO Dan-Huai[1], ZHOU Yuan-Chun[1], and LI Jian-Hui[1]

[1](Scientific Data Center, Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

Abstract Most of the existing spatio-temporal information visualization toolkits cannot process and visualize data efficiently when the size of it is too large. In this paper, we present a visual analytics platform that supports large-scale spatio-temporal data. By redefining task model, data model, and visual mapping strategies, this platform supports processing and visualizing many kinds of Big Data with spatio-temporal attributes. The processing and visualizing can be done in seconds by distributed storage, data reorganization, distributed query, spatial indices, and segmented fetch, even though it has a terabyte of data.

Key words Big Data; spatio-temporal data processing; spatio-temporal visualization; software architecture; model driven architecture

摘要 当前大多数时空数据处理与可视化工具在数据规模增大时，不能够对数据进行快速的处理与可视化。为解决该问题，本文通过对任务模型、数据模型及可视映射策略的重新定义，给出一种大规模时空数据处理与可视化平台。平台能够支持多种不同类型的时空数据，通过分布式的数据存储、数据重新组织、分布式检索、空间索引、分段预取等技术，能够实现大规模数据的快速处理与可视化。

关键词 大数据;时空数据处理;时空数据可视化;软件架构;模型驱动的架构

中图法分类号 TP391

我们处在一个信息爆炸的时代[1-5]。新的数据不断产生，而且产生的速度越来越快，规模越来越大[3]。出现这种现象的原因可总结为两个方面，第一，用于收集数据的软硬件成本逐渐降低；第二，人们对使用收集后的数据进行分析与理解的需求更加迫切。以地理学为例，在多年前政府、企业和一些研究机构已经开始收集各种相关数据。在这些数据集中，带有时间与空间标签的时空数据是非常大的一类。根据数据密集的类型[6]，可以将时空数据分为时间密集型、空间密集型及事件密集型。为了在大规模时空数据中发现更多的模式，研究者开始在 GIS 系统中使用信息可视化的技术[7]。出现了很多可视分析软件[8-14]及可视化组件。然而，彼时不少软件与工具运行在主机系统上，随着大数据时代的到来，其处理能力已不能满足需求。如果不能快速的实现从数据到可视化结果的转换，可视分析软件与工具将失去意义。

在本文中，我们给出了一个面向大规模时空数据的数据处理与可视化平台。该平台部署在 Web 环境下，并且定义了合理的预处理流程及数据模型。在第二节中，文章总结了时空数据可视化的任务模型及相关软件；在第三节中，给出一个自定义的任务模型，作为后续工作的理论基础；在第四节与第五节中，描

收稿日期：
基金项目：本课题得到国家自然科学基金(No.61402435，No. 91224006，No.41371386)；“十二五”国家科技支撑计划 (No. 2012BAK17B01-1)；大数据应用服务技术北京市工程实验室创新能力建设项目(No.Y421021108)；中国科学院十二五信息化专项(XXH12504)资助.

述了该平台的细节。最后，给出文章的结论与下一步工作。

本文的贡献主要在两个方面：第一，总结了当前时空数据分析的任务模型，并给出了改进后的任务模型。该模型从两个维度，对时空数据分析时的任务进行了详细的定义。第二，给出了时空数据处理与可视化平台，该平台能够支持 web 环境下的大规模时空数据的处理与可视化。并且在 TB 级别的数据上对平台进行了验证。

1. 相关研究

时空数据分析有区别于其它数据分析的特点。当前时空数据分析的任务模型分为两类，分别是需求驱动的任务模型以及数据驱动的任务模型。需求驱动的任务模型关注时空数据分析的需求，该模型假设所有分析师在进行时空数据分析时有明确的初始目标。数据驱动的模型以被分析的数据的多个属性为核心进行分析，根据属性的不同，将数据分成不同的种类，此类分析模型更倾向在数据中发现各种未知的模式或异常。Peuquet [6]将时空数据属性进行了分析，将时空数据属性分成了三个不同的类型，分别是地点(where)、时间(when)以及事件(what)。通过该分类方式，Peuquet 给出三个基本的任务，分别是根据给定的时间与地点对事件进行描述、根据给定的时间与事件对地点进行描述以及根据给定的地点与事件对时间进行描述。以此为基础，Block[23]对该任务模型添加了用户需求的部分，通过将用户探索任务考虑在内。例如，如果给定时间点或时间段，如何发现事件及事件的地点。研究人员[25]根据不同的数据类型，将时空数据分析的任务分成了三类，分别是基于点的查询(P-Query)[26]、基于区域的查询(R-Query)[27]以及基于轨迹的查询(T-Query)[28]。相当多的任务模型将数据和用户需求考虑在内。但是，时空数据本身发生了很大的变化。一方面，原始的带有时空标签的数据附加了更多不同类型的属性；另一方面，时空数据的规模变得很大。基于以上两个原因，需要对当前的模型进行改进与优化，使其能够适应当前的大规模时空数据环境。

