

深度解读：Google S2地理位置库实现原理

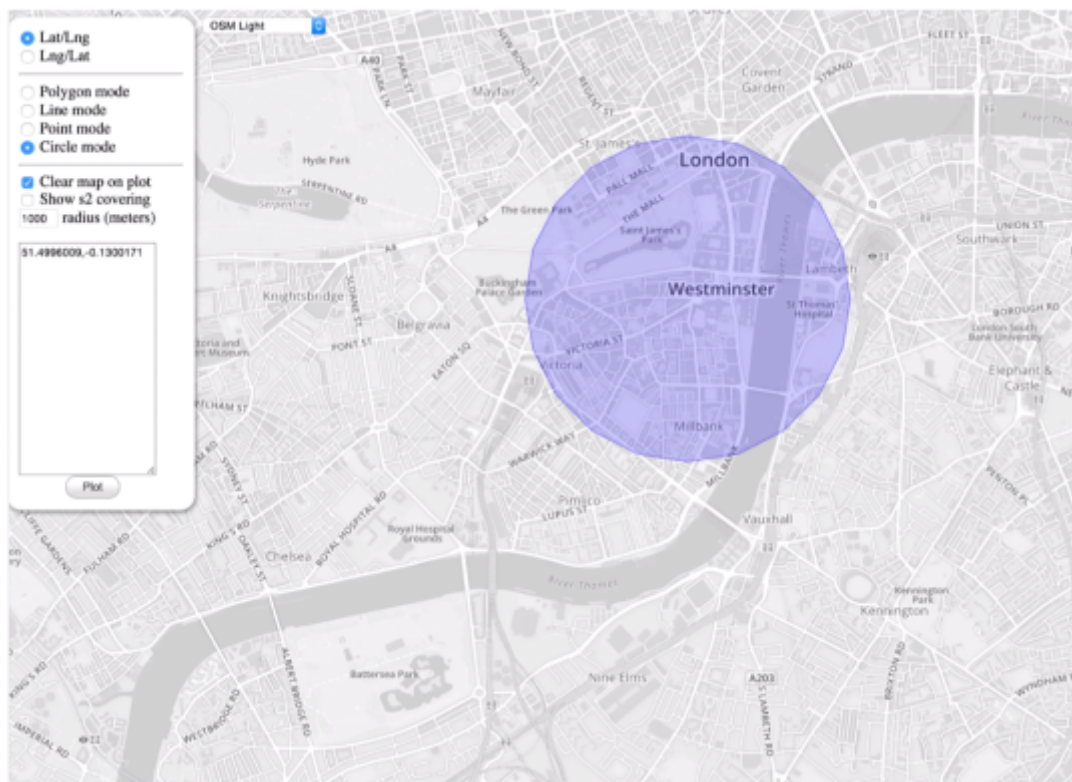
(<http://blog.bittiger.io/post183/>)

2018.01.12 作者：Rong (<http://blog.bittiger.io/author/littlemaorong/>)

今天我们来解读Google底层的地理位置库，这个位置库被用于Uber、Foursquare以及MongoDB等多个平台中，本文来看一看它在Uber中是如何应用的。

