



基于消费者数据的互联网营销

大数据的魅力



经典数据分析案例：啤酒与尿布





Chrome浏览器收集哪些数据？





Google还收集哪些数据？

除了通过Chrome浏览器收集数据，Google还通过其服务收集以下的数据：





Facebook收集哪些数据？



TECH

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Kashmir Hill Former Staff

Welcome to The Not-So Private Parts where technology & privacy collide

Follow

一篇纽约时报的文章指出，当一个女孩怀孕时，美国零售商Target比女孩的父亲先知道这个女孩怀孕。



Target是怎么知道女孩怀孕的？



有人只是去参加了一次北京婚博会，晚上回到家打开微博和微信，发现信息流广告全部变成了婚纱照、婚庆公司、婚礼礼服等。令他感到恐怖的是在此之前从未在手机进行过结婚相关的任何搜索。这一切发生改变的原因仅仅是因为他去了一次婚博会这个地方而已。

有人在知乎看到除甲醛的相关问题，只是百度了一下，结果连一个美食app都开始推荐除甲醛公司。在百度上打开某理财网站，不到半小时推销电话就打过来了。

有时候你在网上搜了一本小說，然后突然就会有很多假网站在百度上显示他们网站有这本小说可以下载，然后让百度把他推在首页，你打开链接一看其实里面没有，但是有其他东西的广告。



即使传统线下商店也在收集你的数据!

问题： 线下商店如何收集个人数据?





讨论：你还知道大数据的哪些应用？



中国银行的对手是谁？



其他国有银行：工商银行，建设银行，农业银行，邮储银行？

股份制银行：招商银行，民生银行，光大银行，浙商银行？

外资银行：汇丰银行，花旗银行，美国银行，渣打银行？

地方银行，城商行，信合，贷款公司...

中国银行的最大对手是谁？



是金融科技（支付宝，微信支付）！

为什么？ 支付宝和微信支付能够取得传统银行无法取得的数据（购物记录，消费记录），从而判断每个人的风险等级，更好的做出借贷决定。而银行往往难以取得这些信息，从而无法与网络支付抗衡。