在时空数据可视化与可视分析应用系统方面，Compieta[30]开发了一款时空数据挖掘系统，该系统能够给领域专家及数据挖掘专家呈现不同的视图，以辅助不同角色的使用者使用系统，并且使用 Isablel 飓风作为测试数据对系统进行了测试。ESV(EarthSystemsVisualizer)[31]是一款辅助分析全球气候变化的可视化工具，该工具使用了一些新颖的交互技术，例如时间刷(temporal brushing)以及事件聚焦(temporal focusing)，来与时间属性进行交互。除此之外，还有一些基于时空数据的可视分析系统，如 Andrienko[33]通过提出一种可视分析流程，对人的迁徙进行了分析；Tominski[34]使用 2D 与 3D 技术相结合，来对出租车运行轨迹进行了可视化；Janoos[34]与 Liao[11]开发的可视分析系统，都利用了机器学习的方法来发现异常事件；

当前大多数基于时空数据的可视化系统，均将多视图、可视化形态等因素考虑在系统内，很好的体现了可视化的优势。但是，目前大多数可视化系统的开发均以某一特定的数据集及应用为基础进行开发，这种开发方式，在目前的数据集增多以及新交互设备层出不穷的环境下，会造成开发成本的提升以及开发效率的下降。

2 时空可视分析任务模型

本文从面向数据与面向操作两个维度给出时空可视分析的任务模型。其中，面向数据的维度根据数据本身的特性进行划分，该维度的划分基于原有的数据驱动的模型，并且能够对新的数据集进行适配。面向操作的维度根据数据处理引擎的能力进行划分，该维度的划分能够适配新数据集下的新任务。

2.1 面向数据的模型

一般来说，我们定义的所有的数据均包含位置信息，大多数的数据包含时间戳信息。对于每一条数据，我们将其划分为四个部分：

- 位置或位置集合(where)
- 时间点或时间集合(when)
- 对象或对象集合(who)
- 属性集合(what)

以 2012 年 1 月至 8 月中国某城市的出租车 GPS 数据为例，每条数据集可表示为：

{ taxiId, latitude, longitude, time, hasPassenger, orientation, speed}

以该数据为例，每条数据均有 latitude 与 longitude 两个属性标明位置，time 属性作为时间戳标明了时间点，一个字符串或一串数字来唯一标识该对象(出租车)，以及一系列属性来标识该条数据的特征(速度、方向等)。

通过分析时空数据的四个组成部分，将四个部分根据分析层级划分为四层。第一层使用三个部分来分析其中的一个部分，第二层使用两个部分来分析另外两个部分，依次类推。表一给出了四个层级的任务的描述。

表 1 任务的四个层级

Level	Tasks contained in each task level
level one	where + when + who → what
	where + when + what → who
	where + who + what → when
	who + what + when → where
level two	where + when → what + who
	what + when → who + where
	where + what → who + when
	when + who → what + where
	where + who → when + what
	who + what → where + when
level three	where → what + when + who
	what → who + when + where
	when → where + who + what
	who → where + what + when
level four	NULL → what + where + when + who

第一层模型包括四个分析任务：

1.根据给定的位置或位置集合、时间点或时间集合、以及对象或对象集合，发现事件的发生，以及发生情况。例如，发现汽车 O 在一天中处在位置 L 时的车速，或发现汽车 O 在这个城市中的时间 T1 到 T2 之间是否有乘客。在第一个例子中，“位置 L”表示“where”，“一天中”表示“when”，“汽车 O”表示“who”，而“车速”表示“what”。在第二个例子中，“这个城市”表示“where”，“时间 T1 到 T2”表示“when”，“汽车 O”表示“who”，而“是否有乘客”则表示“what”。

