|  |  |
| --- | --- |
| Thinking1 | 当我们思考数据源的时候，都有哪些维度，如果你想要使用爬虫抓取数据，都有哪些工具 |
| 答： | 1.开放数据源，ex：Netflix，MovieLen。。。行业数据库。 2.爬虫。Beautiful soup，八爪鱼，Scrapy，Selenium，Puppeteer。 3.传感器 4.日志采集 |
| Thinking2 | 今天讲解了时间序列预测的两种方式，实际上在数据库内建时间属性后，可以产生时序数据库，请思考什么是时序数据库？为什么时间序列数据成为增长最快的数据类型之一（请思考并分享到班级微信群中） |
| 答： | 存放带有时间标签数据的数据库为时序数据库。 时间是我们可以感受到却无法控制的维度。我们无法将时间倒流，也无法穿越未来。而记录下时间点发生的事情，通过对历史的学习，对未来的遇见是我们人类在时间这个维度唯一可以做的，比如每个人做出的每个抉择都是自己人生经验汇总的结果。人类一直在追求确定稳定的东西来增加内心的安全感和抵抗不确定的风险，星座算命就是。从这个角度时序数据非常重要。 再从时序数据库需要的场景来看，一般需要展现历史趋势，周期规律，异常预警，对未来预测分析。而几乎任何的数据分析都可以加入时间的维度，因为任何事件行为的发生都有时间点。 |
| Thinking3 | 开源是当前重要的Trend，我们使用的statsmodels.tsa，tensorflow/keras都是开源工具 1）你都知道有哪些和AI相关的开源工具？ 2）阿里，微软，百度 都有哪些和AI相关的开源工具（包括LightGBM） 3）了解和使用这些工具，对于我们有哪些价值？ |
| 答： | 1.Tensorflow,pytorch,sklearn,surprise,deepctr… 2.LightGBM->MS, MNN->alibaba,Paddle ->baidu 3.open source一直是技术的趋势，原因在于开源社区的开放使得技术的发展速度一定快于闭源技术。科技公司比较热忠于开源技术，原因公司技术团队的实力有能力走在技术前沿并在遇到问题时解决。在一些传统企业和金融业开源的技术的使用没有那么广泛主要原因在于业务的重要性是需要有在线专家团队和有担保能力的技术公司来支持的，而且有些企业本身的技术实力较弱不适合开源技术运维，支持热线可以更容易更快速的解决问题。即使是这样在很多相对保守的企业里很多并没有那么重要的应用也有对开源技术的尝试和应用。对于我们一个可以拥抱开源技术的工程师他解决问题和独立思考解决问题的能力相对会比较强。 |