



# 智能推荐系统-课程简介

讲师：靳彦召



## 课程地位

---

开发语言

人工智能基础

人工智能应用

## 参考教材及学习资料



[https://blog.csdn.net/Magical\\_Bubble/article/details/88371289](https://blog.csdn.net/Magical_Bubble/article/details/88371289)

# 参考教材及学习资料



<https://zhuanlan.zhihu.com/wangzh-enotes>

# 考核方式

- 考核方式
  - 平时成绩：60%
    - 平时表现：12%（课堂提问、课堂纪律、课堂出勤）
    - 雪梨平时作业：48%
  - 竞赛成绩：40%
- 雪梨作业地址：<http://www.edu2act.cn/team/2018-ji-zhi-neng-tui-jian/>



## 第一章

# 推荐系统概述



# 教学目标

- 了解推荐系统
- 推荐系统的应用
- 推荐系统的评测指标和评测维度

## 1、什么是推荐系统

3、推荐系统在电商、社交、音乐、视频、阅读、服务等产品中的应用

5、好的推荐系统应该有哪些评测指标和评测维度

2、推荐系统的作用和意义

4、推荐系统的架构

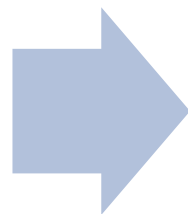




## 1.1 什么是推荐系统

- 如何找到自己想要的商品?
  - 如何挑选一台电视?
  - 如何挑选一部电影?

信息匮乏



信息过载

## 1.1 什么是推荐系统

- 为了解决信息过载问题
  - 分类目录：雅虎、DMOZ、Hao123
  - 搜索引擎：谷歌、百度
- 推荐系统也是一种帮助用户快速发现有用信息的工具。和搜索引擎不同的是，推荐系统不需要用户提供明确的需求，而是通过分析用户的历史行为给用户的兴趣建模，从而主动给用户推荐能够满足他们兴趣和需求的信息

