

什么是预测

What is forecasting?

李丰

北京大学光华管理学院

<https://feng.li/forecasting-with-ai>

历史中的预测

Forecasting in the history

巴比伦人的“肝卜” (Hepatoscopy)



- 巴比伦人(600 B.C)认为肝脏是血液的源泉，因此也是生命的基础。
- 有经过专门训练的祭司通过仔细观察绵羊的肝脏预测。
- 严格意义上讲不是预测，充其量是占卜。

预测师 (forecaster) 的命运

- 罗马帝国君士坦提乌斯二世(357 AD)颁布法令，禁止任何人“咨询占卜者、数学家或预测者……愿对预测未来的好奇心永远被压制”。
 - *“to consult a soothsayer, a mathematician, or a forecaster ... May curiosity to foretell the future be silenced forever.” -- Codex Theodosianus 9.16.4*
- 1824年，英国也曾有过类似的预测禁令，规定“任何假装或自称会算命的人”都“被视为流氓和流浪汉”。最高可判处三个月监禁和苦役！
 - *“every person pretending or professing to tell fortunes” was “deemed a rogue and vagabond”. The punishment was up to three months’ imprisonment with hard labour! -- Vagrancy Act, 1824, Section 4, repealed in 1989.*

二十四节气：周期性地预测季节和气候的变化



- 夏商(1600 B.C)时期只有四个节气。
- 秦汉(220 B.C)年间，二十四节气已完全确立。
- 春雨惊春清谷天，夏满芒夏暑相连，
- 秋处露秋寒霜降，冬雪雪冬小大寒。
- 每月两节不变更，最多相差一两天，
- 上半年来六廿一，下半年是八廿三

预测与预测傻傻分不清楚

Forecasting and Prediction

Forecasting vs Prediction

- Forecasting 是基于历史数据和既定模式，对未来事件进行量化的、概率性的估计。
- Prediction 是对任何未发生事件的定性的、判断性的推论。
- 所有 Forecasting 都是 Prediction，但并非所有 Prediction 都是 Forecasting。

可预测性

What can be forecast?

为什么有一些场景比其他场景更容易（难）预测？

- 你几乎不可能预测你明天能否中彩票。
- 但是你要是知道本场别人都买了什么，你的策略就是**就买别人不买的**。
- 但是你的中奖概率不会变。



可预测性取决于以下几个因素

- 我们对影响预测的因素的了解程度
 - 什么决定了新能源电力的需求？
- 有多少数据是可以用做预测支持
 - 新冠后中国出境游的复苏程度？
- 未来与过去的相关程度
 - 上周我中了刮刮乐，这周我是否还能中奖？

什么是好的预测？

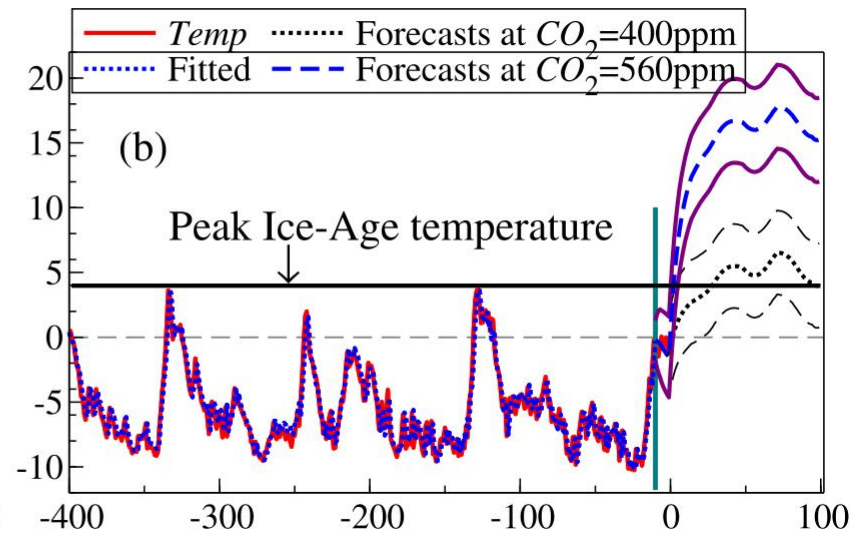
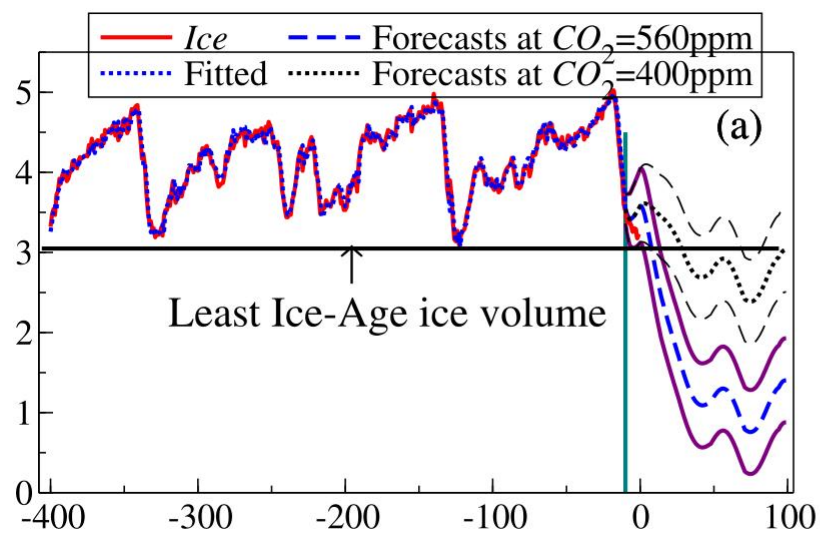
What are good forecasts?

