



推荐系统新零售领域的应用

尼奥

袋鼠云高级算法专家



目录

CONTENTS

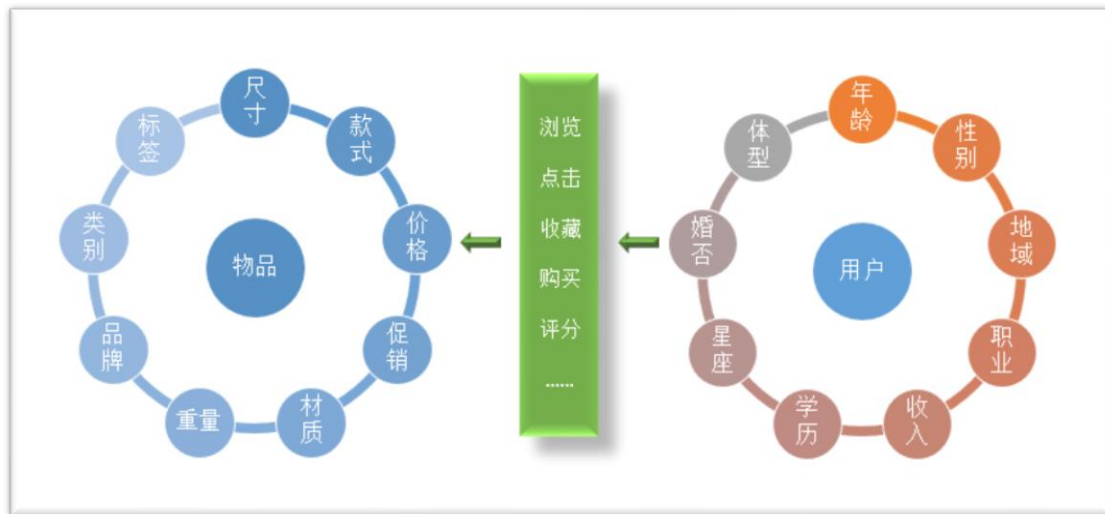
- 推荐系统概述
- 推荐算法简介
- 推荐算法的落地



推荐系统概述



- 什么是推荐系统？
 - 推荐系统和推荐算法的区别
- 推荐系统的应用场景
 - 电商平台
 - 社交平台
 - 影音平台
 - 资讯平台



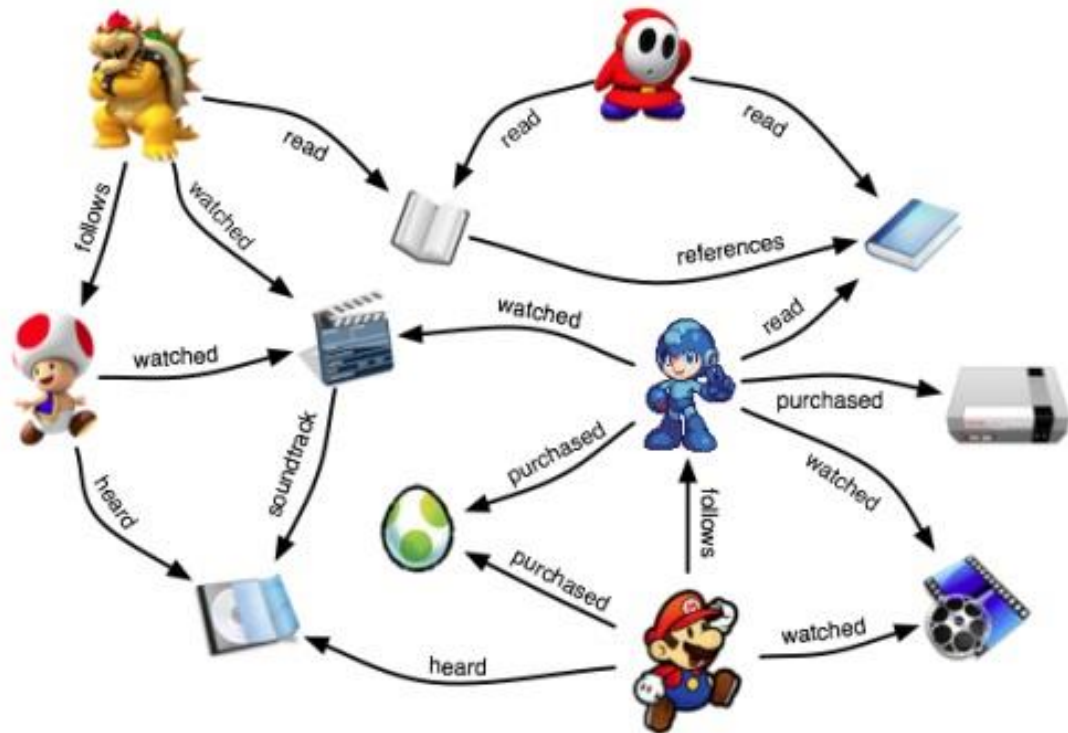
什么是推荐系统？



- Predict preference or rating of an 'item' from a user

——wiki

- 千人千面
 - 增加点击
 - 发掘长尾



推荐系统应用场景



IT大咖说
知识共享平台

- 电商平台
- 社交平台
- 影音平台
- 资讯平台
- 新零售

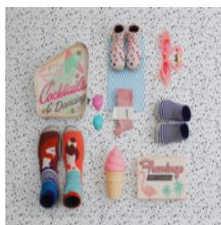
猜你喜欢



欧式真皮转角沙发高档皮艺沙发
橡木雕花组合L型转角实木沙发

¥6200

销量:6