## 1.1 什么是推荐系统

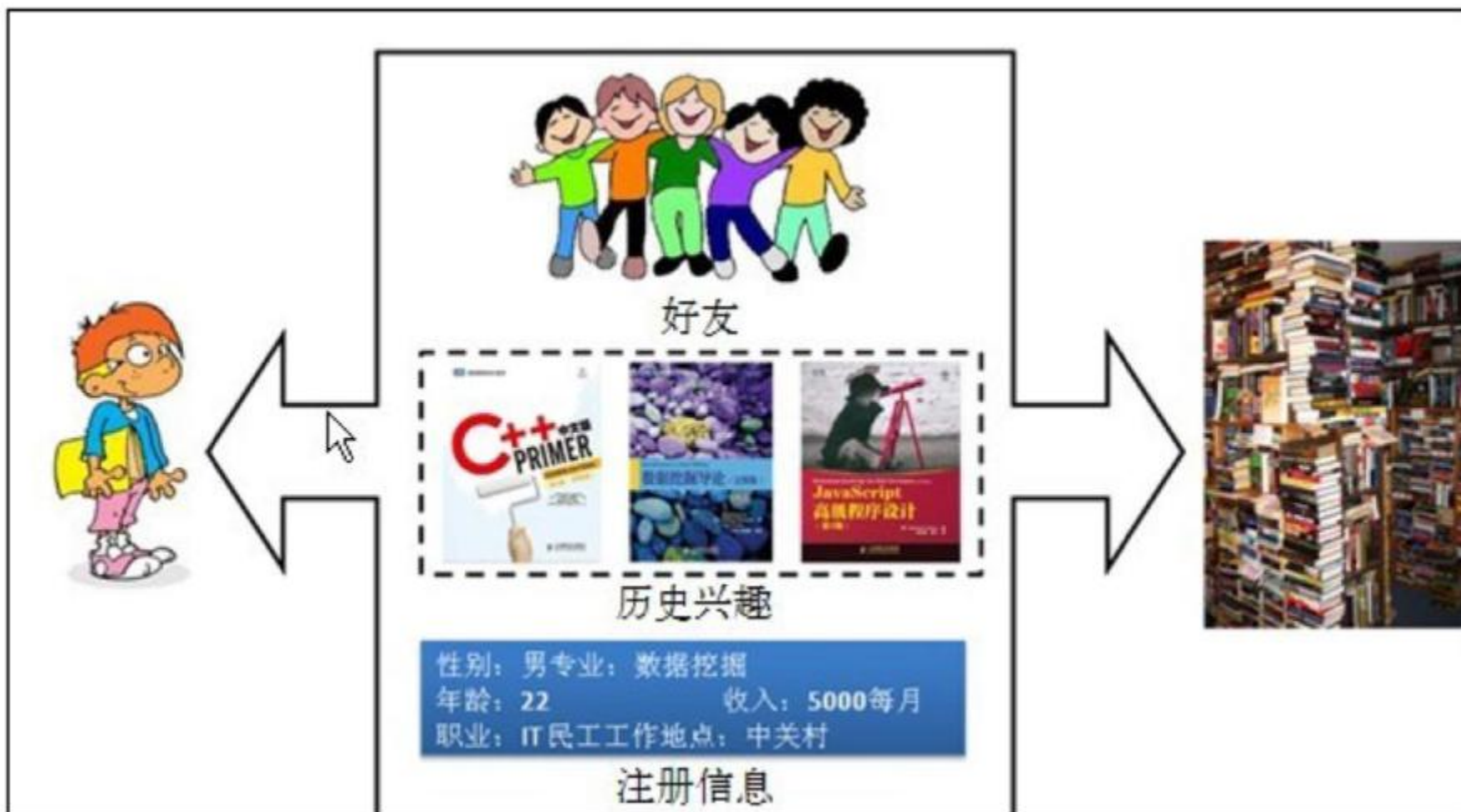


## 1.1 什么是推荐系统

- 现实社会中用户面对很多选择时做决定的过程：
  - 向朋友咨询
  - 个人喜好：喜欢的演员和导演
  - 综合排名

## 1.1 什么是推荐系统

- 推荐算法的本质是通过一定的方式将用户和物品联系起来，而不同的推荐系统利用了不同的方式。



# 目录

1、什么是推荐系统

3、推荐系统在电商、社交、音乐、视频、阅读、服务等产品中的应用

5、好的推荐系统应该有哪些评测指标和评测维度

2、推荐系统的作用和意义

4、推荐系统的架构



## 1.2 推荐系统的作用和意义

**用户角度：**推荐系统解决在“信息过载”的情况下，用户如何高效获得感兴趣信息的问题

**公司角度：**推荐系统解决产品能够最大限度地吸引用户、留存用户、增长用户黏性、提高用户转化率的问题，从而达到公司商业目标连续增长的目的。

# 目录

1、什么是推荐系统

3、推荐系统在电商、社交、音乐、视频、阅读、服务等产品中的应用

5、好的推荐系统应该有哪些评测指标和评测维度

2、推荐系统的作用和意义

4、推荐系统的架构





## 1.3 推荐系统的应用

- 和搜索引擎不同，个性化推荐系统需要依赖用户的行为数据，因此一般都是作为一个应用存在于不同网站之中。在互联网的各类网站中都可以看到推荐系统的应用，而个性化推荐系统在这些网站中的主要作用是通过分析大量用户行为日志，给不同用户提供不同的个性化页面展示，来提高网站的点击率和转化率。
- 广泛利用推荐系统的领域包括电子商务、电影和视频、音乐、社交网络、阅读、基于位置的服务、个性化邮件和广告等。
- 推荐系统的组成：
  - 前台的展示页面
  - 后台的日志系统
  - 推荐算法系统

## 1.3 推荐系统在电商中的应用

- 电子商务网站是个性化推荐系统的一大应用领域。著名的电子商务网站亚马逊是个性化推荐系统的积极应用者和推广者。
- 最主要的应用有个性化商品推荐列表和相关商品的推荐列表。

### Today's Recommendations For You

Here's a daily sample of items recommended for you. Click here to [see all recommendations](#).