评估预测的朴素逻辑

- 预测要在大量的数据上表现优异，偶尔一次预测的好不是真的好。
- 多条时间序列的交互学习能提升综合预测精度。
- 如果没有足够的历史数据，创造数据也要预测。
- 速度是现代商业预测的核心竞争力。

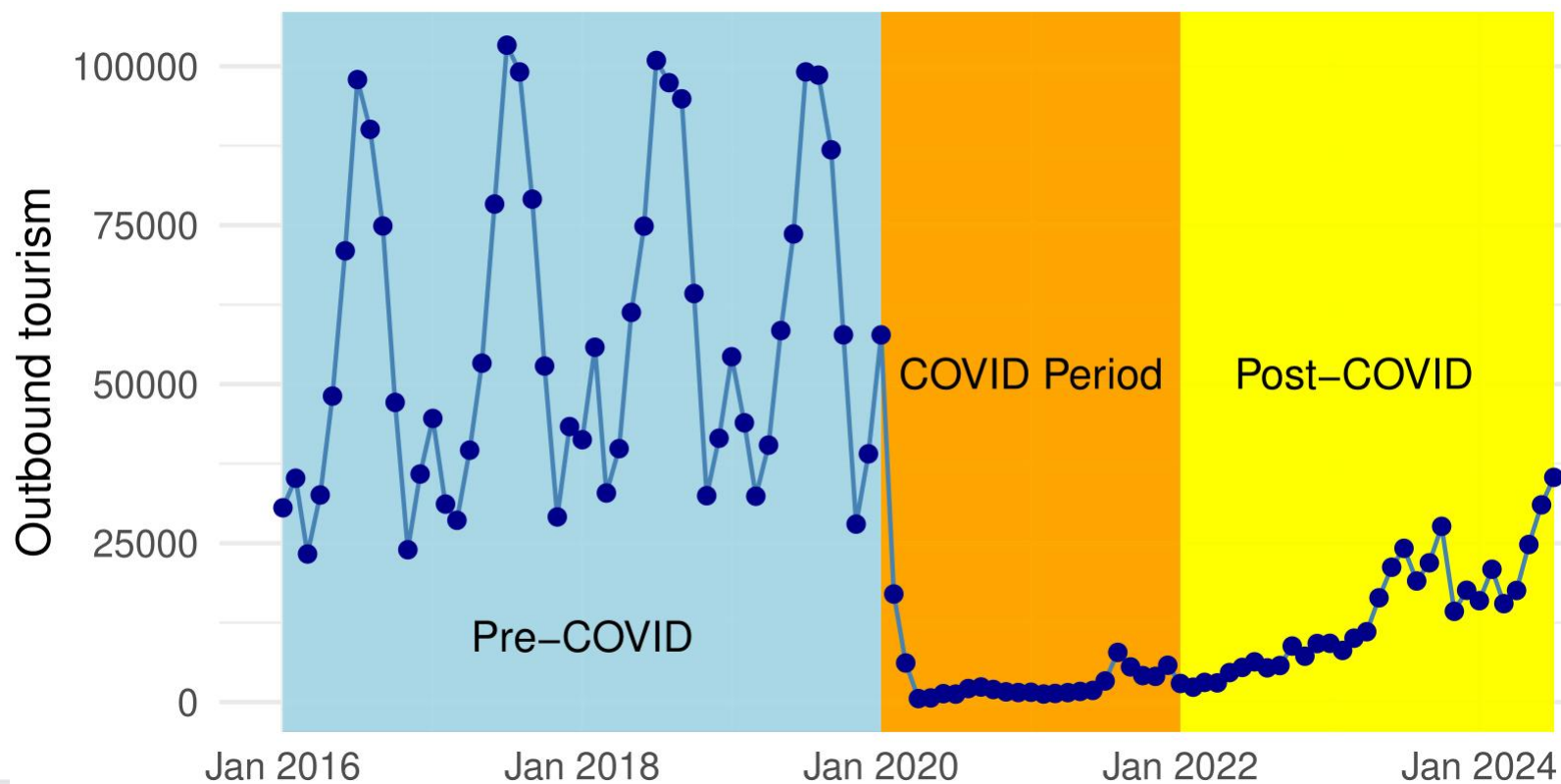
远期预测精度

- 短期预测也许精度可以，准确的长期预测更难，预测误差会逐步累积。
- 冰山融化时间预测



未来的不确定评估

- 因为我们永远都不会知道未来，时间序列预测的评估就变得非常困难
- 五个专家给出五个不同旅游恢复预测，我们该相信谁？



现代时间序列预测

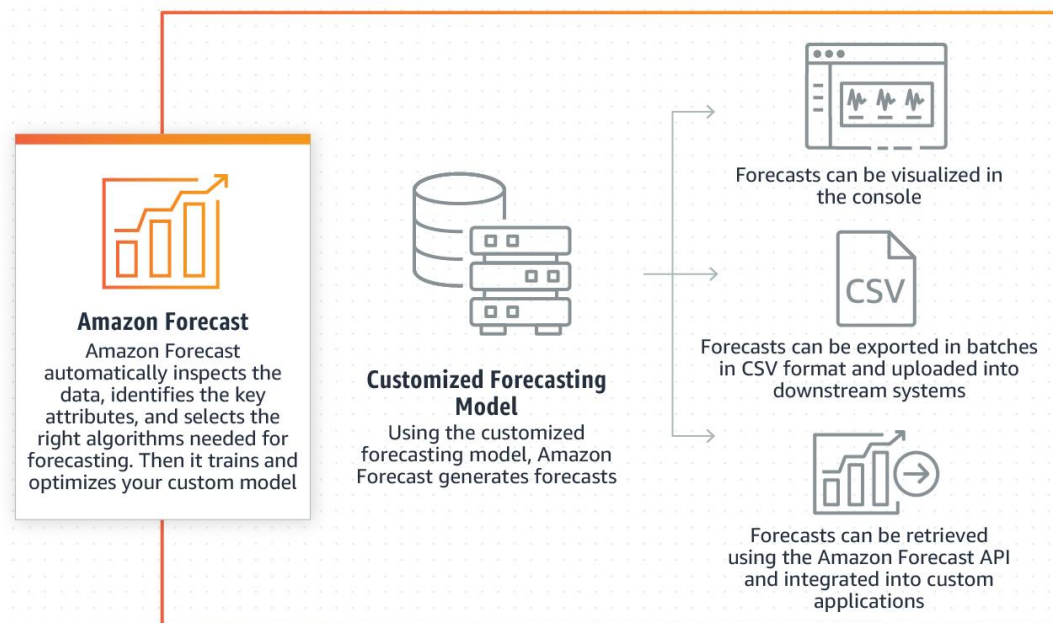
Modern time series forecasting

现代时间序列预测

- 是面向未来的统计思维+计算机工程+人工智能

亚马逊预测平台

- 亚马逊在 180 多个国家销售超过 4 亿种产品。
- 预测团队由2007年 10 人变到2020年的 200人。
- 2013年，自然语言处理、图像识别、深度学习算法被引入预测平台。
- 2020年，Transformer 架构被引入并构建了全自动预测框架。

