

机器学习在电商行业的实践

Machine learning in E - commerce

演讲人：零一

跨界互联
数聚未来

第四届中国数据分析师行业峰会

CHINA DATA ANALYST SUMMIT

北京 中国大饭店 2017.07



零一（陈海城）

- 零一会 创始人
- 杭州沐垚科技有限公司 CEO
- RefLink 数字营销集团 联合创始人
- CDA数据分析研究院 资深讲师
- 历职：老A电商学院讲师部总监、谷点电商运营总监、1024课讲师部总监、狐妖电商运营总监
- 服务客户：苏宁易购、微软、顾家家居、金电联行、ECMS、85度C、浙江师范大学、TGA、特力和乐、百草味、红星美凯龙

- 1.人工智能、机器学习、数据挖掘
- 2.行业应用场景和价值

1.人工智能、机器学习、数据挖掘

腾讯体育5月24日讯 昨天，世界围棋冠军、中国天才围棋选手柯洁对阵人工智能阿法狗，在经过6个小时激烈的鏖战后，终因四分之一的微弱劣势输给了阿法狗，引发了大众广泛讨论和热议。



人工智能英文缩写为AI。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。人工智能就其本质而言，**是对人的思维的信息过程的模拟。**



机器学习：计算机程序如何随着经验积累自动提高性能

数据挖掘：从大量的数据中通过算法搜索隐藏于其中信息的过程

2.行业应用场景和价值

商业的变革是信息流的革命



预测是企业最宝贵的资产



1972年，法国人皮埃尔·瓦克预测了石油危机，让壳牌成为唯一一家能够抵挡这次危机的大石油公司。从此，壳牌公司从世界七大石油公司中最小最丑的一个，一跃成为世界第二大石油公司。

如何应用机器学习来实现商业预测？

业务背景

电商是零售业的渠道之一，传统电商是以采销的形式运作，售罄率是核心指标，只有预测好销量，才能把控住售罄率。而电商有很多活动促销，会严重影响到销量。

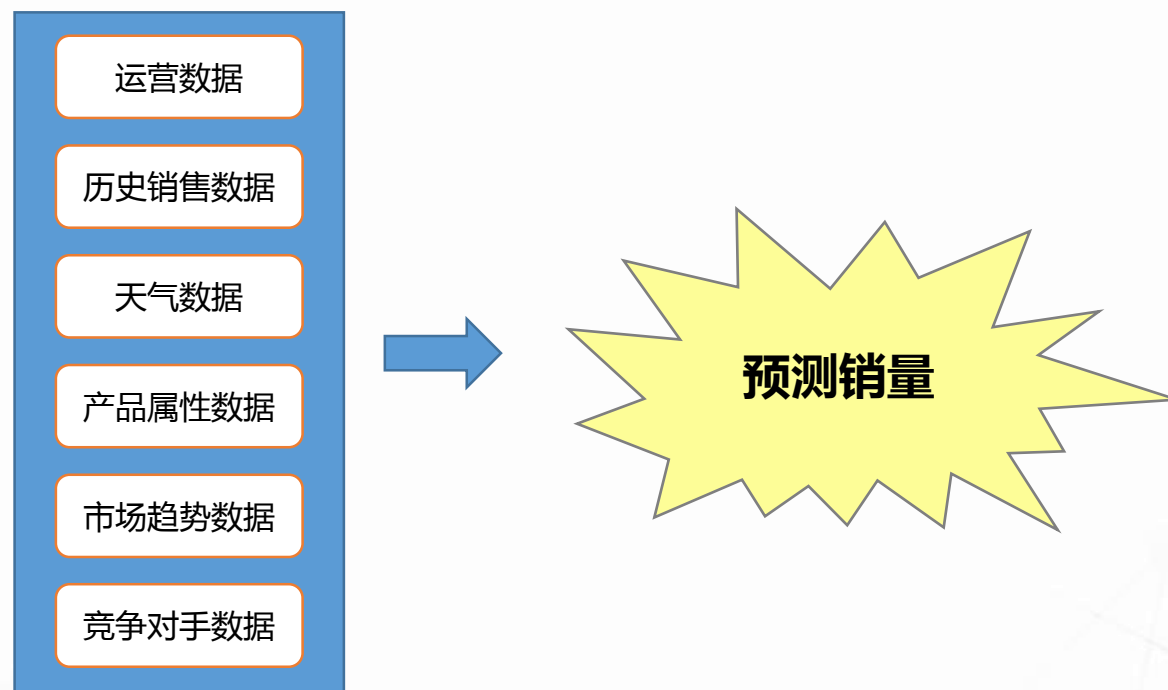
预测难点

- 1.数据本身的噪音多；
- 2.影响因子多（天气、时间、价格、市场投入、促销活动、消费者需求、竞争情况等）；
- 3.涉及到多仓库的分点预测；
- 4.不同品类影响因子不同。

