

zhanlijun

首页新随笔联系订阅管理

随笔 - 49 文章 - 0 评论 - 370

Visitors

CN

177,201

US

15,556

HK

5,779

JP

3,721

TW

3,008

SG

1,356

CA

816

FR

805

GB

775

AU

695

DE

670

KR

336

Pageviews: 377,081

Flags Collected: 83

FLAG counter

个人经历

2015 至今 阿里巴巴

2013-2015 美团

2010-2013 中科院（硕士）

2006-2010 浙大（本科）

阿里巴巴RDC长期招聘Java研发工程师，有意者站内联系！

昵称：zhanlijun

园龄：4年10个月

粉丝：664

关注：5

+加关注

最新随笔

1. 一个复杂系统的拆分改造实践

2. mysql死锁问题分析

3. 近期code review几处小问题集锦

4. 你应该知道的RPC原理

5. 如何健壮你的后端服务？

6. 如何用消息系统避免分布式事务？

7. 一个故事讲清楚NIO

8. 地图匹配实践

9. 利用模拟退火提高Kmeans的聚类精度

10. 空间插值文献阅读（Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall）

地图点聚合优化方案

<http://www.cnblogs.com/LBSer/p/4417127.html>

一、为什么需要点聚合

在地图上查询结果通常以标记点的形式展现，但是如果标记点较多，不仅会大大增加客户端的渲染时间，让客户端变得很卡，而且会让人产生密集恐惧症（图1）。为了解决这一问题，我们需要一种手段能在用户有限的可视区域范围内，利用最小的区域展示出最全面的信息，而又不产生重叠覆盖。



图1

二、已尝试的方案---kmeans

视觉上用聚类算法能较好达成我们目标，因此采用简单的kmeans聚类。根据客户端的请求，我们知道了客户端显示的范围，并到索引引擎里取出在此范围内的数据，并对这些数据进行kmeans聚类，最后将结果返回给客户端。

但是上线之后发现kmeans效果并不如意，主要有以下两个缺点。

a) 性能问题

kmeans是计算密集型算法，需要迭代多次才能完成，而且每次迭代过程中都涉及到复杂的距离计算，比较消耗cpu。

我们上线后遇到load较高的问题。

b) 效果问题

kmeans未能彻底解决重叠覆盖问题！可以看到有些聚合后的图标会叠合在一起。



三、优化方案

再次回顾我们的目的：我们需要一种手段能在用户有限的可视区域范围内，利用最小的区域展示出最全面的信息，而又不产生重叠覆盖。

3.1. 直接网格法

解决地理空间相关问题时，对空间划分网格这种方法往往屡试不爽。

随笔分类(57)
java(3)
LBS(10)
paper阅读笔记(2)
大数据(6)
定位原理/算法(3)
发表的SCI/SSCI(4)
服务治理(4)
空间索引原理(7)
数据库(5)
推荐相关(1)
线上问题定位及解决(2)
消息系统(2)
信息检索算法/实践(6)
应用服务器(2)

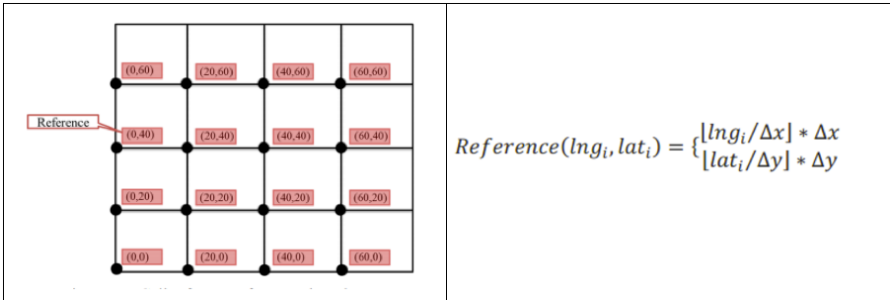
积分与排名
积分 - 115075
排名 - 2612

最新评论
1. Re:如何设计实现一个地址反解析服务？ 如果仅仅是为了将用户坐标解析到道路级别的话，也未必需要用栅格。对于任意一条道路，根据历史记录，可以得到定位于这条道路的所有点，根据这堆点可以得到一个外包多边形，以后所有落在这个多边形内的点都可以认为是..... --张可纯biubiu
2. Re:GeoHash核心原理解析 lucene里面使用了geohash，但是计算距离的时候貌似还是用经纬度计算距离，那使用geohash还有什么意义呢？ --casterQL