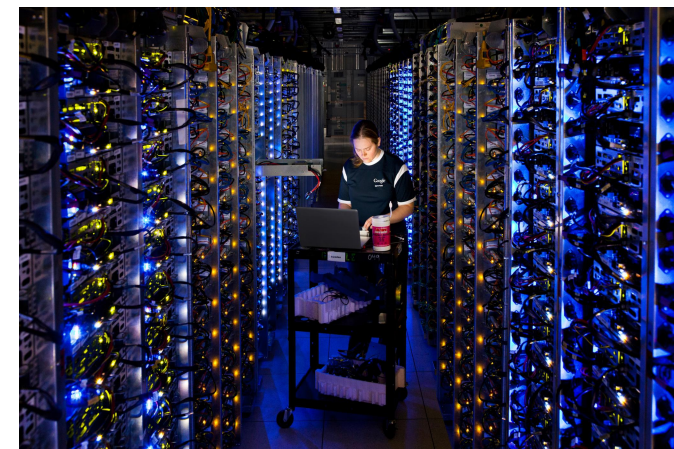


Google 预测实践

- 2010–2015
 - 重点： 搜索增长预测
 - 目标： 单变量/局部模型、因果模型
 - 工具： Fargo（内部）、贝叶斯结构时间序列（BSTS）
- 2021–2025
 - 重点： 产能管理、缺货预警、复杂性应对、基准表现
 - 目标： SLO驱动的产能、可解释性、跨资源精度、预训练/免训练
 - 工具： BQML ARIMA+、TiDE、Starry-Net、PreStarry、TimesFM、聊天机器人
- 未来方向 (??)
 - 可扩展性、自动化、LLM与多模态预测、时间序列基础模型



Chris Fry, Director of Data Science at Google Cloud. He dedicated to driving efficient and reliable utilization of Google's datacenter infrastructure



讨论

- 你所在(所了解)的行业/公司的预测是怎么做的？
- 你们的预测准吗？
- 你们的决策多大程度上受预测影响？