包邮 国内现货 法国Collegien正
品手工鞋室内鞋学步鞋 不退不

¥200



欧式全实木沙发 简欧沙发客厅
沙发 新古典沙发 美式法式沙发

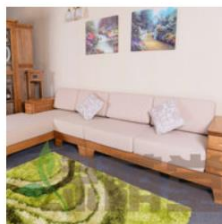
¥3980



中式沙发组合实木禅意现代新中
式沙发样板房间售楼处简约沙发

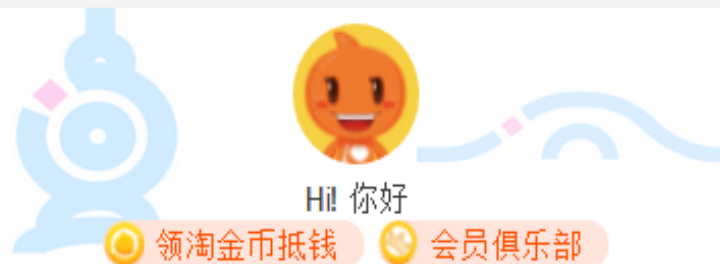


红木沙发红木家具花梨木色菠萝
格木实木现代客厅沙发豪华万字



原单欧式橡木实木贵妃榻沙发

生活 苏宁易购



领淘金币抵钱

会员俱乐部

登录

注册

开店

网上有害信息举报专区

公告

维权

论坛

安全

公益

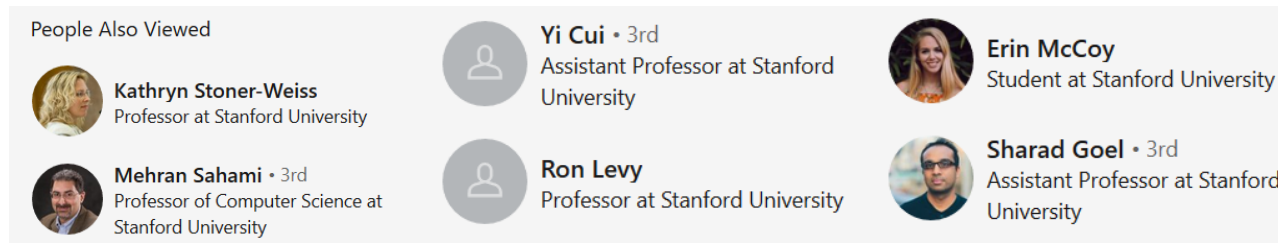
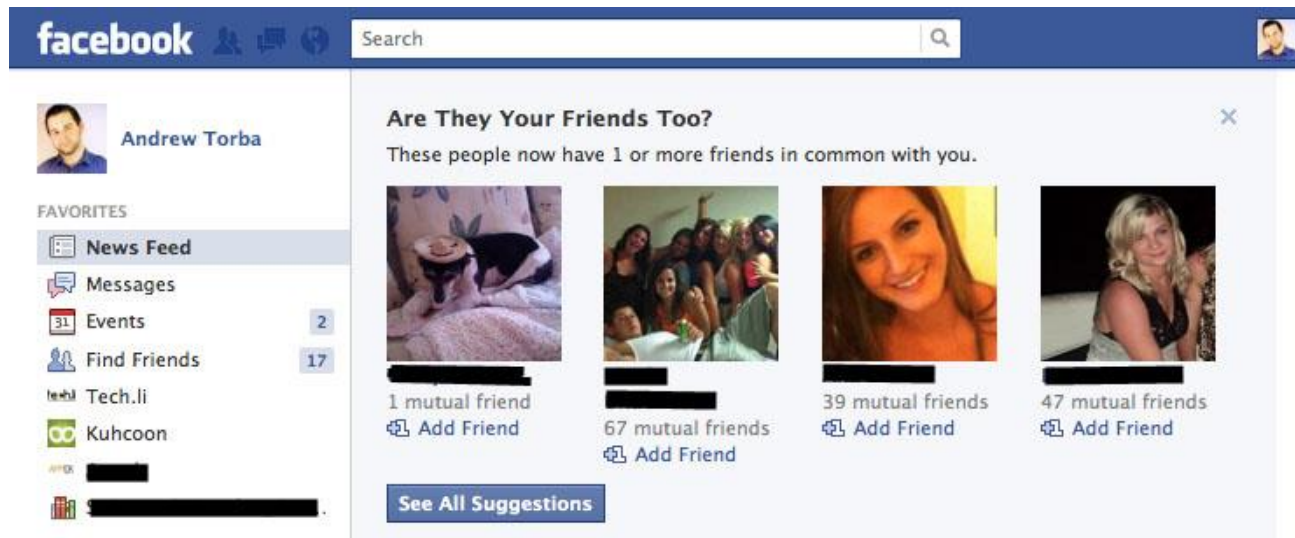


AICUG人工智能技术沙龙 (杭州站)

推荐系统应用场景



- 电商平台
- 社交平台
- 影音平台
- 资讯平台
- 新零售



推荐系统应用场景



IT大咖说
知识共享平台

- 电商平台
- 社交平台
- 影音平台
- 资讯平台
- 新零售

《新时代 向心力》
上传于 2018-02-19 优酷资讯 [+订阅](#)

你可能还喜欢这些节目



罗辑思维“时间的朋友”
分享创业投资的那些事



牛人说
全力聚焦互联网金融



罗振宇“时间的朋友”
分享创业投资的那些事



财经连环画
与你聊聊钱包的那些事



天竞择物 第一季
生活可以更美好



第263期 为啥印度买武器总比中国贵



经典传奇 84年春晚幕后大揭秘 1992-14 高清



移动的皇宫, 揭开梦幻“东方快女”的神秘面纱



宸冰发现 - 马连良的戏曲与人生 01



AICUG人工智能技术沙龙 (杭州站)



- 电商平台
- 社交平台
- 影音平台
- **资讯平台**
- 新零售

资讯推荐系统本质上要解决用户，环境和资讯的匹配： $y = F(x_i, x_u, x_c)$



推荐系统应用场景



IT大咖说
知识共享平台

- 电商平台
- 社交平台
- 影音平台
- 资讯平台
- 新零售

图书推荐 (更新时间: 2017-01-01, 每周更新一次)

