

在这里搜索...

首页 今日话题 SEO优化 网站制作 网络营销 运营思维 互联网 职场人生 资源中心

马海祥博客 > 互联网 > 互联网技术 > 今日头条的个性化推荐算法

今日头条的个性化推荐算法

时间：2016-01-20 文章来源：马海祥博客 访问次数：12973 收藏到： 5

互联网给用户带来了大量的信息，满足了用户在信息时代对信息的需求，但也使得用户在面对大量信息时无法从中获得对自己真正有用的那部分信息，对信息的使用效率反而降低了，而通常解决这个问题最常规的办法是推荐系统。

推荐系统能有效帮助用户快速发现感兴趣和高质量的信息，提升用户体验，增加用户使用产品时间，并有效减少用户浏览到重复或者厌恶的信息带来的不利影响，通常推荐系统越精准，用户体验就越好，用户停留时间也会越长，也越容易留住用户。



今日头条在北京国家会议中心举办了“算数·年度数据发布会”，数据发布会的主题名为“算数”，实际上指的是“算法”与“数据”，我也参加了这场发布会，有些收获不敢私藏与大家分享！

众所周知，今日头条是个个性化的新闻推荐引擎，在今日头条CEO张一鸣看来，算法是《今日头条》这款兴趣推荐搜索引擎应用的核心，这也是与传统媒体最本质的区别，今日头条之所以能够非常懂用户，精准推荐出用户所喜好的新闻，完全得益于算法，而正是精准推荐，使得今日头条在短短两年多的时间内拥有了2.2亿用户，每天有超过2000万用户在今日头条上阅读自己感兴趣的文章。

那么今日头条是怎么做到为每一个用户按兴趣精准推荐新闻的呢？在发布会上，今日头条的技术副总裁杨震原给我们揭秘了今日头条的个性化推荐核心“算法”。

1、今日头条的个性化推荐算法原理

如下图，有3篇文章，让3个用户投票[注：这3个用户是一类人，有相同属性（喜好和偏好）]，那第4个用户应该被系统推荐的文章是那篇呢？第4个用户与前3个用户都是一类人，答案显而易见是第一篇文章。

分类目录

SEO新闻	SEO思维
移动端SEO	SEO问答
医疗SEO	淘宝SEO
企业SEO	站外SEO
网站设计	交互设计
网站策划	网页制作
营销策划	营销案例
竞价技巧	数据分析
写作技巧	微信微博
自媒体	新媒体
内容营销	网站运营
O2O模式	App运营
产品运营	网赚教程
创新思维	电子商务
名人访谈	创业故事

热门推荐



基于眼球追踪技术对用户调研的探讨研究



运营思维

更多>>



教你写出提高客户转化率的6个文案策略

立即访问



收集客户关系管理数据的策略和需求分析

立即访问

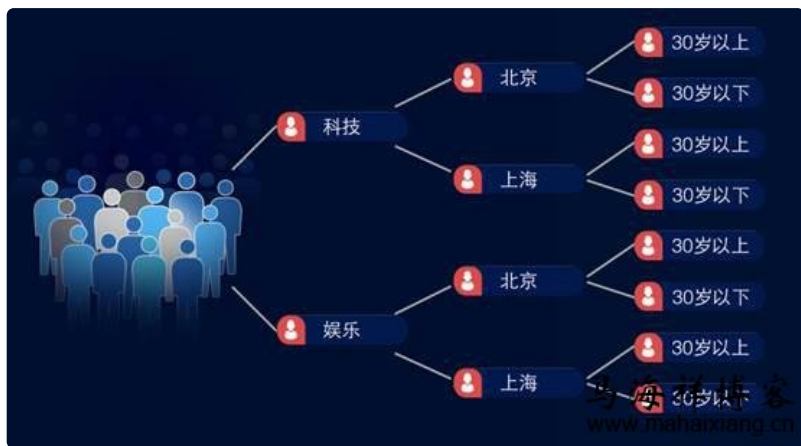


没错，今日头条的个性化推荐算法原理就是基于投票的方法，其核心理念就是投票，每个用户一票，喜欢哪一篇文章就把票投给这篇文章，经过统计，最后得到结果很可能是在这个人群下最好的文章，并把这篇文章推荐给同人群用户过程就是个性化推荐，实际上个性化推荐并不是机器给用户推荐，而是用户之间在互相推荐，看起来似乎很简单，但实际上这需要基于海量的用户行为数据挖掘与分析。