2.根据给定的位置或位置集合、时间点或时间集合、以及属性集合，找到相关的对象或对象集合。例如，找到在位置 L 与时间 T 时载有乘客的出租车，或找到在这个城市中从 T1 到 T2 时间中车速超过 S 的出租车。

3.根据给定的位置或位置集合、对象或对象集合以及属性结合，发现关注的时间点或时间集合。例如，在所有出租车中，发现在位置 L 车速为 0，并且没有乘客的时间范围。

4.根据给定的对象或对象集合、时间或时间集合、以及属性集合，发现相关的位置或位置集合。例如，在所有出租车中，找到在 T1 到 T2 时间段中没有乘客的区域 L。

第二层模型包括六个分析任务：

1.根据给定的位置或位置集合、以及时间点或时间集合，找到相关的对象与属性集合。例如，找到在时间 T、位置 L 的所有的车辆 O 以及它们的各种状态 A。在该例中，“位置 L”表示“where”，“时间 T”表示“when”，“车辆 O”表示“who”，而“状态 A”表示“what”。

2.根据给定的时间点或时间集合、以及属性集合，找到相关的对象集合与位置集合。例如，找到在时间 T 中所有载客的车辆 O 以及它们的位置。

3.根据给定的位置或位置集合、以及对象或对象集合，找到对象的属性集合以及相关的时间点或时间集合。

4.根据给定的时间或时间集合、以及对象或对象集合，找到对象的属性集合以及相关的位置或位置集合。

5.根据给定的位置或位置集合、以及属性集合，找到相关的对象或对象集合及相关的时间点或时间集合。

6.根据给定的对象或对象集合、以及属性或属性集合，找到相关的位置或位置集合及相关的时间点或时间集合。

第三层模型包括四个分析任务：

1.根据给定的对象或对象集合、时间点或时间集合、以及属性或属性集合，找到相关的位置或位置集合。例如，找到在时间 T 中车速超过 S，并且载有乘客的车辆 O 的位置 L。在本例中，“位置 L”表示“where”，“车辆 O”表示“who”，“车速超过 S，并且载有乘客”表示“what”，“时间 T”表示“where”。

2.根据给定的对象或对象集合、时间点或时间集合、以及位置或位置集合，找到相关的属性集合。

3.根据给定的对象或对象集合、位置或位置集合、以及属性或属性集合，找到相关的时间点或时间集合。

4.根据给定的位置或位置集合、属性或属性集合、以及时间点或时间集合，找到相关的对象或对象集合。

第四层模型只包含一项任务，与其它三层的模型不同，它不需要输入约束即进行分析。在本层模型下，经常用于发现异常的、未知的模式。它表示在不需要其它约束的情况下，找到感兴趣的对象或对象集合、属性集合、时间点或时间集合、以及位置或位置集合。本层模型经常会有大规模的数据的输入与输出。

总体来说，随着任务从第一层变化到第四层，更多的统计与分析工作需要进行处理。分析过程中的不确定性增加，更有利于分析师发现数据背后的异常。

2.2 面向操作的模型

由于数据规模的增加，不同操作所花费的时间出现了很大的差别。根据计算所需的复杂程度对任务模

型进行建模，将任务分成三个部分，分别是简单查询、统计以及预测。

简单查询的目标是从存储中直接检索到相关的数据。这种操作一般需要较少的计算量，并且该类数据一般较容易被理解。以出租车 GPS 数据为例，一个简单的任务是找到感兴趣的点(points of interest)，这些点能够满足某些特定的时空约束。除了该任务，我们给出简单查询中其它经典的任务描述：

- 根据给定的时空数据，找到轨迹片段。
- 根据指定的区域，找到时空数据。
- 根据给定的区域，找到特定的轨迹。
- 根据给定的约束，找到相关的对象。

统计的目标是通过一些计算、测量等统计方式，来找到相关的统计结果。以出租车 GPS 数据为例，统计特定时间段出租车的流量等。我们给出该类目标中的经典的任务描述：

- 给定检索结果数据集，计算结果中数据的条数。
- 给定特定区域的检索结果数据集，计算一个特定属性的平均值。
- 给定特定时间的检索结果数据集，计算一个特定属性的最大或最小值。
- 给定检索结果数据集，计算根据不同时空属性的分类结果。

预测是高级别的任务。预测的目标是利用给定的数据来找到潜在的事情或属性。在时空分析的很多领域中，预测都起到了非常重要的作用。例如，我们利用给定的出租车 GPS 数据，预测接下来的一个小时或一天中的交通情况。我们也可以利用给定的突发事件数据，预测指定区域的潜在的拥堵事件。我们给出该部分的经典任务描述：