袋鼠云新华书店大数据项目

添加图书

书名	作者	出版日期	定价	ISBN	分类	近30天销量	门店库存	推荐采购量	推荐原因
<input type="checkbox"/> 深海危机/中国原创奇幻动物小说	唐池子	2015-12-01	16.0	9787534289637	中国儿童文学	3	0	10	相似门店热销
<input type="checkbox"/> 爱学拼音(上)/魔力铅笔	爱学系列...	2009-06-01	3.0	9787534252747	学前教育	15	0	30	相似门店热销
<input type="checkbox"/> 科学野战营(史无前例的发明)	纸上魔方	2016-08-01	20.0	9787534293665	少儿百科词典	45	5	10	相似门店热销
<input type="checkbox"/> 小学生生字卡(3上)	林彤	2015-06-01	10.5	9787534285868	小学语文	18	0	20	相似门店热销
<input type="checkbox"/> 三国演义/世界少年文学经典文库	罗贯中	2009-05-01	14.5	9787534253881	儿童文学	37	0	20	相似门店热销

全选

删除

加到我的关注

推送至多门店ERP

推送至门店ERP

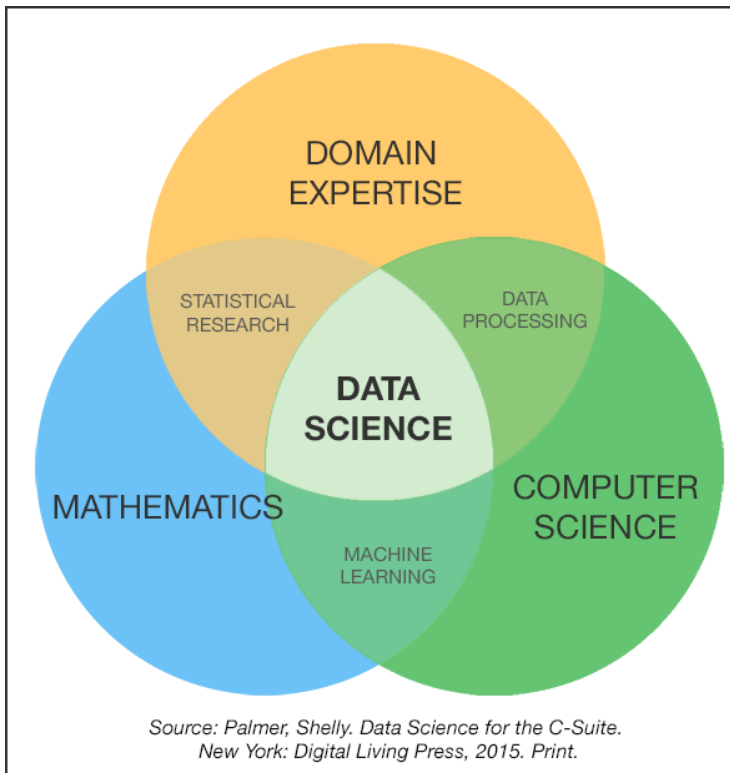


AICUG人工智能技术沙龙 (杭州站)

推荐系统概述



- 技术主导
- 业务主导



目录

CONTENTS

- 推荐系统概述
- 推荐算法简介
- 推荐算法的落地



推荐算法介绍

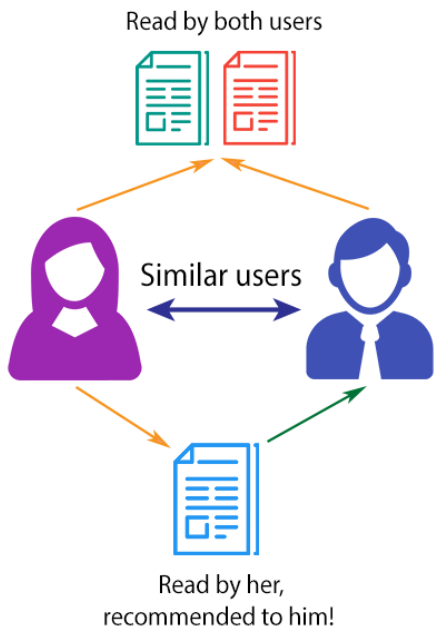


- 基于内容 Content-based Recommendation
- 协同过滤 Collaborative Filtering Recommendation
- 关联规则 Association Rule-based Recommendation
- 基于效用 Utility-based Recommendation
- 基于知识 Knowledge-based Recommendation
- 融合推荐 Hybrid Recommendation