2、今日头条是如何来划分人群和文章

搞技术的朋友都知道，内容推荐必须基于受众人群，只有在确定受众人群之后才能确定推荐内容，有的放矢才叫精准推荐。

上面那个例子只是很小的人群，但是现实生活中不会有这样小的人群，而且太小的人群通过推荐投票的方式也很难推荐出很精细的内容，只能推荐出一些热门内容，那今日头条是如何来划分人群？



我们假定上图这个人群有8万人，只有两类：一类叫喜欢科技的人群；一类喜欢娱乐的人群，我们先分成两类，每类有4万人，我们又找到另一个维度地域：上海和北京，喜欢科技的人有上海和北京的，喜欢娱乐的也有北京和上海的，我们把8万人群分成四组，每组两万，我们找年龄，30岁以上和以下，我们分成8个人群，每个人群1万人，第1个是喜好科技，位置在北京，年龄30岁以上，这有1万人，我们再以下面的喜欢娱乐，地点上海，人群1万。

根据人群不同维度，我们可以细分到一个合理的力度。特别需要注意的是：在细分的领域里面投票数既不要太多，但也不要太少，因为太多没有个性化，仅是一个大众关心的话题而已，而投票数太少又没有统计意义，支持度就低了。



社区O2O兴起的本质与未来发展方向

[立即访问](#)



10个改变未来的科技产品

[立即访问](#)



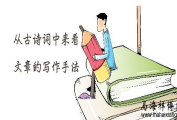
自媒体运营的规范准则

[立即访问](#)



传统企业电商该如何制定网络销售渠道策略

[立即访问](#)



从古诗词中来看文章的写作手法

[立即访问](#)



如何才能写出一篇优质文章?

[立即访问](#)



一个顶尖的产品经理要具备那些能力?

[立即访问](#)



伪原创文章的方法技巧、等级和作用

[立即访问](#)

互联网

[更多>>](#)



基于眼球追踪技术对用户调研的探讨...

眼球追踪技术就是当人的眼睛看向不同方向时，眼部会有细微的变化，这些变化会产生可以提取的特征，计算机可以.....



互联网思维究竟是一种什么样的思维?

但凡做企业的，不管是创业的还是在互联网冲击下转型升级的传统行业企业家，“互联网思维”已经成为了大家共同.....





人群特征可以是阅读姓名，年龄，地域，职业等等，同理，文章也是类似的，文章有提到哪些名人，文章发布的时间，文章所属的地区等等，也都有很多特性，这些都是可以去组合起来。

3、今日头条是怎样判断一个人属于怎样的人群

判断一个用户属于怎样的人群相对简单，比如地域，用户的手机在什么区域，就可以认为用户是什么地域的；比如说用户兴趣，可以根据用户的阅读习惯去判断，用户会经常去看科技的文章，那就可以判断用户属于科技的人群；再比如说用户的好友关系，用户在今日头条上面注册了，用户的好友都是娱乐圈的人，则该用户很有可能也是娱乐圈的人。



所以通过用户在今日头条上行为，以及用户客观的地理位置信息，就能判断用户到底属于一个什么样的人群。

4、今日头条的个性推荐算法公式

精准推荐最难的不是划分人群，也不是判断用户人群归属，更不是文章属性判断，一个人可以属于多个人群，也有多个文章候选，选那个推荐才是最难的，而今日头条的推荐内容是怎么算出来的呢？

如何开启苹果系统的两步验证机制，...

首先，你需要登录至苹果的网页版Apple ID管理系统，你需要点击“管理你的Apple ID”，随后输入帐号密码信息。在登录.....