讨论：什么是大数据？



什么是大数据？



新数据举例：眼球追踪





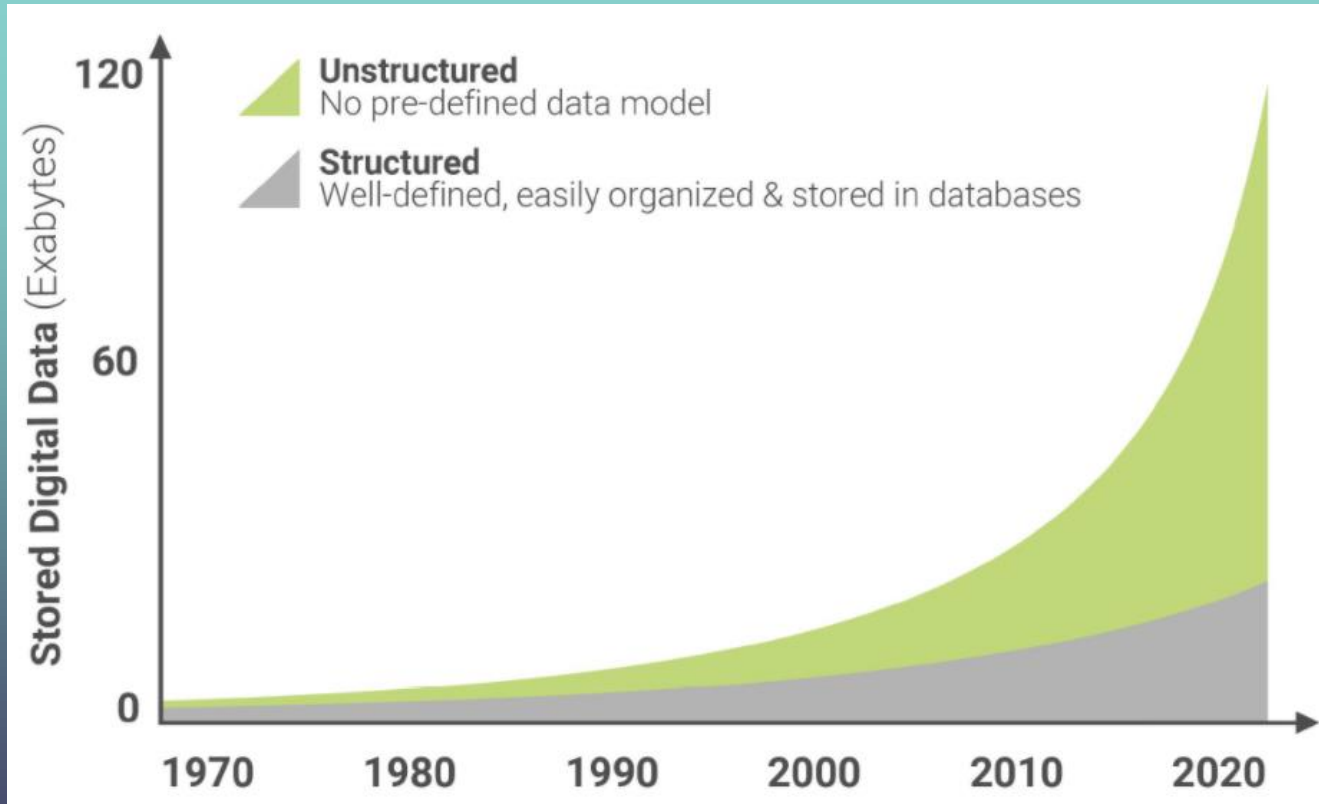
大数据分类

结构化数据：我们传统认知中的“数字”，如销量，利润，产量，浏览量。

非结构化数据：无法用数字规律展示的数据，如顾客评论，视频网站的视频，淘宝顾客上传的照片，眼球追踪数据。

半结构化数据：在一定程度上可以归档分类的数据，如电子邮件，日志文件，XML文档等。







讨论： 如何分析文本数据？





文字分析

最初步的文字分析：我们可以分析文字的长度（例如中文的字数，英文的单词数，句子数等），文字的长度越长，代表文字的内容越丰富。

一项2006年的著名研究(Chevalier and Mayzlin, 2006)分析了网购评论数据，这一研究表明，一个较长的评论往往比一个较短的评论更有说服力，对产品销量的影响往往也更大。



文字分析

进阶的情感分析：我们可以判断文字的极性（正面还是负面），例如

正面：这家酒店地理位置非常好，周围购物十分方便。

负面：这家酒店躲在犄角旮旯里，害的我找了半天。

讨论：如何自动化的判断句子的极性？

文字分析：字典法

进阶的情感分析：为了实现情感分析，我们可以构建两个字典，分别包含褒义词和贬义词。例如：

褒义词：好，方便，便宜，实惠，给力，棒，...

贬义词：差劲，可恶，糟糕，恶心，倒霉，...

当一段话出现的褒义词多而贬义词少，这段话一般也是正面的；当一段话出现的贬义词较多而褒义词少，这段话一般也是负面的。

文字分析

进阶的情感分析：我们可以判断文字的极性（正面还是负面），例如

正面：这家酒店地理位置非常**好**，周围购物十分**方便**。

负面：这家酒店躲在**犄角旮旯**里，**害**的我找了半天。

[这里](#)下载中文情感词典

问题：情感分析有什么用？

情感分析的应用举例：

预测股票收益：我们可以通过分析投资人在网站论坛上的发言，判断大家是否看好一支股票，进而预测这支股票未来的涨跌。

基于情感分析的交易策略：加密对冲基金如何利用AI实现绝对收益能力




问题：情感分析有什么用？

情感分析的应用举例：

判断消费者满意度：我们可以通过分析用户评论，判断哪些消费者对产品有意见，从而积极解决问题，留住消费者。

产品推荐：我们可以分析每个用户对不同电影的好感，判断这个消费者喜欢哪种电影，更好的为消费者进行推荐。此外，电影院线也可以通过分析电影评论，预测电影票房，决定上映哪部电影。