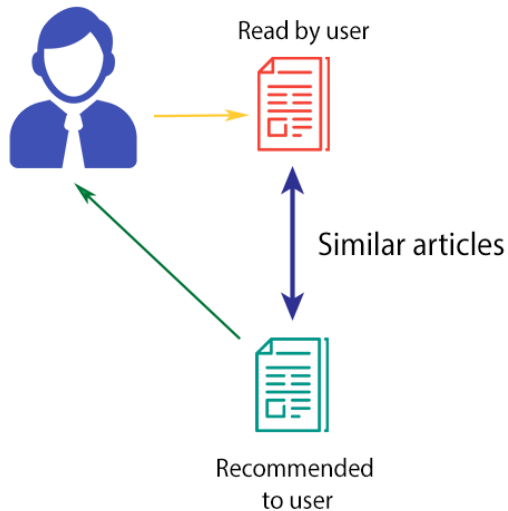
(最) 常用推荐算法



COLLABORATIVE FILTERING



CONTENT-BASED FILTERING





- 方法
 - 数据分析为主
 - 依靠显性标签
- 基本机器学习算法
 - 协同过滤 (etrec, MF)
 - 构建标签
- 前沿算法
 - 深度学习
 - 时间序列标签
 - 标签向量化
 - . . .

图书推荐 (更新时间: 2017-01-01, 每周更新一次)

袋加云新华书店大数据项目 添加图书

书名	作者	出版日期	定价	ISBN	分类	近30天销量	门店库存	推荐采购量	推荐原因
<input type="checkbox"/> 深海危机/中国原创奇幻动物小说	唐池子	2015-12-01	16.0	9787534289637	中国儿童文学	3	0	10	相似门店热销
<input type="checkbox"/> 爱学拼音(上)/魔力铅笔	爱学系列...	2009-06-01	3.0	9787534252747	学前教育	15	0	30	相似门店热销
<input type="checkbox"/> 科学野战营(史无前例的发明)	纸上魔方	2016-08-01	20.0	9787534293665	少儿百科词典	45	5	10	相似门店热销
<input type="checkbox"/> 小学生生字卡(3上)	林彤	2015-06-01	10.5	9787534285868	小学语文	18	0	20	相似门店热销
<input type="checkbox"/> 三国演义/世界少年文学经典文库	罗贯中	2009-05-01	14.5	9787534253881	儿童文学	37	0	20	相似门店热销

全选 删除 加到我的关注 推送至多门店ERP 推送至门店ERP



区别



1.基本推荐方法

- 逻辑判断
 - ETL的逻辑
 - 基于统计分析的逻辑
- 人工判断
 - 信息变化特别快或者信息不全
 - 业务经验主导
- 优缺点
 - 结构清晰，难度低，灵活性强
 - 受数据质量和数量影响较强，对业务依赖较大，可复制差





2.基础推荐算法

- 协同过滤：

- ETREC (阿里云)
- MF (评分)
- *SVD

- 标签加工：

- 标签分类
- 标签构造

- 优缺点：

- 可扩展性强；可复制性高
- 对算法的理解有要求，对业务的理解要求较高

购买此商品的顾客也同时购买





3. 前沿算法

- 深度学习：
 - LSTM
- 时空信息的提取：
 - 时间序列分析
 - LBS
- 标签向量化：
 - 隐性标签
- 优缺点：
 - 深度定制，千人千面
 - 模型庞大，对算法基础要求高，对业务要求非常高