网络营销

更多>>



图片社交的痛点和定位



腾讯微博为什么会败给新浪微博？



如何在行业中打造个人品牌的影响力



内容营销的方法步骤

网站制作

更多>>



计算机语言的发展简史



CSS常用代码使用技巧大全



2012年网站体验设计趋势回顾

SEO优化

更多>>



如何利用QQ空间做关键词排名
说到QQ空间，只要使用腾讯QQ的都有



通过上图的计算公式: $W1*候选1的投票率+W2*候选2的投票率+W3*候选3的投票率+……=最高分$, 最后能计算出一个得分, 按得分的高低来排序, 就可以得到推荐文章的一个候选, 这个过程实际上是一个比较简单的算法, 而这在今日头条内部叫逻辑回归。

总的来说, 个性化推荐技术本身并不神秘, 归根到底推荐算法关键还是在于对海量用户行为的数据分析与挖掘, 也许各家算法略有不同, 但最终目的都是殊途同归, 为实现最精准的内容推荐而努力中 (对于排序的算法, 大家可查看一下马海洋博客《[九大排序算法的实现方法及算法分析](#)》的相关介绍)。

5、关于推荐算法的三大类

推荐算法大致可以分为三类: 基于内容的推荐算法、协同过滤推荐算法和基于知识的推荐算法。

(1)、基于内容的推荐算法

基于内容的推荐算法的原理是用户喜欢和自己关注过的Item在内容上类似的Item, 比如你看了哈利波特I, 基于内容的推荐算法发现哈利波特II-VI, 与你以前观看的在内容上面 (共有很多关键词) 有很大关联性, 就把后者推荐给你, 这种方法可以避免Item的冷启动问题 (冷启动: 如果一个Item从没有被关注过, 其他推荐算法则很少会去推荐, 但是基于内容的推荐算法可以分析Item之间的关系, 实现推荐)。

弊端在于推荐的Item可能会重复, 典型的新闻推荐, 如果你看了一则关于MH370的新闻, 很可能推荐的新闻和你浏览过的, 内容一致; 另外一个弊端则是对于一些多媒体的推荐 (比如音乐、电影、图片等) 由于很难提内容特征, 则很难进行推荐, 一种解决方案则是人工给这些Item打标签。

(2)、协同过滤算法

协同过滤算法的原理是用户喜欢那些具有相似兴趣的用户喜欢过的商品, 比如你的朋友喜欢电影哈利波特I, 那么就会推荐给你, 这是最简单的基于用户的协同过滤算法 (user-based collaboratlve filtering), 还有一种是基于Item的协同过滤算法 (item-based collaborative filtering), 这两种方法都是将用户的所有数据读入到内存中进行运算

[一个QQ空间, ……](#)

[从百度经验的页面代码结构来解析网站](#)

[最近看到很多的讨论群内都在讨论百度经验平台……](#)

[企业网站的产品页面优化要点](#)

[做网站不在乎规模的大小, 并不是说草根站长就……](#)

[如何做好网页中Meta标签的SEO优化设置](#)

[在做SEO优化的过程中, 网页代码中的Meta标签可以……](#)

[百度排名11位现象的判定特征](#)

[百度排名11位是指你的站点中流量不错的主要关键……](#)

[抓取网站的搜索引擎蜘蛛是不是越多](#)

[不论哪个搜索引擎的爬虫, 来抓取你网站的页面……](#)

[网站页面标题的SEO优化及布局要点](#)

[对于一个刚入行的站长或SEO来说, 首先要搞明白……](#)

[视频推广的最新方法和转化技巧](#)

[随着移动互联网的蓬勃发展, 视频又焕发出新的……](#)

本月热点文章

- 关于大型网站架构的负载均衡技术详解
- HTTPS建设使用的方案教程解析
- 自然语言处理的单词嵌入及表征方法
- 基于贝叶斯推断应用原理的过滤垃圾…
- HTTP、SSL/TLS和HTTPS协议的区…
- 基于高斯模糊原理的模糊图片的研究
- 如何收集和存储服务器运营的数据
- 深入解析互联网协议的原理
- 详解内存数据库中的索引技术
- 基于眼球追踪技术对用户调研的探讨…

标签云

seo 推广 qq 思维
优化 文章写 SEM 团队管理
百 转化 演 wordpress 盟网站 搜索 巧
Jay seo 设计 SEO问题 新手 文 与 降 权
企 响应式设 SMC 社交媒体 技术 法
网 博 手机网站 Java 谷歌
医疗 文案 织梦 交营销 数据 链
腾 网赚 信营销 十