销量预测实施过程

工具：Python

机器学习库：pandas, statsmodels, sklearn



销售预测的机器学习算法实现

1) 时间序列分析法

时间序列分析法是利用变量与时间存在的相关关系，通过对以前数据的分析来预测将来的数据。在分析销售收入时，将销售收入按照年或月的次序排列下来，以观察其变化趋势。

2) 回归分析法

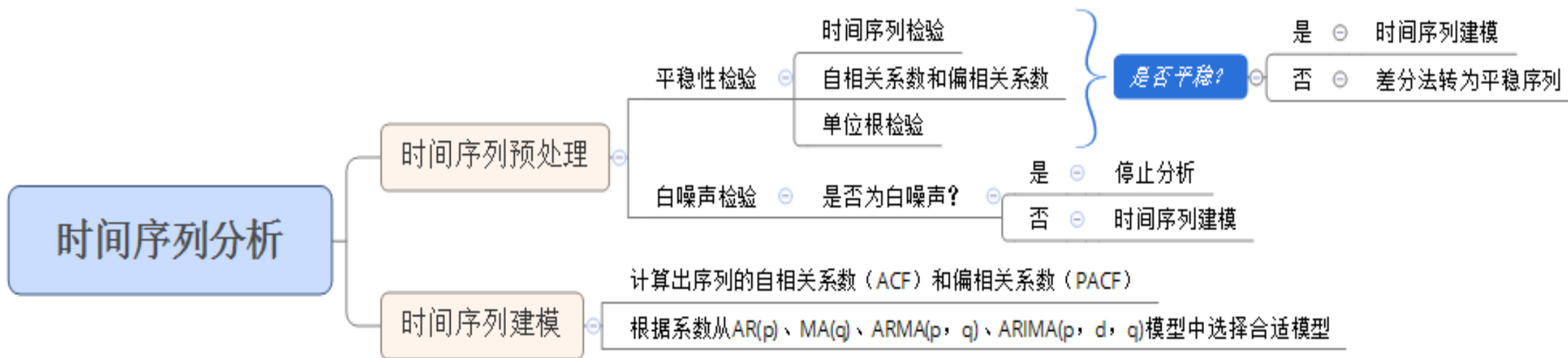
各种事物彼此之间都存在直接或间接的因果关系。同样的，销售量亦会随着某种变量的变化而变化。当销售与时间之外的其他事物存在相关性时，就可运用回归和相关分析法进行销售预测。

销量预测 - 时间序列分析法

工具: Python

机器学习库: pandas, statsmodels

训练集: 历史销售数据



销量预测 - 回归分析法

工具: Python

机器学习库: pandas, sklearn

训练集: 相关特征数据 (右图)



业务背景

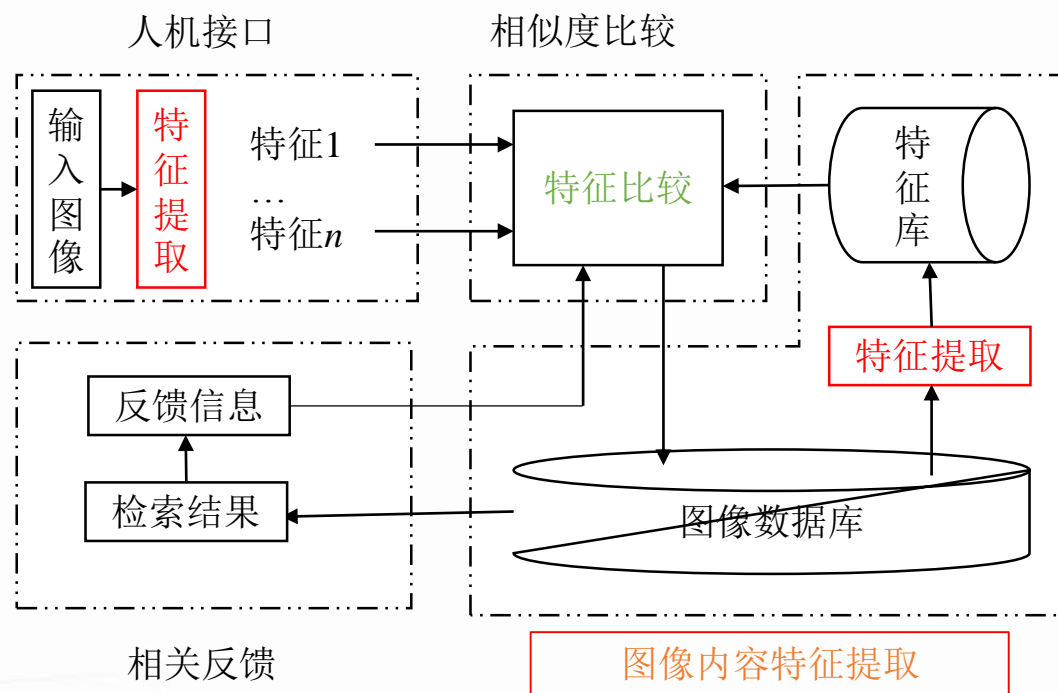
服装已经成为电商最主要的品类之一，其销售额也迅速增长，相应的竞争也越来越激烈。卖家如何快速挖掘热销品类服装的款式、流行的色彩、流行的元素，已经成为商家的新战场。

实施过程：

工具：Python

机器学习包：Scikit-learn、TensorFlow、Caffe、Statsmodels

流程设计：



模型应用

输入类目和产品关键词



输出爆款产品所具有的特征

本模型的应用基于自建的产品特征库，每次输入提取的图片特征都会跟产品特征库进行匹配，并进行自我迭代学习，最终输出符合要求的产品流行特征元素。



CDA 数据分析师
www.cda.cn

THANKS

跨界互联 数聚未来

第四届中国数据分析师行业峰会
CHINA DATA ANALYST SUMMIT