- 1、业务效果好的算法并不一定是复杂的算法
- 2、做项目要比光看书的收获大得多
- 3、推荐是个系统工程，算法很重要，但不是全部



CONTENTS

目录

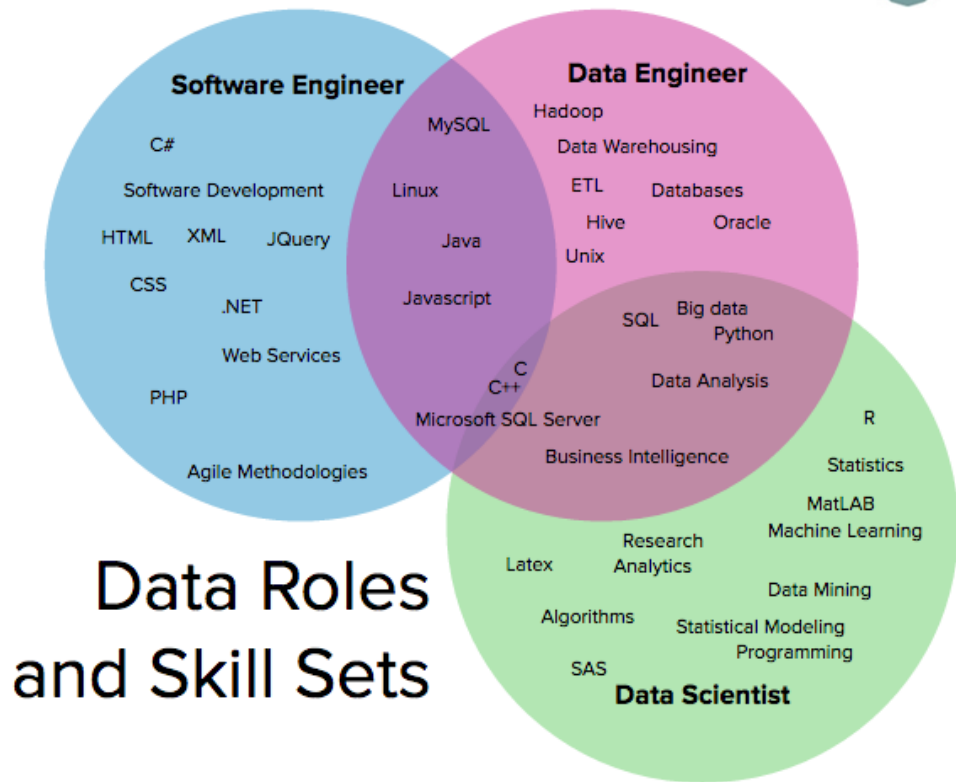
推荐系统概述

推荐算法简介

推荐算法的落地



- 推荐系统的部署
- 推荐系统的评估
- 推荐系统的维护



一个有意思的案例



IT大咖说
知识共享平台

Netflix Awards \$1 Million Prize and Starts a New Contest

BY STEVE LOHR SEPTEMBER 21, 2009 10:15 AM



Jason Kempin/Getty Images Netflix prize winners, from left: Yehuda Koren, Martin Chabbert, Martin Pottie, Michael Jahrer, Andreas Toscher, Chris Volinsky and Robert Bell.

Update | 1:45 p.m. Adding details announced Monday about the extremely close finish to the contest.

Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)

Arvind Narayanan and Vitaly Shmatikov

The University of Texas at Austin

February 5, 2008

Abstract

We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge.

We apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous movie ratings of 500,000 subscribers of Netflix, the world's largest online movie rental service. We demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, we successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.