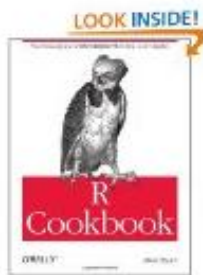
Page 6 of 44 (Start over)



[Networks, Crowds, and Markets: R...](#) (Hardcover) by David Easley

★★★★☆ (5) \$39.50

[Fix this recommendation](#)



[R Cookbook \(O'Reilly Cookbooks\)](#) (Paperback) by Paul Teetor

★★★★☆ (11) \$31.49

[Fix this recommendation](#)



[Introduction to Machine Learning...](#) (Hardcover) by Ethem Alpaydin

★★★★☆ (6) \$38.00

[Fix this recommendation](#)



[Programming Collective Intelligence...](#) (Paperback) by Toby Segaran

★★★★☆ (72) \$26.39

[Fix this recommendation](#)

## 1.3 推荐系统在电商中的应用

个性化推荐列表采用了一种基于物品的推荐算法 (item-based method)，该算法给用户推荐那些和他们之前喜欢的物品相似的物品。



The screenshot displays the Amazon.com interface with a 'Recommended for You' section. The top recommendation is the book 'Introduction to Machine Learning (Adaptive Computation and Machine Learning series)' by Ethem Alpaydin. It shows the book cover, title, author, price (\$46.33), and options to 'Add to Cart' or 'Add to Wish List'. To the right of the book details, there is a feedback section titled 'Rate this item' with a star rating (5 stars) and two checkboxes: 'I own it' and 'Not interested'. A red arrow points from the text '对推荐结果的反馈方式' to the 'Not interested' checkbox. Below this, the 'Because you rated...' section shows two more book recommendations: 'Probabilistic Graphical Models: Principles and Techniques (Adaptive Computation and Machine Learning series)' by Daphne Koller and Nir Friedman, and 'Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)' by Ian H. Witten and Eibe Frank. Each recommendation includes a book cover, title, author, and a star rating. The 'Don't use for recommendations' checkbox is visible for both books in this section. The Amazon.com logo and 'Help | Close window' links are also present.

amazon.com<sup>®</sup> [Help](#) | [Close window](#)

**Recommended for You**

**LOOK INSIDE!**

**Introduction to Machine Learning (Adaptive Computation and Machine Learning series)**  
by Ethem Alpaydin (Author)  
**Our Price: \$46.33**  
**Used & new** from \$31.97

[Add to Cart](#) [Add to Wish List](#)

**Rate this item**  
[X] ★★★★★  
☐ I own it  
☐ Not interested

**Because you rated...**

**Probabilistic Graphical Models: Principles and Techniques (Adaptive Computation and Machine Learning series)**  
(Hardcover)  
by Daphne Koller (Author), Nir Friedman (Author)  
[X] ★★★★★  
☐ Don't use for recommendations

**Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)**  
(Paperback)  
by Ian H. Witten (Author), Eibe Frank (Author)  
[X] ★★★★★  
☐ Don't use for recommendations

[Help](#) | [Close window](#)

## 1.3 推荐系统在电商中的应用

另一种个性化推荐方式：利用Facebook的好友关系，给用户推荐他们的好友在亚马逊上喜欢的物品。

### Popular Among Your Friends on Facebook

Page 1 of 12





The Beatles 1 ~ The Beatles

★★★★☆ (1,189) \$13.88

7 friends like this:



> [See more](#)



The Matrix DVD ~ Keanu Reeves

★★★★☆ (3,019) \$5.49

6 friends like this:



> [See more](#)



NINETEEN EIGHTY-FOUR by George Orwell

★★★★☆ (1,668)

4 friends like this:



> [See more](#)



Michael ~ Michael Jackson

★★★★☆ (228) \$13.35

6 friends like this:



> [See more](#)



Harry Potter Hardcover Boxed Se... by J. K. Rowling

★★★★☆ (857) \$114.07

4 friends like this:



> [See more](#)





## 1.3 推荐系统在电商中的应用

亚马逊另一个重要的推荐应用是**相关推荐列表**：当你在亚马逊购买一个商品时，它会在商品信息下面展示相关的商品。

亚马逊有两种相关商品列表，一种是购买了这个商品的用户也经常购买的其他商品。

Customers Who Bought This Item Also Bought Page 1 of 17

					
CaseCrown Faux Suede Case (Navy) for Pandigital Novel Digital Bo...	HDE 7\" data-bbox="205 631 321 731"/>	3 Pack of Universal Touch Screen Stylus Pen (Red +... by Bargaincell	CaseCrown Faux Suede Case (Red) for Pandigital Novel Digital Boo...	3 Pack of Premium Crystal Clear Screen Protectors for Apple iPad	CaseCrown Faux Suede Case (Fuschia) for Pandigital Novel Digital...
★★★★☆ (4)	★★★★☆ (15)	★★★★☆ (636)	★★★★☆ (5)	★★★★☆ (539)	★★★★☆ (9)
\$5.21	\$15.48	\$1.99	\$5.21	\$1.95	\$11.21

## 1.3 推荐系统在电商中的应用

另一种是浏览过这个商品的用户经常购买的其他商品。

### What Other Items Do Customers Buy After Viewing This Item?



3 Pack of Universal Touch Screen Stylus Pen (Red + Black + Silver) by Bargaincell

★★★★☆ (636)

\$1.99



Summer of Fire (Yellowstone series) by Linda Jacobs Kindle Edition

★★★★☆ (5)

\$0.00

SUPERSALE.99



USA SELLER

New 7 inch android 2.2 VIA8650 epad tablet pc 3.0MP Camera, SUPPORTS EXTERNAL3G MODEM+WIFI  
Screen+Flash 10.1 with PRIORITY SHIPPING FROM SUPERSALE.99 USA SELLER by SuperPad

★★★★☆ (28)

\$89.99



SKYTEX Skypad Alpha 7" Touch Screen Cortex-A8 Tablet Android OS 2.3 by SKYTEX

★★★★☆ (6)

\$150.92

> [Explore similar items](#)


## 1.3 推荐系统在电商中的应用

相关推荐列表最重要的应用就是**打包销售**（cross selling）。

### Frequently Bought Together



**Price For All Three: \$108.05**

 Add all three to Cart

Add all three to Wish List

[Show availability and shipping details](#)

- ☒ **This item:** R Cookbook (O'Reilly Cookbooks) by Paul Teetor Paperback **\$25.95**
- ☒ R in a Nutshell: A Desktop Quick Reference by Joseph Adler Paperback **\$35.99**
- ☒ ggplot2: Elegant Graphics for Data Analysis (Use R) by Hadley Wickham Paperback **\$46.11**

## 1.3 推荐系统在电商中的应用

- 这些推荐的应用，究竟给亚马逊带来了多少商业利益？

**亚马逊有20% ~ 30%的销售来自于推荐系统。**

搜狗实验室：“如果我们能为搜狗的用户推荐更合适的广告，让广告的点击率增长1%，就能为公司增加上千万的利润。”

