



推荐系统（1）-业界推荐系统架构

Alan
数据分析、挖掘、机器学习

关注他

135 人赞同了该文章

一、推荐系统意义、架构介绍

推荐系统的意义目的：

- **提升用户体验。**通过个性化推荐，帮助用户快速定位感兴趣的信息。
- **提高产品销售。**帮助用户和产品建立精准连接，从而提高产品转换率。
- **发掘长尾价值。**根据用户兴趣，清空不是很热门的商品销售给特定人群。
- **方便移动互联网交互。**减少用户操作，主动帮助用户找到感兴趣内容。

1.1业界主流推荐系统架构

试用模式

XMind:ZEN



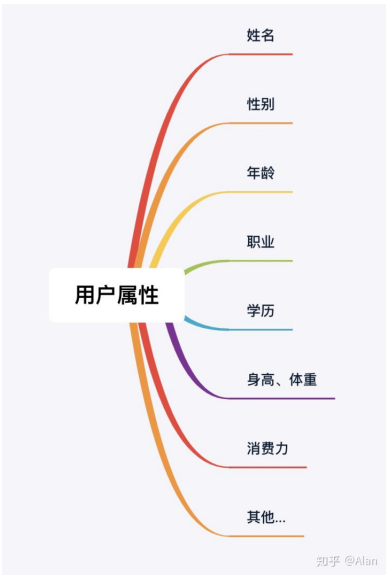
业界主流推荐系统架构



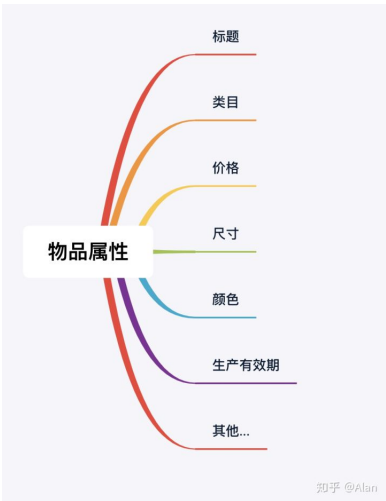


业界推荐系统通用架构

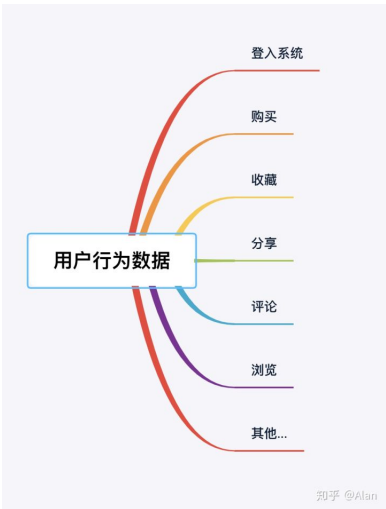
1.2用户本身数据



1.3物品本身数据



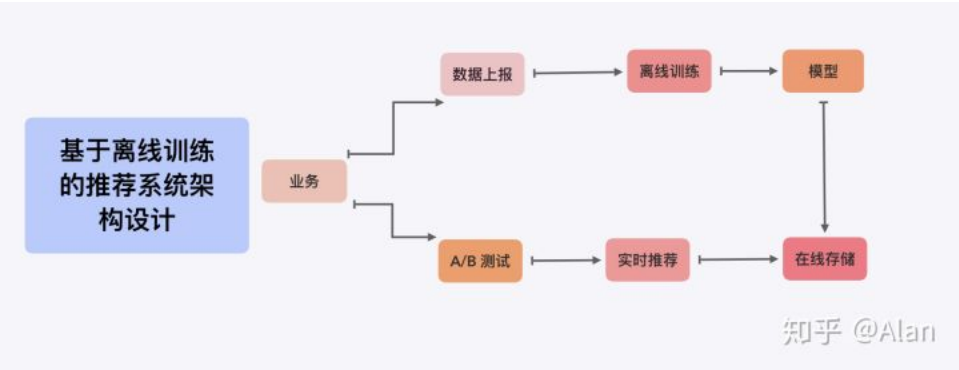