Uber面临的挑战：如何查找我身边的司机？

当一个用户在城市中心发出用车需求，Uber后台需要找到离用户几公里以内（如下图圆圈范围内）所有空闲的司机，来为用户提供服务。那如何实现这个功能呢？



挑战

- 如何唯一表示地球上的一块空间？

有了空间以后就可以把出租车司机放入进行查找。

- 如何将地球切分成大小近似的区块，并支持不同粒度的表示？

区块如果不等大，很可能发生负载不均问题。是否有可能实现如大洋洲大小的粒度，同时也能实现如沙粒大小的粒度。

为了解决上述两个问题，我们需要三个步骤。

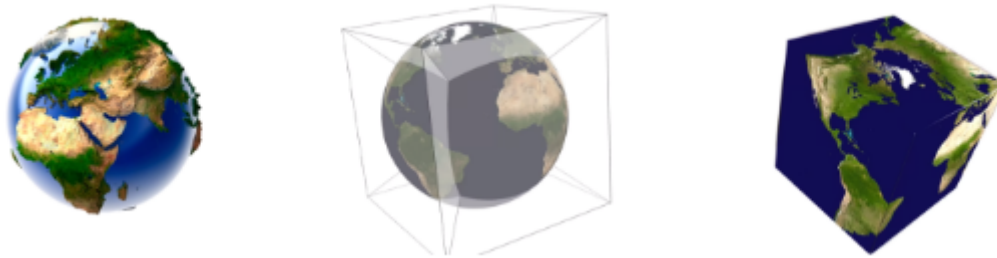
第一步，将三维地球变成二维；

第二步，将二维再转成一维；

最后一步，将一维表示成二进制码存储。

如何将三维变二维？

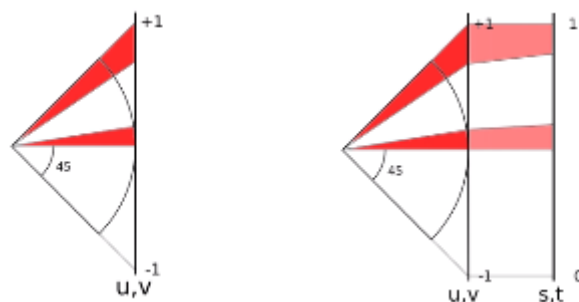
地球是一个三维的球体，我们只要把这个球体放在一个正方体中，想象从地球的中心向外发光，地球表面的点会投射在正方体上，地球表面就变成如下图所示的正方体。我们可以用0-5这6个数字来标记每一面，通过这个方法将三维变成二维。



投射的区间比例不同怎么办？

上面的方法实际上会产生一个问题：投射区间比例不同。在下图一中我们可以看到，虽然投出的角度数相同，但上方投出的区间却远大于下方。投射范围会出现中间短两边长的问题。如果投射范围不一样，不同区块的面积会差很多。

我们想到的方法是加入区间转换。在得到第一步投射之后，再进行二次变换，将上面长的拉短、将下面短的拉长，尽量让区间变相同。



如何选择转换方法？

我们希望选择的方法能够使投射出的区间尽量均匀。什么叫均匀呢？用投射出的最大面积除以投射出的最小面积，比例越大，说明越不均匀；比例越近于1，说明越均匀。

另外，我们还要考虑算法的效率。如果算得很慢，则不适合实际应用。

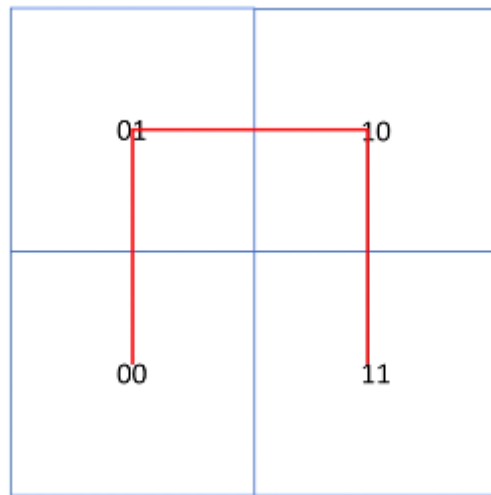
综合考虑这两个因素，可以进行以下考虑：

- 1.最简单的方法：线性方法，这是没有任何变换的。我们可以得到平衡度是5.2，即最大面积是最小面积的五倍多；但计算效率却很高，因为不需要变化，是0.086。
- 2.如果用正切的方法。因为考虑到角度问题，所以平衡度非常好，平衡度是1.41，即只差41%；但效率很差，耗时是线性方法的三倍以上。
- 3.考虑拟合，做二次性变换可以达到很好的效果。误差顶多到达2倍，但效率却和线性方法近似。因此是非常合适的转换方法。

候选方案	平衡度（最大面积/最小面积）	计算效率
线性	5.20	0.086
正切	1.41	0.279
二次	2.08	0.102

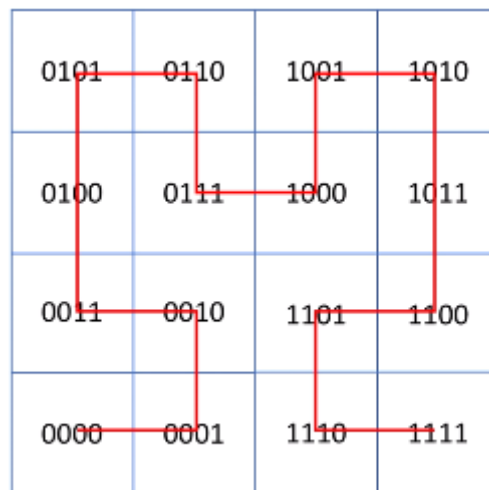
如何将二维变一维？

通过刚才的方法，我们能够将地球的表面转换成二维空间的平面。那接下来要将二维转变成一维。如果切割二维空间，可以切割出很多正方形。如何表示这个正方形呢？最简单的方法是在平面上进行遍历。每遍历到一个点，就给它标注一个值，比如00、01、10、11，随着二进制数字增加，相当于遍历面上不同的位置。