文字分析

高级的文字分析：找出文字对应的话题

这家酒店躲在犄角旮旯里，害的我找了半天。

这家酒店脏兮兮的，是不是床单没洗啊？

上面两个评论都是负面的，但是表达的话题完全不同：第一则评论讨论的是酒店的地理位置，而第二则评论讨论的是酒店的卫生。

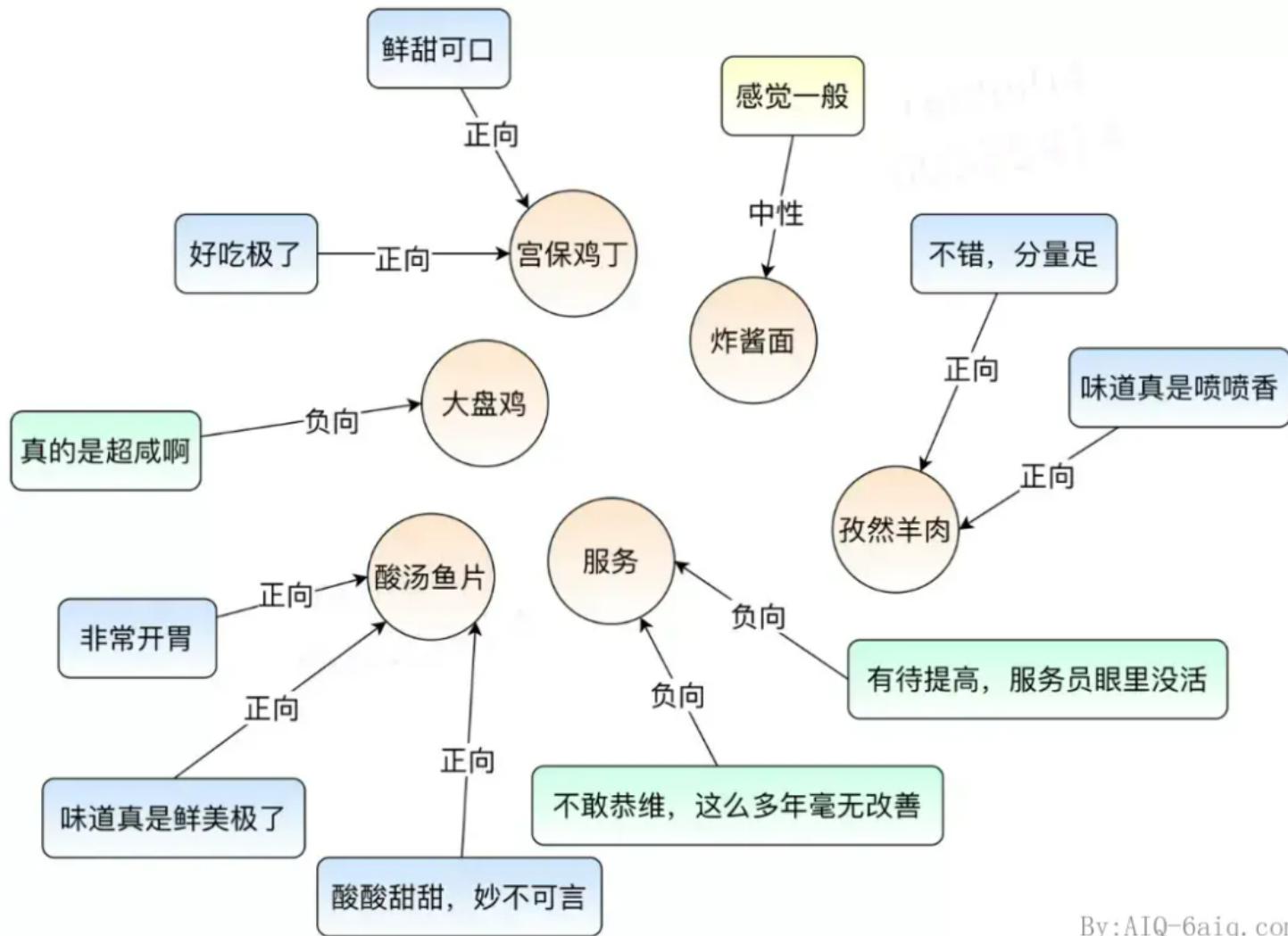
文字分析

在机器学习中，我们将这一问题称为主题模型(Topic Modeling):我们输入不同的文件，通过机器学习算法，将文件划归不同的主题。

我们可以将酒店评论划归价格，位置，登记，房间大小，早餐等

我们可以将餐馆评论划归菜单，停车，装修，上菜时间，服务态度，味道等

通过主题模型，我们可以具体分析消费者对哪一点满意，哪一点不满意，有的放矢的进行改善。



如何进行主题分析？

主题分析是一个艰难的任务。现在的主题模型一般是使用LDA (Latent Dirichlet Allocation 隐狄利克雷分配模型)模型实现的。

你可以点击[这里](#)查看LDA模型的原文

LDA的实现非常复杂，但好消息是：你并不需要从头到尾编程实现。常用的软件如Python和R已经设计好了主题模型包，当你需要使用时直接调用这些程序包即可。



小游戏

随着机器学习技术的发展，我们可以进行越来越复杂的文本分析。

下面这个小游戏使用人工智能帮助我们对对联。





讨论：如何分析图片？



图片是由一个个像素构成的





如何进行图像分析？

我们可以依据图像的像素提取一些基本的数据：

图像的大小（像素数），长宽，方向（横向还是纵向）

图像的色调（深色还是浅色），灰度

但这些数据类似于文字的长度，往往只能无法反应图像的具体内容



如何进行图像分析？

近年来，随着机器学习技术的发展，研究人员开发了一种新的图像分析方法：卷积神经网络(Convolutional Neural Network)。简单来说，卷积神经网络可以判断图像的内容：这个图像究竟是一只猫，一只狗，还是其他？

卷积神经网络的实现同样非常复杂。但是，我们可以利用网上的一些图片分析服务商直接分析图像。[Imagga](#)就是其中的一个例子。



讨论：图像分析在互联网营销有什么用？





图片分析应用举例

这里介绍我的一篇基于图片分析的学术论文：