1.4用户行为数据





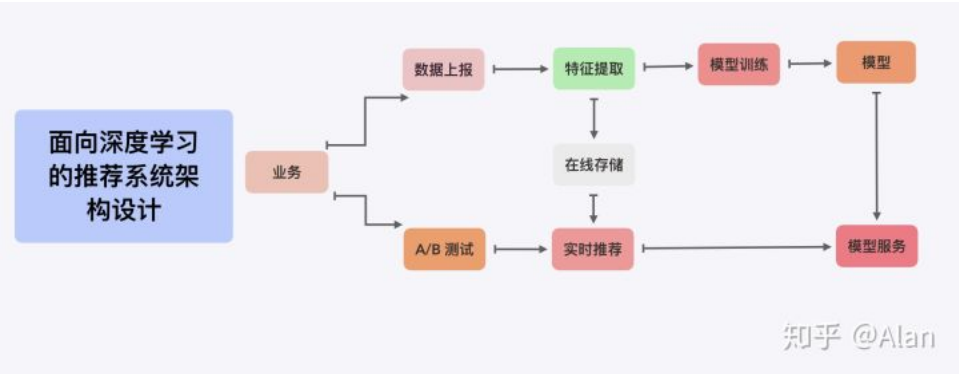
2.1基于离线训练的推荐系统架构设计

- 主要应用场景：手机应用市场、音乐推荐、短视频推荐、资讯、购物
- 常用算法：逻辑回归（Logistics Regression）、梯度提升决策树（GBDT）和因式分解（FM）
- 架构设计：



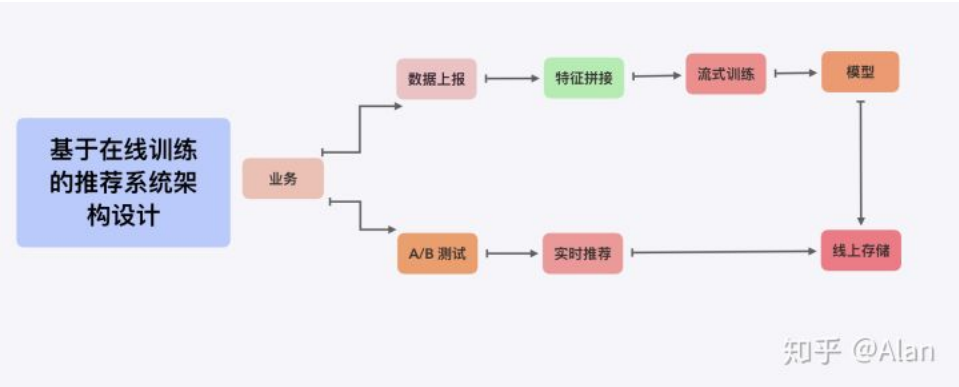
2.2面向深度学习的推荐系统架构设计

- 主要应用场景：图像处理、自然语言理解、语音识别、在线广告等具有时序行为、多维数据的推荐系统
- 常用算法：受限玻尔兹曼（RBM）、自编码器（AE）、卷积神经网络（CNN）、深度神经网络（DNN）、宽度学习（Wide&Deep）等
- 架构设计：



2.3基于在线训练的推荐系统架构设计

- 主要应用场景：广告和电商等高纬度大数据量、实时性要求很高的推荐系统，例如，新闻事件
- 常用算法：FTRL-Proximal、AdPredictor、Adaptive Oline Learning 和PBODL等
- 架构设计：