- 利用给定的时间密集型统计数据，预测特定的属性情况。
- 利用给定的非常规数据，预测潜在的事件发生。
- 利用给定的特定空间区域的数据，预测特定的属性情况。

在该维度中，三个不同层次的任务需要不同的计算资源。在第一层次的简单查询中，在大数据环境下，对性能的约束在存储与索引。该层次需要好的存储机制及定义良好的索引来保证性能。在第三层次的预测中，大数据环境下的约束在模型与算法。为保证任务的高效执行，需要设计良好的预测模型与算法。而第二层的统计需要两者的结合来完成。为了满足多层次的任务，需要优化时空数据环境下的存储与索引，并且给出如何使用好的预测模型的解决方案。

3 平台架构

本节给出平台的架构如图 1。平台由四个部分组成，分别是原始数据层、预处理与数据转换成、操作层、显示层。

3.1 原始数据层

原始数据通过各种不同的传感器收集，由于收集的数据的特性、开发人员的知识等原因，原始数据存在较大差异。第一，原始数据以不同的数据格式进行存储。例如，大多数传统的 GPS 传感器收集的数据以文本文件存储，而一些商业数据多保存在关系数据库、数据仓库中。即使两种数据集均以文件的形式存储，也会存在一种数据集是时间密集型，而另一种数据集是位置密集型。第二，原始数据以不同的粒度存储，由于数据获取的需求的变化，以及数据获取设备的约束，原始数据集的粒度差别很大。由于原始数据的差异性及复杂性，需要实现不同的数据访问方法来支持数据的提取。所有的数据提取方法用来支持数据预处理与数据转换。

3.2 预处理与数据转换层

在该部分中，操作流程分成了两个阶段，分别是预处理与数据转换。在第一阶段由三个部分组成，分别是数据分割、数据过滤以及数据生成。数据分割是将原始数据从某个维度，以一种合适的粒度对原始数据进行分割，来在大的粒度上解决数据访问效率的差异性。原始数据中，不可避免会出现错误数据，数据过滤则用来解决该问题。在数据分析与可视化的过程中，有时会对不同粒度的数据进行分析，而数据生成则是用来解决该问题的。它通过传统的差值算法来生成新的数据。该部分的第二阶段是数据转换。在原始数据预处理完成之后，平台将预处理后的数据转换成预先定义的数据格式。该格式充分将时空数据的特性考虑在内，并且使用 MongoDB 作为数据库存储数据。MongoDB 是一款面向文档的数据库，它支持传统数据库的检索、索引等功能。除此之外，它还支持空间索引与查询，这在时空数据分析中非常有意义。

3.3 操作层

根据计算中时间的消耗，将本层分为三个不同的部分。分别是查询、统计分析与预测。根据前述文章定义的任务模型可知，查询比预测消耗更小的时间与计算资源。第一部分是查询，它从数据库中检索到数据，并返回感兴趣的结果。以 Ke Deng[25]提出的分类为参考，我们将时空数据查询分成三类，分别是 P-Query，R-Query 与 T-Query。其中，P-Query 是查询满足特定约束的点集合，R-Query 是查询满足某些条件的特定的区域或形状的集合，T-Query 根据指定的时间、区域来查询特定的轨迹或路线。由于大多数

该类查询只需要直接从数据库中进行数据的检索，需要较少的计算量，大多数此类查询通过数据库管理系统完成。在本平台中，我们部署了一套分布式的 MongoDB 集群，能够高效的实现此类查询。

统计分析是第二个部分，它能够实现聚合、分类等统计分析功能。该部分比第一部分需要更多的计算工作。在一般情况下，统计分析是在查询结果的基础上进行的。例如，如果要找到人类的聚集区，或地球上的其它信息，需要使用某些聚类算法。首先，需要从数据库中检索到相关的数据，并将其转换成特定的格式，该格式能够作为算法的输入。第二，使用聚类算法来对数据进行聚类。为满足大数据下的统计分析的计算需求，平台利用了 R 语言作为统计分析的主要工具，并搭建了一个分布式的 R 语言计算环境。该部分的操作需要数据库的查询与平台的计算环境协作实现。

