

R 和 Python，以及开源的讨论

1. R v.s. Python

- R 语言：开源，极其丰富的统计包，更有一大波开源群体贡献的高质量包，spark 和 python 都能直接调用 R，SAP 内存数据库可以直接调用 R 包，有一些包被微软收购。非常好的画图体验和强大的矩阵计算能力，还可以扩展为 R 集群。会 Matlab，R 上手也就几个小时强化训练就够了。R 遵循 GPL 协议
- Python 语言：工程化语言
- 总结：R 和 Python 还是没法相互替代的；哪种合适用哪个 R 统计上的处理更方便，包又多；Python 的系统整合上比 R 又强很多

2. 开源的功过：

- 正方：首先开源是一种潮流，不可阻挡；中国大数据行业有 Spark, Java, R, Python, C, MATLAB,... 在大数据行业 R 只是一个语言；Java 也是 Open Source 但每天公司用；大数据行业使用开源工具，并不意味着处于产业链的末端，而是投身到了开源潮流之中；开源潮流是即成事实，并且已经让大众享受到了科技带来的便利；山寨其实本身就是对知识产权的侵犯。不能因为同情而支持采用不正当手段获得利益；山寨是有意义的。但并不值得提倡、学习。开源同样也能起到便利大众的作用，也是光明正大的。而且可以形成一种良好的竞争氛围。同样的基础设施下，大家其实都是平等的，有实力者自然会胜出。这本来就是竞争的环境 你做的东西不好 自然要被淘汰 不论对方是否开源
- 反方：OpenCV 库开源让中国一些做图像算法的公司受到沉重打击；中国大数据行业使用开源工具，处于产业链的末端；对任何开源系统都应该保持警惕；优势企业会通过开源加大领先优势，弱势企业会因此丧失使用其他专利武器保护自身研发的能力；山寨给大众带来了便利，但开源让山寨受到了伤害；深圳很多企业应该既是开源的受益者，又是受害者。十几年前，OpenCV 还是真正开源的时候，里面真的有很多好算法，有些是尝试性的探索。Intel 收购了，那几个算法就被去掉了。同样的事情也出现在其他的开源库中，一旦被大公司收购，都会有类似的行为。其目的性昭然若揭。

3. 开源协议：BSD/AGPL/GPL

- ## 4. 认识：说到竞争，本质是产业链之争。这个理念在制造业，尤其是汽车企业非常先进。而计算机，互联网发展速度太快，以至于管理模型还没形成，就面临淘汰。这是互联网企业的优势，也是劣势。互联网是个朝阳的行业，却难以形成持久的企业。未来企业一定是要面对一个被互联网撕碎的市场，没有来得及反应的企业必将被淘汰。很多传统行业的既定思维方式或套路，在互联网面前都可能会失效。

微信群：数据分析和挖掘实战经验交流群

参与者：@Davy, @向东, @mengjiao, @John Heath, @Frankenstein Pei, @xiaowei,

@lwty-Raymond, @平行宇宙, @小东, @二哥