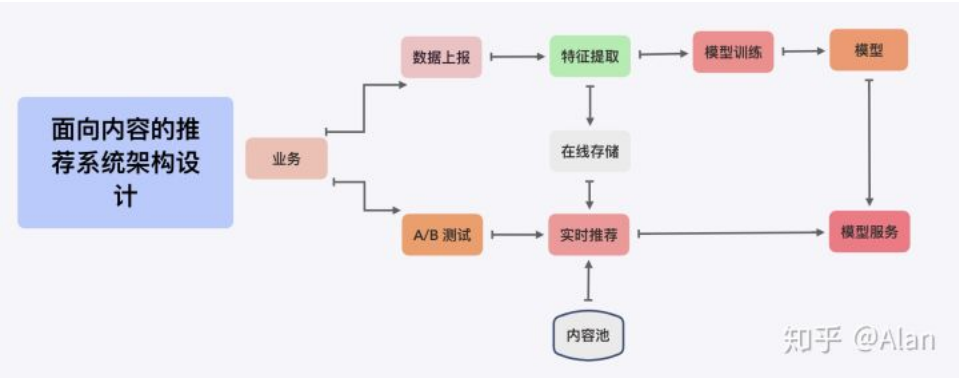


2.4面向内容的推荐系统架构设计



网站等

- 常用算法：NLP自然语言、word2vec、深度学习DNN、OCR等
- 架构设计：



三、推荐系统常用组件

3.1数据上报常用组件

- Apache Kafka 开源流处理平台，分布式实现，通过主题管理。

3.2离线存储常用组件

- HDFS分布式文件系统，Hive是一个基于Hadoop的数据仓库，提供比较完整的SQL功能，使用HDFS作为底层存储，数据规模可达上百PB，并且支持结构化数据。

3.3离线计算常用组件

- Apache Spark是一个基于内存数据处理的高性能分布式计算框架。Spark 提供MapReduce计算模型，Spark使用内存的中间数据结构，使它可以更好的多轮迭代的工作负载。
- TensorFlow是一个开源软件框架，对数据流图进行计算，开发可以变得简单、配置调用接口来进行参数设置。

3.4在线存储常用组件

- Redis是一个基于内存的数据结构的存储系统。
- Memcached是一种高性能的分布式内存缓存系统。
- RocksDB是另外一个高性能的key-value数据库

3.5模型服务常用组件

TensorFlow Serving可以搭建机器学习模型的服务，与TensorFlow模型无缝连接，可自动加载新模型、批量处理请求、可水平扩展等。

3.6实时计算常用组件

Apache Storm是一个开源的分布式实时计算系统。可以与Apache Kafka 开源流处理平台很好适配，常见用于数据实时统计、聚合分析、模型预测等。

Spark Streaming是对Spark核心API的一个扩展，它提供实时数据流的可扩展、高吞吐、高可靠的流处理，方便开发者对流式数据进行处理。

四、效果评估

1、实时性

在基于基于离线训练的推荐系统架构，模型迭代过程需要至少以小时为周期，仅适用于对数据时效性不高的业务场景。



解决方案：

针对这个问题，可以使用基于在线训练的推荐系统架构对特征进行实时提取和实时拼接，这样基本可以作到秒级的特征反馈，适用于捕捉用户的短时兴趣。

2、多样性

问题：

推荐系统多样性缺乏，推荐结果越来越同质化，用户新鲜感降低、很难激发用户购物需求。

解决方案：

针对这个问题，常见用热传导、二次优化、社会网络化等。通过设定合理的相似性、集中指数和覆盖度，来提高推荐结果的多样性和新颖性。

3、评估测试

现实生活中生产系统最常用的评估方法-ABTest，一般先通过方法上下线保留更好的方案，同时不断上线新的召回、排序特征等，迭代优化模型，提升线上效果。

离线算法评估指标：准确率、覆盖度、多样性、新颖性和AUC等。

在线评估指标：点击率、用户停留时间、广告收入等

注意点：短期指标与长期指标相结合，从产品本身、用户体验、业务出发优化推荐系统。

发布于 2019-11-22

「真诚赞赏，手留余香」

赞赏

还没有人赞赏，快来当第一个赞赏的人吧！

[技术架构](#) [推荐系统](#) [推荐算法](#)

文章被以下专栏收录



Data Analyst
一起揭开数据背后不为人知的秘密吧！

关注专栏

推荐阅读

【思考与讨论】Top-K推荐中两种考虑问题的方式

▲ 赞同 135 ▼ 5 条评论 分享 喜欢 收藏 申请转载 ...