One Picture is Worth 258 Characters Using Photo Mining to understand
The Role of a Camera in Online Word of Mouth

一张图片相当于258个字：通过图片挖掘来理解照片的在口碑中的作用



背景信息

Yelp是美国的主流餐饮评论网站，类似我国的美团大众点评。在Yelp，用户不仅可以发文字评论和给餐馆打分，也可以上传自己的照片。






研究问题

相对于文字来说，照片是不是更有效果？

照片和文字的关系是什么？有了照片，用户还会看文字吗？

什么样的照片收到用户的欢迎？



研究成果

相对于文字来说，照片是不是更有效果？

平均而言，一张照片的效果相当于258个英文字符。

照片和文字的关系是什么？有了照片，用户还会看文字吗？

有照片时，文字可能更重要。换句话说，文字和照片的结合最有效果。

什么样的照片受到用户的欢迎？

我们将照片分成建筑物，菜单，人像，肉食和蔬菜几类，发现肉食的照片最受欢迎，而建筑物的照片最不受欢迎。



研究的意义

对于平台来讲，我们帮助平台分析哪些评论更有用，平台可以优先将这些评论推荐给新的用户。

对于商家而言，他们可以鼓励用户分享就餐是的照片，并且为用户拍照提供技术指导。

对于用户而言，我们教育他们怎么写出一篇有影响了的评论。



推荐阅读

Balducci and Marinova (2018) *Journal of the Academy of Marketing Science*
Unstructured data in marketing

Tirunillai and Tellis (2014) *Journal of Marketing Research*
Mining marketing meaning from online chatter: Strategic brand analysis of big data using Latent Dirichlet Allocation

Liu, Dzyabura, and Mizik (2020) *Marketing Science*
Visual Listening In: Extracting Brand Image Portrayed on Social Media