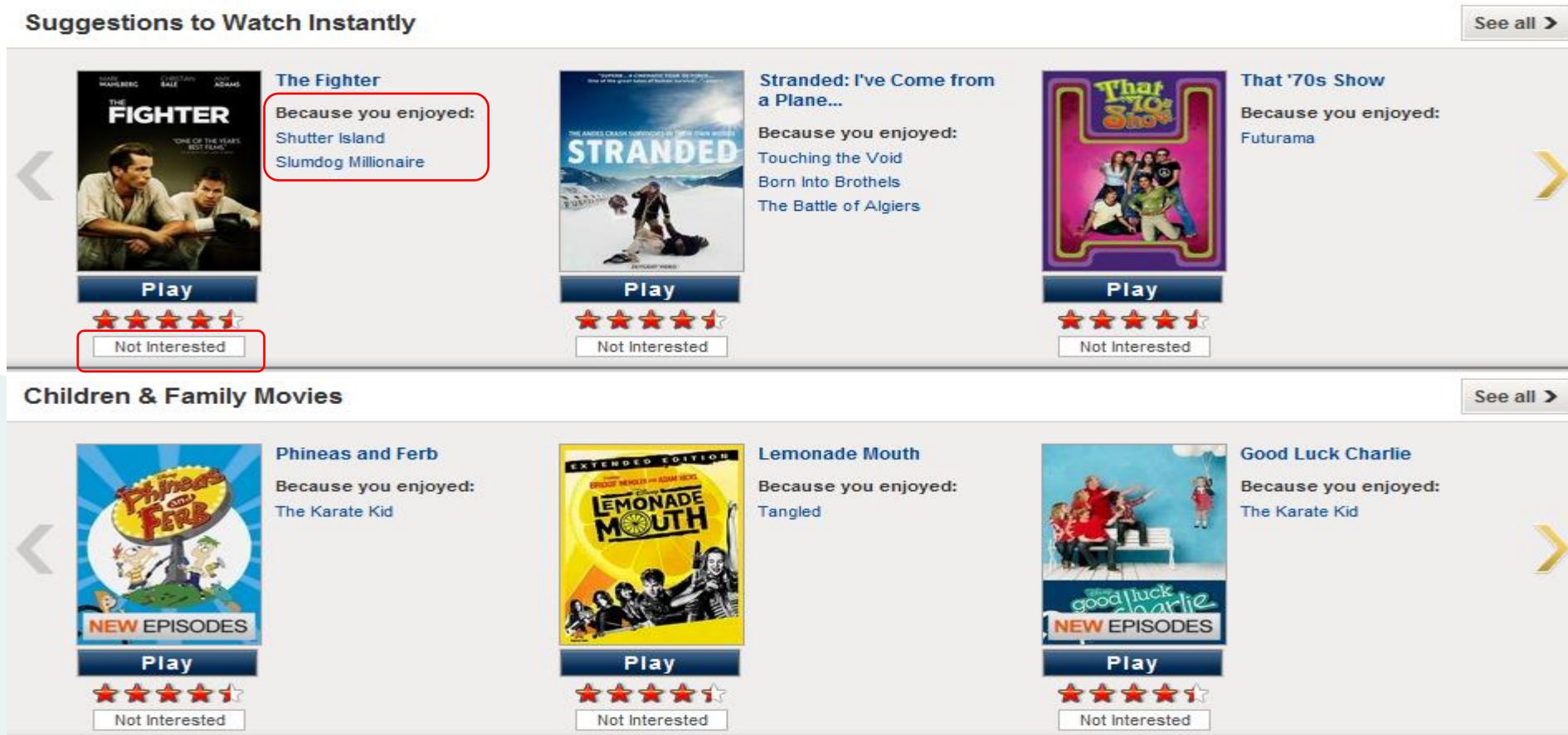


## 1.3 推荐系统在视频类网站中的应用

- 在电影和视频网站中，个性化推荐系统也是一种重要的应用。在该领域成功使用推荐系统的一家公司就是Netflix，它和亚马逊是推荐系统领域最具代表性的两家公司。

## 1.3 推荐系统在视频类网站中的应用

- 电影的标题和海报。
- 用户反馈模块——包括Play（播放）、评分和Not Interested（不感兴趣）3种。
- 推荐理由——因为用户曾经喜欢过别的电影