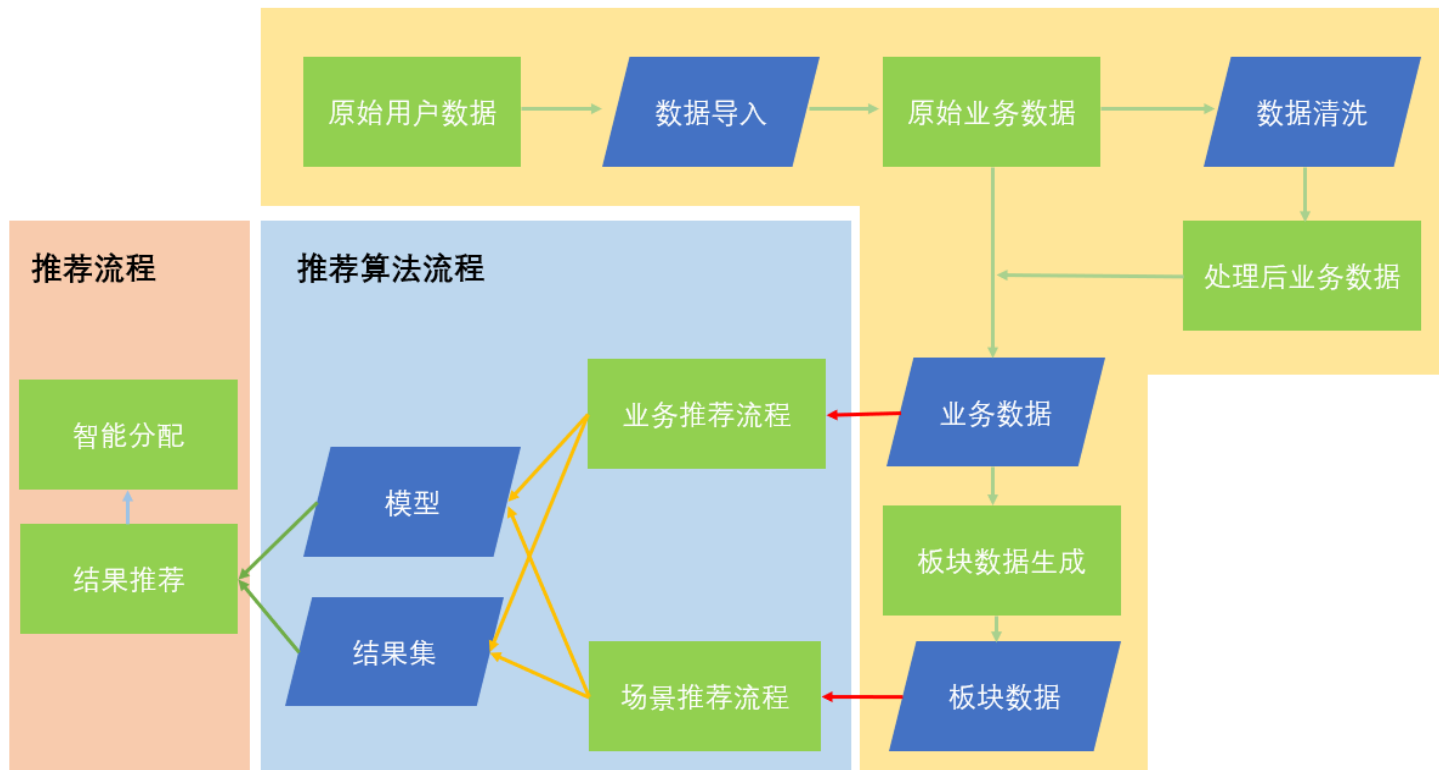
And in fact, it was. Despite all the plaudits and case studies, [Netflix announced this week](#) that despite paying \$1 million dollars to a winning team of multinational researchers in 2009, they never bothered to implement their solution. Why? Because, according to Netflix the "additional accuracy gains that we measured did not seem to justify the engineering effort needed to bring them into a production environment."



AICUG人工智能技术沙龙（杭州站）



数据处理流程





- 准确与否
 - A/B TEST
 - 业务人员的认可
- 影响结果的因素：
 - 数据质量（数据量，数据维度，脏数据量）
 - 评估周期（小时，天，周）
 - 评估方式（销量量/浏览量增长多少，5%？10%？）
 - 业务人员的期望



Thanks!

Q & A