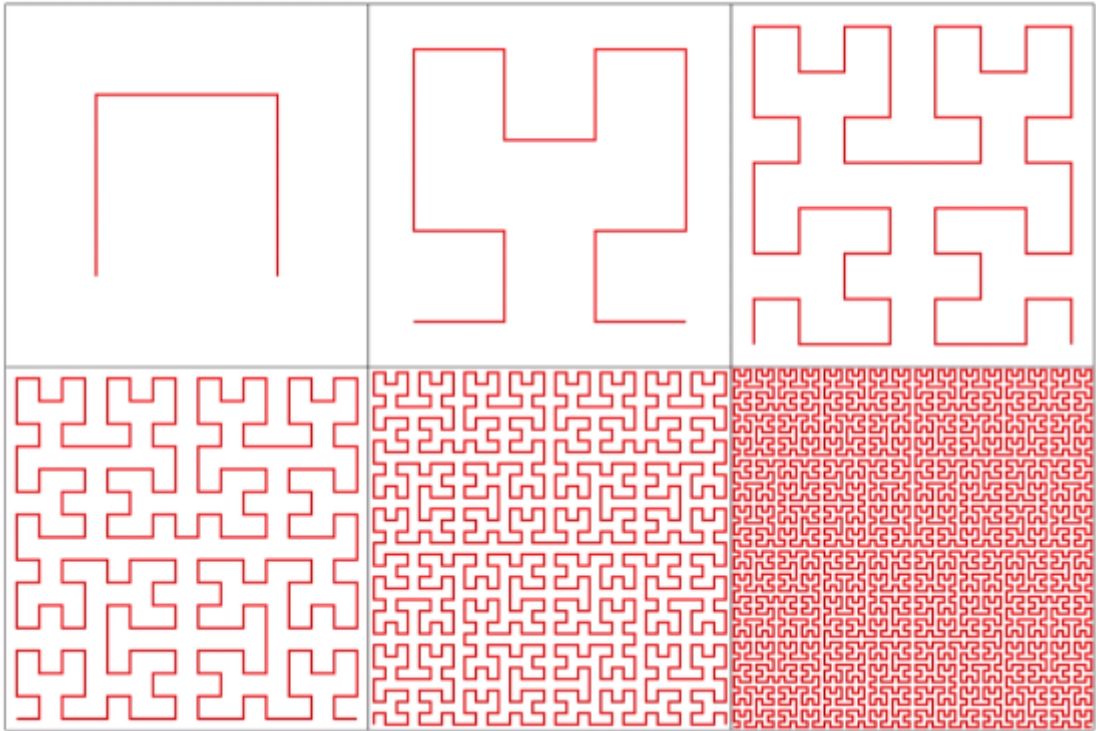
如何表示不同的粒度？

刚才只是四个正方形，我们能表示16个正方形吗？当然可以。我们从0000开始走一遍。因为现在方块多了，所以需要四bit来表示。0000、0001、0010、0011.....这样不断地走下去，就可以得到所有的位置。



更多的粒度如何表示？Hilbert curve（希尔伯特曲线）

如果还有不同的粒度，我们如何表示呢？大家可以看一下Hilbert曲线。从1个、2个、3个、4个、5个、6个不断地扩展，可以不断地自循环，进行迭代。这是一个非常美妙的数学图形，它能够将二维变换成一维。而且它还有一个优势，在这样一个变换过程中，地理位置相对邻近的点，在线上的距离相对也是邻近的。



Google S2的粒度是多少？

最高粒度Level 0，实际上是一个正方形的全面积85,011,012km²；如果是最小的粒度Level 30，只有0.48cm²。基本可以用来表示日常生活中可能用到的任何一个位置和空间。