的，因此成为 Memory-based Collaborative Filtering，另一种则是 Model-based collaborative filtering，包括 Aspect Model，pLSA，LDA，聚类，SVD，Matrix Factorization等，这种方法训练过程比较长，但是训练完成后，推荐过程比较快。

(3)、基于知识的推荐算法

最后一种方法是基于知识的推荐算法，也有人将这种方法归为基于内容的推荐，这种方法比较典型的是构建领域本体，或者是建立一定的规则，进行推荐。

混合推荐算法，则会融合以上方法，以加权或者串联、并联等方式尽心融合。

当然，推荐系统还包括很多方法，其实机器学习或者数据挖掘里面的方法，很多都可以应用在推荐系统中，比如说LR、GBDT、RF（这三种方法在一些电商推荐里面经常用到），社交网络里面的图结构等，都可以说是推荐方法。

马海祥博客点评：

个性化推荐的主要思想八个字概括之：物以类聚、人以群分。主要的方法及变种应该有很多，像协同过滤、基于内容的推荐、基于标签的推荐等等，以后推荐系统一个趋势就是基于社交的，利用用户反馈进行实时推荐，多种推荐算法进行混合，从而达到一个较好的推荐结果。

本文发布于马海祥博客文章，如想转载，请注明原文网址摘自于
<http://www.mahaixiang.cn/internet/1418.html>，注明出处；否则，禁止转载；谢谢配合！

打赏

相关标签搜索： 今日头条

上一篇： HTTP与HTTPS的区别

下一篇： HTTP、SSL/TLS和HTTPS协议的区别与联系

相关文章推荐：

- 1 如何开启苹果系统的两步验证机制，避免
- 2 HTTPS建设使用的方案教程解析
- 3 关于大型网站架构的负载均衡技术详解
- 4 自然语言处理的单词嵌入及表征方法
- 5 今日头条的个性化推荐算法
- 6 HTTP、SSL/TLS和HTTPS协议的区别与联系
- 7 基于眼球追踪技术对用户调研的探讨研究
- 8 HTTP服务的七层架构技术解析及运用
- 9 HTTP与HTTPS的区别
- 10 深入解析互联网协议的原理



您可能还会对以下这些文章感兴趣！



如何开启苹果系统的两步验证机制，避免iCloud帐号遭到攻击

首先，你需要登录至苹果的网页版Apple ID管理系统，你需要点击“管理你的Apple ID”，随后输入帐号密码信息。在登录之后，你需要从左侧导航栏中选择“密码和安全”选项，在这里，你将需要验证安全问题，随后下拉至“两步验证”区域，点击蓝色的“开始”链接并阅读其中的.....【[查看全文](#)】

阅读：1407 关键词： 苹果系统 验证机制 icloud攻击 icloud帐号 icloud 日期：2014-09-



计算机的开机启动原理

计算机从打开电源到开始操作，整个启动可以说是一个非常复杂的过程。总体来说，计算机的整个启动过程分成四个阶段：第一阶段：BIOS；第二阶段：主引导记录；第三阶段：硬盘启动；第四阶段：操作系统；直至执行/bin/login程序，跳出登录界面，等待用户输入用户名和密码。.....

[【查看全文】](#)

阅读：3039 关键词：计算机 计算机启动 计算机原理 开机启动原理 日期：2014-01-16

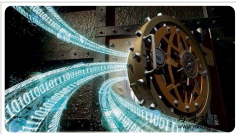


HTTP、SSL/TLS和HTTPS协议的区别与联系

HTTPS是为了安全性而设置的，要验证很多的信息，相对应http请求的速度肯定有点慢，如果使用HTTPS的话很麻烦的，无意给服务器和客户端增加了很大的压力，所以平时最好不要使用HTTPS，如果牵扯到个人隐私或者是其他的什么重要信息就一定要这么做了，很多的时候你感觉有点问题，..... [【查看全文】](#)