阅读排行榜
1. GeoHash核心原理解析(43980)

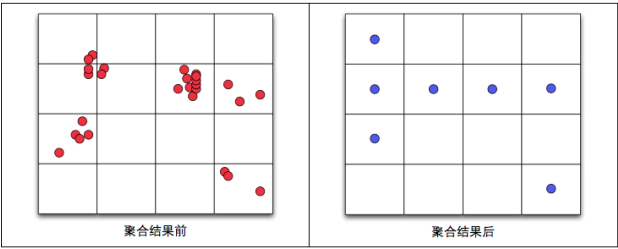
原理：将地图范围划分成指定尺寸的正方形（每个缩放级别不同尺寸），然后将落在对应格子中的点聚合到该正方形中（正方形的中心），最终一个正方形内只显示一个中心点，并且点上显示该聚合点所包含的原始点的数量。

如何将点落到正方形内呢？我们将空间人为指定100*100大小，通过这个公式进行映射。



优点：运算速度较快，每个原始点只需计算一次，没有复杂的距离计算。

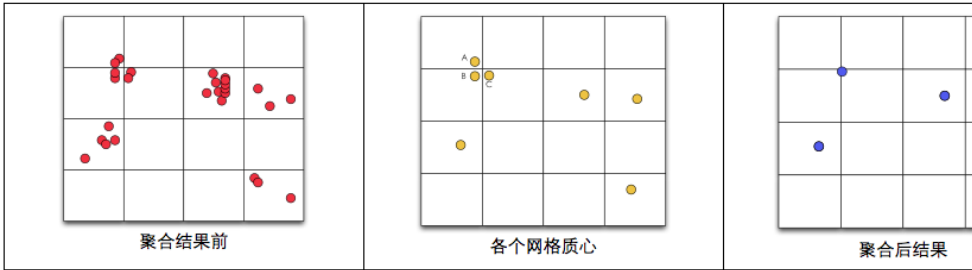
缺点：有时明明很相近的点，却仅仅因为网络的分界线而被逼分开在不同的聚合点中，此外，聚合点的位置采用的是该网格的中心，而非该网格的质心，这样聚合出来的点可能不能较精确反映原始点的信息。



3.2. 网格距离法

原理：沿用方案一思想，1）将各个点落到相应正方形内；2）求解各个网格的质心；3）合并质心：判断各个质心是否在某一范围内，如果在某一范围内则进行合并。

如何判断各个质心点是否需要合并呢？以A点为例，画一个矩形或者圆范围，落在此范围内的合并，B、C均落在范围内，因此A、B、C三点合并。



优点：运算速度同样较快，相对于方案一，多了求解质心以及质心合并两个步骤，但这两个步骤都较为简单，能很快完成。

3.3. 直接距离法

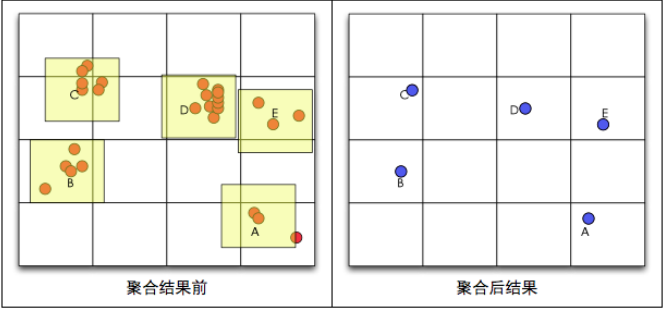
原理：初始时没有任何已知聚合点，然后对每个点进行迭代，计算一个点的外包正方形，若此点的外包正方形与现有的聚合点的外包正方形不相交，则新建聚合点（这里不是计算点与点间的距离，而是计算一个点的外包正方形，正方形的变长由用户指定或程序设置一个默认值），若相交，则把该点聚合到该聚合点中，若点与多个已知的聚合点的外包正方形相交，则计算该点到到聚合点的距离，聚合到距离最近的聚合点中，如此循环，直到所有点都遍历完毕。每个缩放级别都重新遍历所有原始点要素。