第三部分是预测，该部分是基于当前的数据集对将来的趋势或即将发生的事件进行预测。与第二部分相似，该部分也基于平台搭建的 R 语言分布式计算环境。与第二部分不同的是，该部分不需要太多的数据库查询作为基础。

4.4 显示层

缺乏好的展现与交互，即使能够在合理的时间处理，时空数据也很难被理解。为了使得平台有良好的可用性及可扩展性，在平台基础上添加了可视化部分，该部分作为操作层的主要组成部分。可视化是显示层的主要组成部分。在该部分中，使用了模型驱动的方法，根据需求对可视化呈现进行建模。在呈现时，根据时空数据的特点，给出了两种不同类型的呈现方式，一种是以基于地图或地球的呈现，该呈现方式符合时空数据环境的分析需求。它的核心部分是地球或地图的呈现。另一种方式是抽象数据的呈现方式。该方式使用传统的数据可视化组件进行呈现，如使用饼图、折线图、散点图等。与基于地图的呈现方式相比，抽象数据的呈现方式更利于呈现统计结果，如根据年份变化的平均值。为增加可用性，平台中添加了交互功能，如可使用拉锁在地图上进行所见即所得的选择等。

4 平台实现

为了验证平台，本文基于一款时空数据对该平台进行了实现。选取了中国某一线城市出租车持续 8 个月的 GPS 数据，在该城市中，约有 23000 辆出租车，数据从 2012 年 1 月 1 日至 2012 年 8 月 28 日。原始数据使用文本文件，按天存储，每个文件约 4000 万条记录。原始数据总体的规模是 800GB。

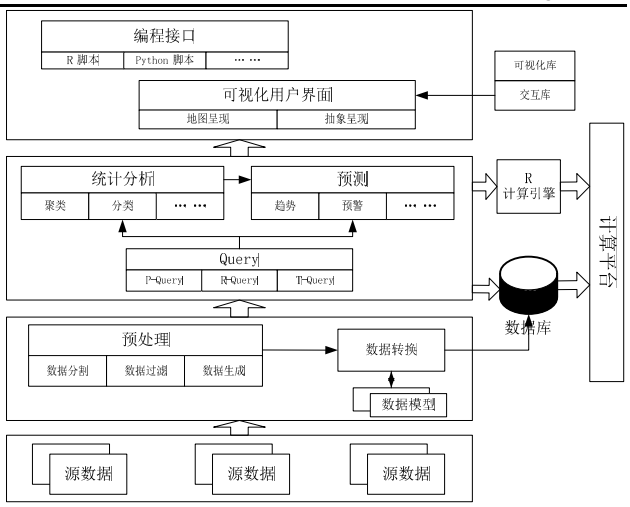

图 1 平台整体架构图。在本图中，实线箭头表示“使用”，宽箭头表示“数据转换”，圆柱表示数据库，矩形表示模块或子模块

系统使用 MongoDB 来对数据进行重新组织与存储，搭建了一个 5 个分片的 MongoDB 集群。在实现过程中，使用 J2EE 对系统原型进行建设。为了实现基于地图的可视化，利用了开放的地图组件如 Open Street Map 及 Google Map。在本节中，介绍五个平台实现的细节，分别是数据组织、分布式数据查询机制、分段数据预取、基于地图的可视呈现及基于统计的可视呈现。

4.1 数据组织

为满足对数据的高效的获取与访问，需要对原始数据进行重新组织。根据原始数据发现，出租车 GPS 数据可以看做时间密集型数据。利用已定义的任务模型，并考虑 MongoDB 的性能，系统选择时间维度作为主要维度对数据进行组织。在确定主要维度后，确定该维度的数据分割的粒度。为充分利用已完成的并行计算的环境，并保证数据获取的速度，选择小时作为时间维度的粒度进行数据的分割。除此之外，还在空间维度上编制了空间索引，并对其它属性维度也编制了索引，以保证访问的速度。在数据重新组织后，数据访问的速度明显提高。但在每个维度上编制索引带来了一个问题，即数据量的增加。在数据重新组织后，数据的规模约为原始数据规模的四倍，数据尺寸的增加的主要原因来自新的索引的编制。

4.2 分布式数据查询机制

为了实现数据的快速检索，平台实现了分布的查询机制。目前由四台虚拟机构成，每台虚拟机包括 4GB 的内存、2GB Hz 的 CPU。这种方式不仅克服了 R 语言本身“单线程”的限制，与基于 R 的各种并行实现相比，该机制的灵活性也更大。对于一个检索，通过一个通用的分段方法将查询分成多个部分，每个