## 1.3 推荐系统在视频类网站中的应用

多少用户使用了Netflix的推荐系统？

从Netflix的推荐理由来看，它们的算法和亚马逊的算法类似，也是基于物品的推荐算法，即给用户推荐和他们曾经喜欢的电影相似的电影。至于推荐系统在Netflix中起到的作用，Netflix在宣传资料中宣称，有60%的用户是通过其推荐系统找到自己感兴趣的电影和视频的。

## 1.3 推荐系统在视频类网站中的应用


YouTube和Hulu也有自己的个性化推荐页面。

YouTube曾经做过一个实验，比较了个性化推荐的点击率和热门视频列表的点击率，实验结果表明个性化推荐的点击率是热门视频点击率的两倍。

展示推荐结果时也提供了视频标题、缩略图、视频的平均分、推荐理由和用户反馈模块。


Video Type: TV Shows & Movies Filters: ☐ Captions ☐ Free

Top Recommendations 1 of 4




**Brooklyn Rules**  
Recommended because you have shown interest in One Eyed King

Are you interested?  
  [I've seen it](#)




**Morlang**  
Recommended because you have shown interest in Adrift in Manhattan

Are you interested?  
  [I've seen it](#)



**Fatal Memories**  
Recommended because you bookmarked Canvas (2006)


Are you interested?  
  [I've seen it](#)



**The Only Thrill**  
Recommended because you bookmarked Love & Rage


Are you interested?  
  [I've seen it](#)

Science Fiction 1 of 4




**And You Thought Your Parents Were Weird!**  
Recommended because you have shown interest in Trading Mom

Are you interested?  
  [I've seen it](#)




**The Final Cut**  
Recommended because it's highly rated by Hulu users


Are you interested?  
  [I've seen it](#)



**From Beyond**  
Recommended because it's highly rated by Hulu users

Are you interested?  
  [I've seen it](#)



**Caprica**   
Recommended because it's highly rated by Hulu users

Are you interested?  
  [I've seen it](#)