优点：运算速度相对较快，每个原始点只需计算一次，可能会有点与点距离计算，聚合点较精确的反映了所包含的原始点要素的位置信息。

缺点：速度不如完全基于网格的速度快等，此法还有个缺点，就是各个点迭代顺序不同导致最终结果不同。因此涉及到制定迭代顺序的问题。

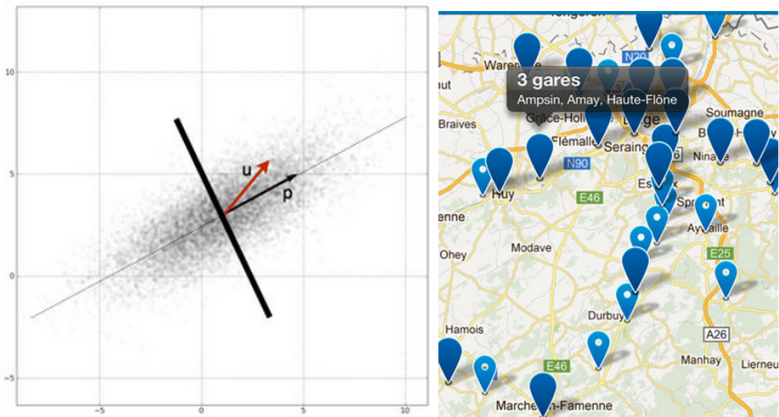
2. 你应该知道的RPC原理(30598)
3. 如何用消息系统避免分布式事务？(23677)
4. mysql死锁问题分析(22275)
5. 位图索引:原理 (BitMap index) (21132)

- 评论排行榜
1. 地图匹配实践(82)
2. 如何用消息系统避免分布式事务？(42)
3. 你应该知道的RPC原理(23)
4. GeoHash核心原理解析(22)
5. 地理围栏算法解析 (Geo-fencing) (20)



3.4. K-D树方法

这种方法需要结合PCA（主成分分析）和K-D树，在效果上比较好，但是性能较差，实现也较为复杂。



(http://applidium.com/en/news/too_many_pins_on_your_map/)

参考文献

<https://developers.google.com/maps/articles/toomanymarkers>
http://applidium.com/en/news/too_many_pins_on_your_map/
基于百度地图的标记点聚合算法研究
在线地图的点聚合算法及现状

转载请标明源地址： <http://www.cnblogs.com/LBSer>

分类: [LBS](#), [空间索引原理](#)

好文要顶

关注我

收藏该文

zhanlijun

关注 - 5

粉丝 - 664

±加关注

« 上一篇：[lucene join解决父子关系索引](#)

» 下一篇：[jvm间歇性崩溃分析](#)

posted @ 2015-04-11 10:29 zhanlijun 阅读(6911) 评论(4) 编辑 收藏

评论列表

#1楼 2016-04-01 11:58 邯郸打不倒

学长你好 我是一名准研究生 导师让我看聚合的东西。请问学长 直接距离法的源码 能让我参考一下吗？我的邮箱是yongfei.jin@qq.com 谢谢学长

支持(0) 反对(0)

#2楼[楼主] 2016-04-01 12:27 zhanlijun

@ 邯郸打不倒

我没有可开源的代码哦，其实知道原理后，建议自己实现下，也是对自己的锻炼

支持(0) 反对(0)

#3楼 2017-03-09 09:42 菜鸟Alex

有用

支持(0) 反对(0)

#4楼 2017-06-05 17:15 Ace8793

楼主，请教个事：

1.这些标记的点的数，是gis以外的服务(非GeoServer服务)提供的吗？

2.客户端拿到这些数据后，再渲染成标记点？

希望楼主指点，^_^

支持(0) 反对(0)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，访问网站首页。

【推荐】超50万VC++源码：大型组态工控、电力仿真CAD与GIS源码库！

【缅怀】传奇谢幕，回顾霍金76载传奇人生

【推荐】业界最快速.NET数据可视化图表组件

【腾讯云】买域名送解析+SSL证书+建站

【活动】2050 科技公益大会 - 年青人因科技而团聚



最新IT新闻：

- 蓝色光标陷劳资纠纷 6年人员成本增逾10倍
 - 投资育碧、腾讯财报 这两件事应该一起看
 - 7小时通宵大搜查！英国隐私监管机构进驻剑桥分析可查服务器
 - 手机厂商群撩小程序，是隔靴搔痒还是釜底抽薪？
 - 外卖平台该不该将“准时送达”服务变成增值业务？
- » 更多新闻...



最新知识库文章：

- 写给自学者的入门指南
 - 和程序员谈恋爱
 - 学会学习
 - 优秀技术人的管理陷阱
 - 作为一个程序员，数学对你到底有多重要
- » 更多知识库文章...