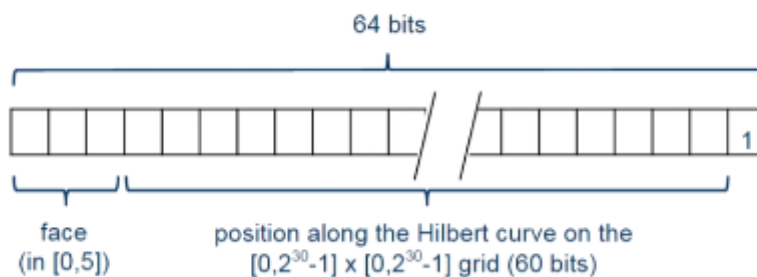
Level	Min Area	Max Area
0	85,011,012 km ²	85,011,012 km ²
1	21,252,753 km ²	21,252,753 km ²
12	3.31 km ²	6.38 km ²
30	0.48 cm ²	0.93 cm ²

↑
smallest cell

如何用二进制表示一个区间？

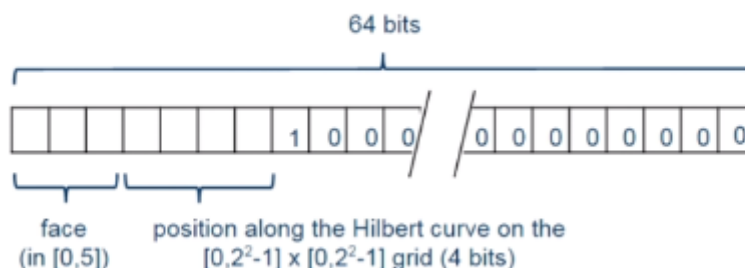
选择了平面和粒度之后，最后你还要想用多少Bit来实现？Bit越多肯定越占空间，Bit越少越好，这里用64 bit就能够实现了。怎么实现呢？比如要表达最高粒度Level30，我们需要用中间60 bit表达曲线的遍历过程，前面3个Bit来表示哪个面，它最多可以表达8个面，这里只需要能表达6个面就够了。最后一bit放置1，表示它前面60bit都是用来表示某个面里的具体位置。

S2 Cell ID of a leaf cell (level 30):



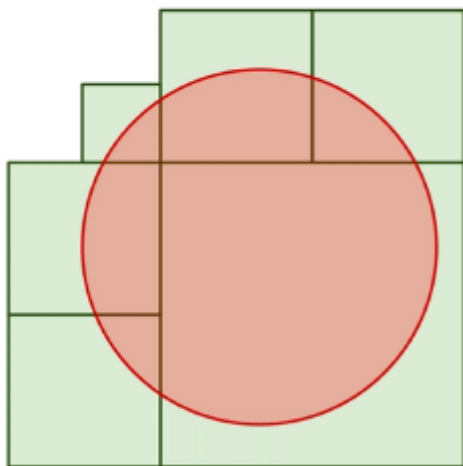
同理，如果粒度选择Level2就更容易。因为Level2中一个面上的不同正方形只需要4个bit来表示，前面3个bit确定哪个面之后用另外4个bit来表示位置，之后所有的bit都是闲置的，标1和一连串的0用来表达占位符。所以，前一部分是来表达level2的数据，后一部分用来占位。通过这个方法，我们知道了不同粒度的数据如何通过二进制数表达。

S2 Cell ID of a level-2 cell:

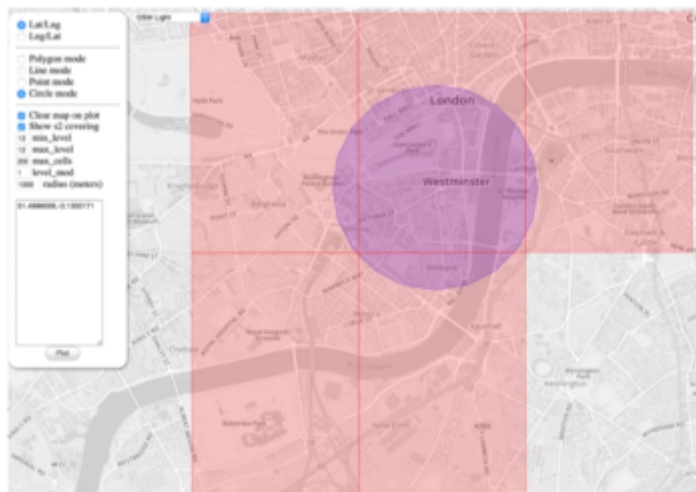


如何覆盖一个非规则区间？

一个方法就是用方块对圆进行覆盖，选择的方块粒度越细，可以覆盖的效果越好，但为了存储覆盖结果需要用的数字也会更多。比如下图的圆形需要用6个区间的数据，才能完全覆盖下来。