## 1.3 推荐系统在音乐类网站中的应用

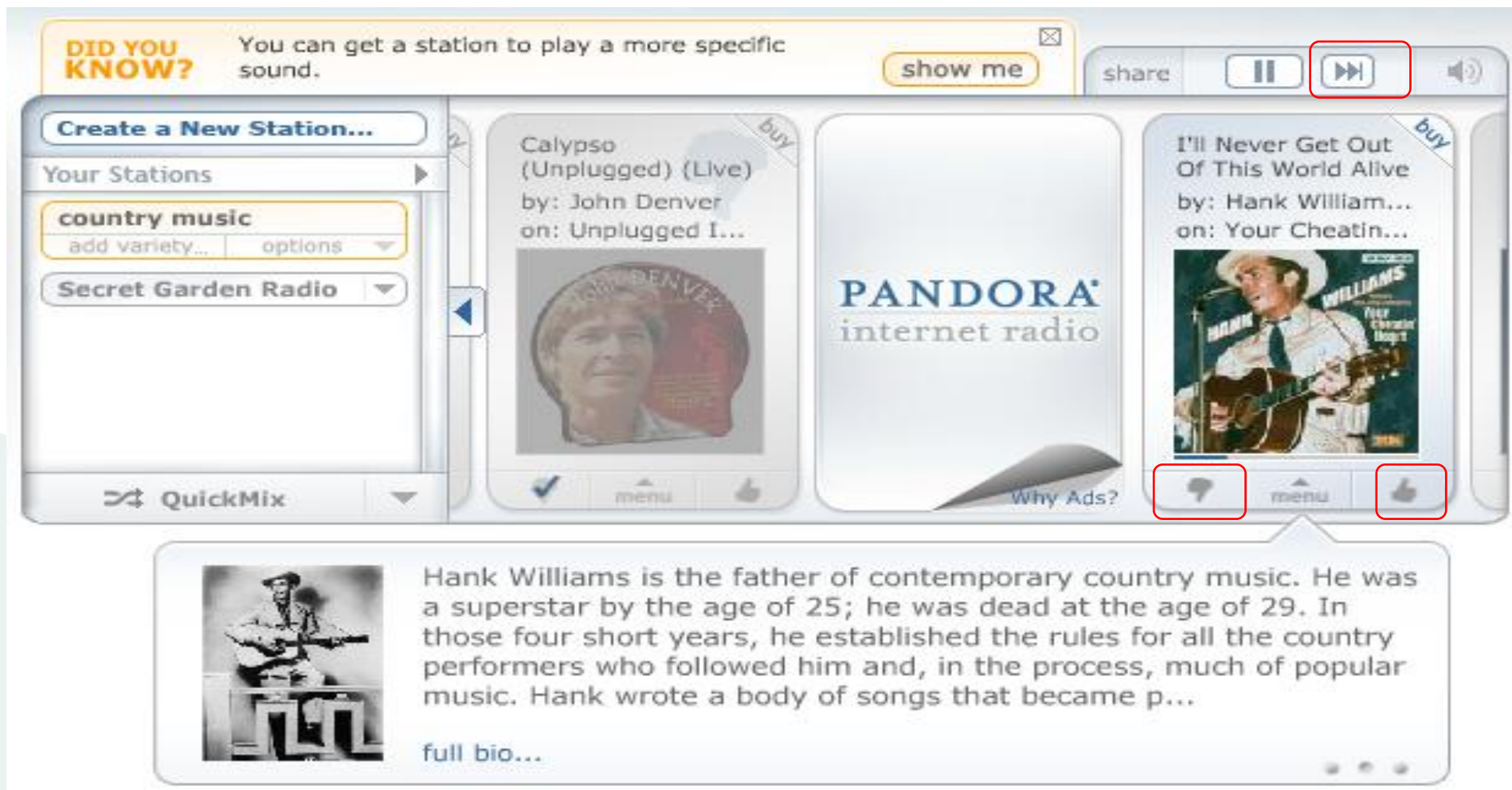
个性化推荐的成功应用需要两个条件。**第一是存在信息过载**，因为如果用户可以很容易地从所有物品中找到喜欢的物品，就不需要个性化推荐了。**第二是用户大部分时候没有特别明确的需求**，因为用户如果有明确的需求，可以直接通过搜索引擎找到感兴趣的物品。

在这两个条件下，个性化网络电台无疑是最合适的个性化推荐产品。首先，音乐足够多（信息过载）；其次，很多时候用户是把音乐作为背景音乐来听（没有明确需求）。



## 1.3 推荐系统在音乐类网站中的应用

国际上著名的个性化音乐网络电台Pandora



## 1.3 推荐系统在音乐类网站中的应用

国际上著名的个性化音乐网络电台Last.fm



## 1.3 推荐系统在音乐类网站中的应用

国内著名的个性化音乐网络电台豆瓣FM





## 1.3 推荐系统在音乐类网站中的应用

从前端界面上看，这3个个性化网络电台很类似。它们都不允许用户点歌，而是给用户几种反馈方式——喜欢、不喜欢和跳过。经过用户一定时间的反馈，电台就可以从用户的历史行为中习得用户的兴趣模型，从而使用户的播放列表越来越符合用户对歌曲的兴趣。

Pandora背后的音乐推荐算法主要来自于一个叫做音乐基因工程的项目。由音乐家和工程师对歌曲进行标注，Pandora会根据专家标注的基因计算歌曲的相似度。因此Pandora的算法主要是**基于内容**。