部分分别分配给不同的虚拟机。图 2 显示了在不同数目的虚拟机情况下，查询 12000、24000、36000 及 48000 条数据的性能。结果表明，随着机器数目的增加，检索的时间明显降低。除此之外，还能够发现，随着数据条数的增加，并行机制对时间开销上的降低效果更加明显。

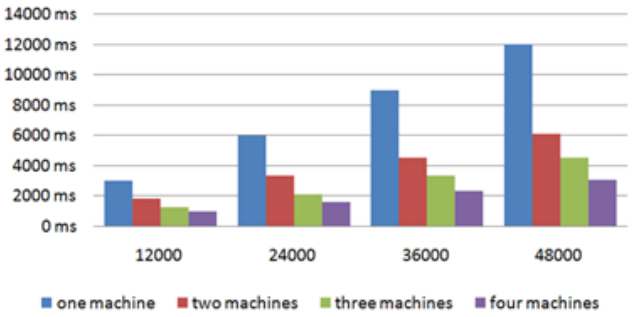

图 2 不同的虚拟机数目下，检索 12000、24000、36000、48000 条数据的性能比较

4.3 分段数据预取

即使数据检索的速度增加，如果检索的数据量太大，系统的性能依然不好。一个主要的瓶颈在于网络的传输。为了改善该问题，平台利用了 JavaScript 中的定时器实现分段数据预取。首先，计算结果数据集的数目，如果该数目大于某个阈值，则将该数据集交由分割方法对结果进行分割。定时器启动，并以分割结果按块获取数据。图 3 给出了该机制的效果。

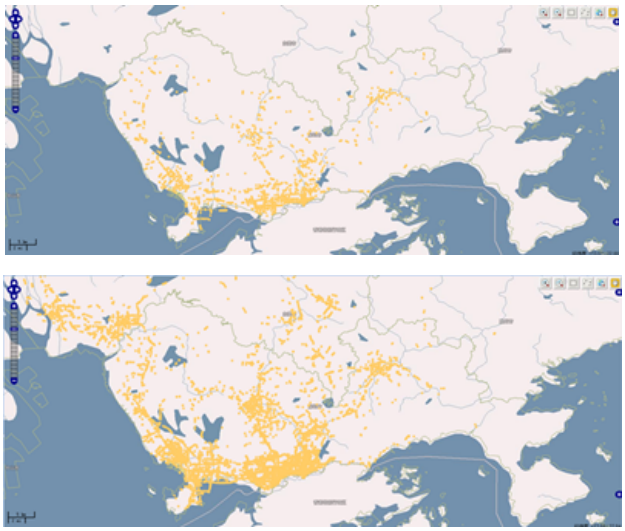

图 3 分段数据预取。上图包含 2000 个数据集，下图包含 18000 个数据集，上图至下图的变化是逐渐变化的

4.4 可视呈现

平台给出了两种可视呈现的方式，分别是基于地图的呈现与基于统计的呈现。基于地图的呈现结果如图 4。该界面的上部分是所有的任务模型。该部分的左边是常用任务，右边是所有维度的集合。在界面的中间是主展示区域，通过地图来将获取的结果进行呈现。系统给出了三种不同的基于地图的可视化方法，分别是热力图、点图及线段图。三种不同的可视化方法效果如图 5。在界面中，我们给出了三种交互工具，分别是动态过滤，可视化配置及可视化工具。用户可以通过动态过滤对真正感兴趣的内容进行过滤，该过滤全部在网页前端完成。可视化配置能够将数据的维度与地图上的图形属性进行关联，可视化工具则是来提高可视化的可用性而实现的一些功能。