文】

阅读：14035 关键词：http ssl https https协议 日期：2016-05-13



详解内存数据库中的索引技术

传统的数据库管理系统把所有数据都放在磁盘上进行管理，所以称作磁盘数据库（DRDB:Disk-Resident Database），磁盘数据库需要频繁地访问磁盘来进行数据的操作，磁盘的读写速度远远小于CPU处理数据的速度，所以磁盘数据库的瓶颈出现在磁盘读写上，基于此，内存数据库的概..... [【查看全文】](#)

阅读：3257 关键词：内存数据库 索引技术 数据库 日期：2015-01-09



基于贝叶斯推断应用原理的过滤垃圾邮件研究

随着电子邮件的应用与普及，垃圾邮件的泛滥也越来越多地受到人们的关注。而目前正确识别垃圾邮件的技术难度非常大。传统的垃圾邮件过滤方法，主要有关键词法和校验码法等。前者的过滤依据是特定的词语；后者则是计算邮件文本的校验码，再与已知的垃圾邮件进行对比。它们..... [【查看全文】](#)

阅读：855 关键词：贝叶斯推断 贝叶斯应用 贝叶斯原理 过滤垃圾邮件 垃圾邮件 日期：



今日头条的个性化推荐算法

互联网给用户带来了大量的信息，满足了用户在信息时代对信息的需求，但也使得用户在面对大量信息时无法从中获得对自己真正有用的那部分信息，对信息的使用效率反而降低了，而通常解决这个问题最常规的办法是推荐系统。推荐系统能有效帮助用户快速发现感兴趣和高质量的信..... [【查看全文】](#)

阅读：12908 关键词：今日头条 日期：2016-01-20



详解大型网站系统的特点和架构演化发展历程

大型网站的挑战主要来自庞大的用户，高并发的访问和海量数据，任何简单的业务一旦需要处理数以P计的数据和面对数以亿计的用户，问题就会变得棘手，大型网站架构主要就是解决这类问题。大型网站不是从无到有一步就搭建好一个大型网站，而是能够伴随小型网站业务的渐进发..... [【查看全文】](#)

阅读：853 关键词：大型网站 网站架构 网站系统 日期：2017-03-02



关于大型网站架构的负载均衡技术详解

负载均衡是将负载（工作任务，访问请求）进行平衡、分摊到多个操作单元（服务器，组件）上进行执行，是解决高性能，单点故障（高可用），扩展性（水平伸缩）的终极解决方案。面对大量用户访问、高并发请求，海量数据，可以使用高性能的服务器、大型数据库，存储设备，高性能W..... [【查看全文】](#)

文】

阅读：809 关键词：大型网站 网站架构 负载均衡 日期：2016-08-05



HTTPS建设使用的方案教程解析

百度已对部分地区开放HTTPS加密搜索服务，随后，百度实行全站化HTTPS安全加密服务，百度HTTPS安全加密已覆盖主流浏览器，旨在用户打造了一个更隐私化的互联网空间、加速了国内互联网的HTTPS化。同时也希望更多网站

加入到HTTPS的队伍中来，为网络安..... [【查看全文】](#)

阅读：42 关键词：seo https 日期：2018-02-01



HTTP服务的七层架构技术解析及运用

一般来说，计算机领域的体系结构普遍采用了分层的方式，从最底层的硬件往高层依次有：操作系统->驱动程序->运行库->系统程序->应用程序等等。从网络分层模型OSI来讲，由上至下为：应用层->表示层->会话层->传输层->网络层->数据链路层->物理层。当然实际应用的TCP/IP协..... [【查看全文】](#)

阅读：4386 关键词：七层架构解析 七层架构运用 七层架构技术 http服务 日期：2014-09-

↓ 点击查看更多 ↓

网站导航

关注微信公众号



SEO优化 网站制作 网络营销 运营思维

SEO新闻 SEO思维 移动SEO 站外SEO 站内SEO
营销策划 竞价技巧 微信微博 内容营销 营销案例
电子商务 O2O模式 App运营 网赚教程 创新思维

关注博主：



Copyright 2012~2020 马海祥博客 (www.mahaixiang.cn) 版权所有 | 网站备案：苏ICP备12048125号 | 站长统计 |
警告：本站禁止镜像、反向代理和采集，文章转载请注明来源网址，否则将追究相关责任。