其实Uber运用了更简单的方法，它选择了统一的粒度，Level 12。



上图中可以看到因为用相同粒度，所以区间面积也是同样大小，总共用五块区间就可以完全覆盖圆形。像这样固定粒度，除了表示Level 12的数字，后面一串占位符是可以删除掉来节省空间。

总结

- 地球发光变平面

从地球中心发光，投射到外面的正方形，就得到一个平面

- 希尔伯特变成线

通过希尔伯特曲线走法，能够表示不同粒度的线，将二维变成一维

- 区间覆盖连成片

不同的区间，不同粒度方块，可以把一个图形覆盖出来，达到我们最终想要的结果

参考资料：《Geometry on the Sphere: Google's S2 Library》



Rong (<http://blog.bittiger.io/author/littlemaorong/>)

♥ Add favorites

🔗 Share (<https://www.addtoany.com/share>)



极品干货：生存模型在信用风险中的应用 ([http://blog.bitti...](http://blog.bittiger.io/%e6%8a%80%e6%9c%af/))

Nov 29, 2018 作者：Yifan

今天来给大家讲一讲如何把生存模型（Survival Model）应用到信用风险领域。众所周知，生存模型顾名思义是用来研究个体的存活率与时间的关系。例如研究病人感染了某种病毒后多长时间会...

技术 (<http://blog.bittiger.io/tag/%e6%8a%80%e6%9c%af/>)

Read more

(<http://blog.bittiger.io/%e6%9e%81%e5%93%81%e5%b9%b2%e8%b4%a7%ef%bc%9a%e7%94%9f>)



技术解读：Twitter实时大数据Streaming进化，从Storm到...

Nov 28, 2018 作者：Yifan

今天我们来解读文章《Twitter Heron: Streaming at Scale》，聊一聊从Storm到Heron，Twitter在大规模实时流处理的逐渐进化。这篇文章的作者是符茂松，他是Twitter的资深工程师、Heron的作者...

技术 (<http://blog.bittiger.io/tag/%e6%8a%80%e6%9c%af/>)

Read more

(<http://blog.bittiger.io/%e6%8a%80%e6%9c%af%e8%a7%a3%e8%af%bb%ef%bc%9a%e7%94%9f>)



面试必考：如何设计承载千万用户的Uber实时架构 (http://...)

Nov 27, 2018 作者：Yifan

今天我们解读的是Uber的首席架构师Matt Ranney所分享的Uber实时架构的从1到万。首先，Uber是什么？Uber是连接乘客和司机的交通平台，它专注的是运输业。当我们输入自己的位置，然后...

[简历 \(http://blog.bittiger.io/tag/%e7%ae%80%e5%8e%86/\)](http://blog.bittiger.io/tag/%e7%ae%80%e5%8e%86/)

[简历书写 \(http://blog.bittiger.io/tag/%e7%ae%80%e5%8e%86%e4%b9%a6%e5%86%99/\)](http://blog.bittiger.io/tag/%e7%ae%80%e5%8e%86%e4%b9%a6%e5%86%99/)

[Read more](#)

[面试 \(http://blog.bittiger.io/tag/%e9%9d%a2%e8%af%95/\)](http://blog.bittiger.io/tag/%e9%9d%a2%e8%af%95/)

(http://blog.bittiger.io/%e9%9d%a2%e8%af%95%e5%bf%85%e8%80%83%ef%bc%9a%e5%a6%82%)

COMMENTS

Add a public comment...

NAME *

EMAIL *

WEBSITE

Comment

About

BitTiger is the lifelong learning platform from Silicon Valley.

✉ contact@bittiger.io

BitTiger Careers (<https://careers.bittiger.io>)

Follow Us



(<http://www.bittiger.io/5592471491071020JINfuaiLB1Q>)
pnref=lhc) refer_flag=1001030102_&is_all=1)

Privacy Policy (/privacypolicy)

Terms of Use (/termsofuse)

Terms of Service (/termsofservice)

Code of Conduct (/codeofconduct)

Copyright Policy (/copyrightpolicy)



All Rights Reserved ©2017 BitTiger, Inc.