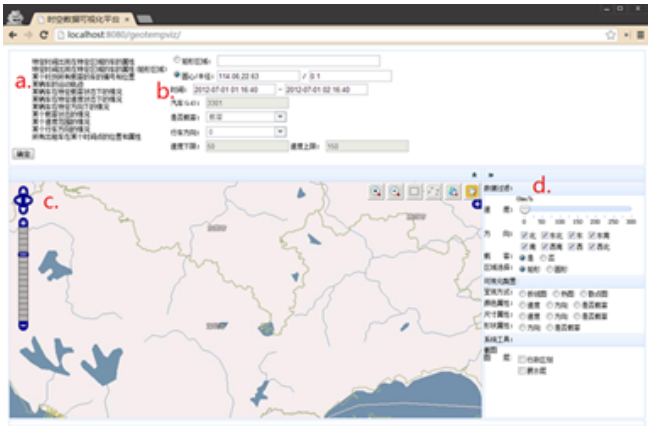

图 4 基于地图的呈现. (a). 预定义的任务. (b). 后台获取数据的约束. (c). 地图主视图. (d). 获取数据后的交互工具

图 5 三种可视组件. 左:50000 个节点的热力图. 中. 10000 个节点的点图. 右. 1000 个节点的线段图

另一种可视呈现的方式是基于统计的可视呈现。在本实现中，该呈现方式包括四个部分，如图 6 所示。由预定义的统计任务、数据选择器、可视化配置及可视化呈现四个部分组成。该图是一个预定义的任务，使用折线图呈现了一天中平均速度的变化。另外的一些预定义的任务可以用来降低分析师的工作强度。对于更自由的分析任务，可以通过数据选择器实现，数据选择器是用来确定哪些数据用来进行统计任务的。可视化配置确定使用哪种可视化组件进行呈现，以及不同的轴对应结果的哪些维度。

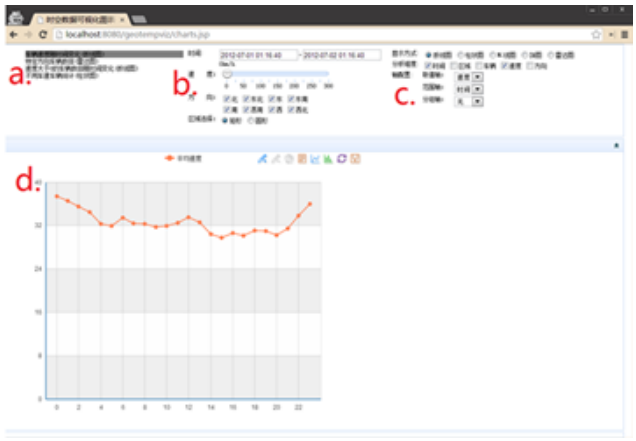

图 6. 基于统计的呈现. (a). 预定义的统计任务. (b). 数据选择器. (c). 可视化配置. (d). 可视呈现.

5 结论与下一步工作

本文给出一个预处理模型及数据模型，并且给出了一款数据处理与可视化平台。我们总结了当前时空数据分析的任务与分析模型，并基于此给出了一个 web 环境下支持大规模时空数据分析与可视化的平台。在下一步工作中，我们将会将平台应用于多个不同的时空数据集。

参 考 文 献

[1]. Manyika, J., et al., Big data: The next frontier for innovation, competition, and productivity. 2011: McKinsey Global Institute.

[2]. Halevi, G. and H. Moed, The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature. Research Trends, 2012(30).

[3]. Lynch, C., Big data: How do your data grow? Nature, 2008. 455(7209): p. 28-29.

[4]. Barga, R., et al., Bioinformatics and data-intensive scientific discovery in the beginning of the 21st century. OMICS: A Journal of Integrative Biology, 2011. 15(4): p. 199-201.

[5]. LaValle, S., et al., Big data, analytics and the path from insights to value. MIT Sloan Management Review, 2011. 52(2): p. 21-31.

[6]. Peuquet, D.J., It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. Annals of the Association of american Geographers, 1994. 84(3): p. 441-461.

[7]. Keim, D., et al. Challenges in visual data analysis. 2006.

[8]. Chi-Chun, P. and P. Mitra. FemaRepViz: Automatic Extraction and Geo-Temporal Visualization of FEMA National Situation Updates. in Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on. 2007.

[9]. Guo, D., et al. WMS-based Flow Mapping Services. in IEEE services 2012. 2012. Hawaii, USA: IEEE.

[10]. Ho Van, Q., T. Astrom, and M. Jern. Geovisual analytics for self-organizing network data. in Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on. 2009.

[11]. Janoos, F., et al. Activity Analysis Using Spatio-Temporal Trajectory Volumes in Surveillance Applications. in Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on. 2007.

[12]. He, L., et al. Visual analysis of route diversity. in Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on. 2011.

[13]. Weaver, C. Multidimensional data dissection using attribute relationship graphs. in Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on. 2010.

[14]. Afzal, S., R. Maciejewski, and D.S. Ebert. Visual analytics decision support environment for epidemic modeling and response evaluation. in Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on. 2011.

[15]. Diansheng, G., Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data. Visualization and Computer Graphics, IEEE Transactions on, 2009. 15(6): p. 1041-1048.

[16]. Peuquet, D.J. and M.-J. Kraak, Geobrowsing: creative thinking and knowledge discovery using geographic visualization. Information Visualization, 2002. 1(1): p. 80-91.

[17]. Casner, S.M., Task-analytic approach to the automated design of graphic presentations. ACM Transactions on Graphics (TOG), 1991. 10(2): p. 111-151.

[18]. Qian, L., et al. Delineating operations for visualization and analysis of space-time data in GIS. in GIS/LIS. 1997.

[19]. Zhou, M.X. and S.K. Feiner. Visual task characterization for automated visual discourse synthesis. in Proceedings of the SIGCHI conference on Human factors in computing systems. 1998. ACM Press/Addison-Wesley Publishing Co.

[20]. Koussoulakou, A. and M.-J. Kraak, Spatia-temporal maps and cartographic communication. Cartographic Journal, The, 1992. 29(2): p. 101-108.

[21]. Peuquet, D.J. and N. Duan, An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. International journal of geographical information systems, 1995. 9(1): p. 7-24.

[22]. Pelekis, N., et al., Literature review of spatio-temporal database models. The Knowledge Engineering Review, 2004. 19(03): p. 235-274.

[23]. Blok, C.A. Monitoring Change: Characteristics of Dynamic Geo-spatial Phenomena for Visual Exploration. Spatial Cognition 2000; 16-30].

[24]. Andrienko, N., G. Andrienko, and P. Gatalsky, Exploratory spatio-temporal visualization: an analytical review. Journal of Visual Languages & Computing, 2003. 14(6): p. 503-541.

[25]. Zheng, Y. and X. Zhou, Computing with Spatial Trajectories. 2011.

[26]. Chen, L., M.T. Özsu, and V. Oria. Robust and fast similarity search for moving object trajectories. in Proceedings of the 2005 ACM SIGMOD international conference on Management of data. 2005. ACM.

[27]. Jeung, H., et al., Discovery of convoys in trajectory databases. Proceedings of the VLDB Endowment, 2008. 1(1): p. 1068-1080.

[28]. Lee, J.-G., J. Han, and K.-Y. Whang. Trajectory clustering: a partition-and-group framework. in Proceedings of the 2007 ACM SIGMOD international conference on Management of data. 2007. ACM.

[29]. Yu, C. and D.J. Peuquet, A GeoAgent‐based framework for knowledge‐oriented representation: Embracing social rules in GIS. International Journal of Geographical Information Science, 2009. 23(7): p. 923-960.

[30]. Compieta, P., et al., Exploratory spatio-temporal data mining and visualization. Journal of Visual Languages & Computing, 2007. 18(3): p. 255-279.

[31]. Harrower, M., A. MacEachren, and A.L. Griffin, Developing a geographic visualization tool to support earth science learning. Cartography and Geographic Information Science, 2000. 27(4): p. 279-293.

[32]. Andrienko, G.L. and N.V. Andrienko, Interactive maps for visual data exploration. International Journal of Geographical Information Science, 1999. 13(4): p. 355-374.

[33]. Zicheng, L., Y. Yizhou, and C. Baoquan. Anomaly detection in GPS data based on visual analytics. in Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on. 2010.

[34]. Tominski, C., et al., Stacking-Based Visualization of Trajectory Attribute Data. Visualization and Computer Graphics, IEEE Transactions on, 2012. 18(12): p. 2565-2574.

[35]. Maciejewski, R., et al. LAHVA: Linked Animal-Human Health Visual Analytics. in Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on. 2007.

[36]. Sye-Min, C., et al. Maintaining interactivity while exploring massive time series. in Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on. 2008.

[37]. Ko, A.J., et al., The state of the art in end-user software engineering. ACM Comput. Surv., 2011. 43(3): p. 1-44.

杜 一，男，1988 年生，博士，助理研究员，研究方向：数据可视化，人机交互技术。

郭旦怀，男，1973 年生，博士，副研究员，研究领域：数据挖掘、数据可视化。

周园春，男，博士，研究员，研究领域：数据挖掘、云计算。

黎建辉，男，博士，研究员，研究方向：数据挖掘、云计算。