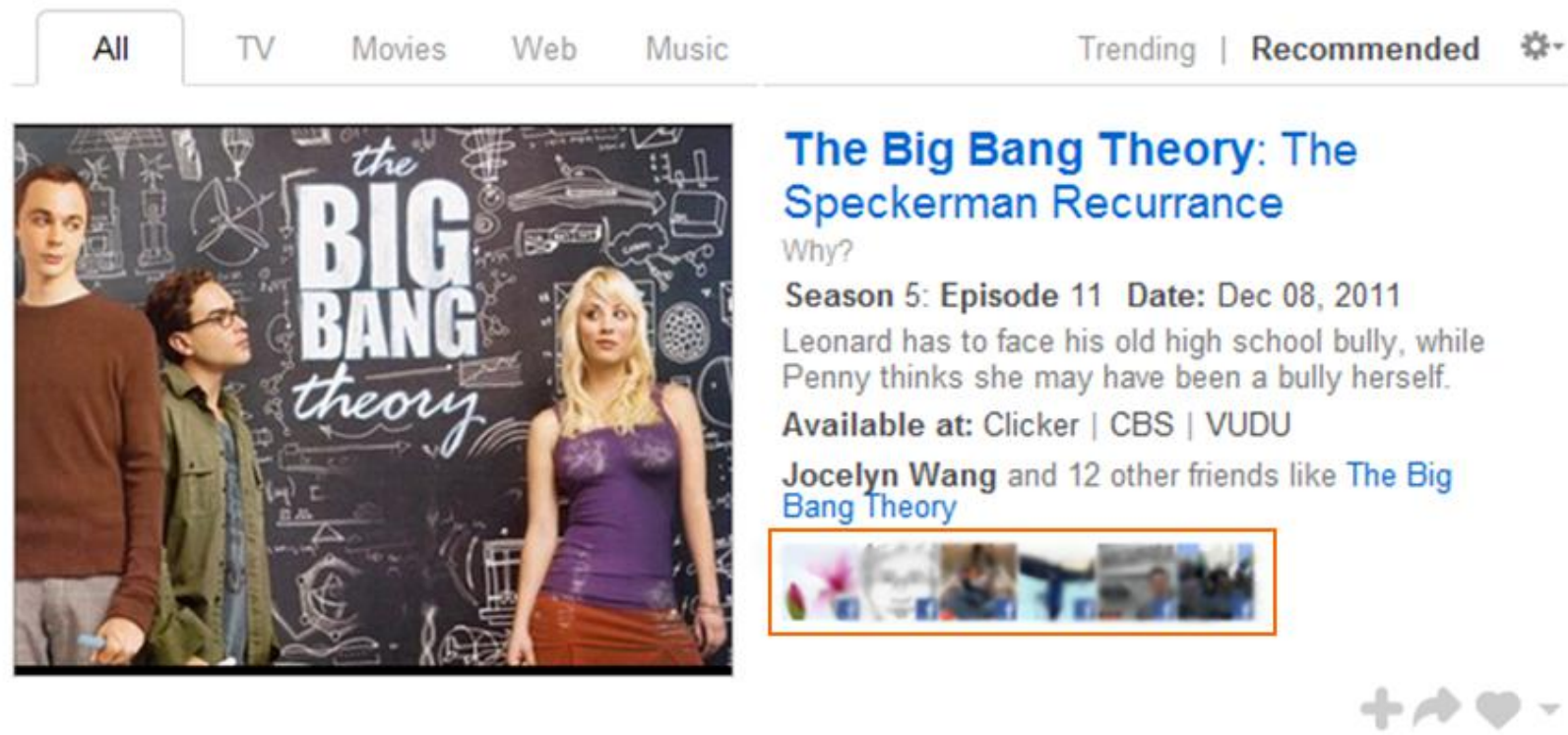
Last.fm记录了所有用户的听歌记录以及用户对歌曲的反馈，在这一基础上计算出不同用户在歌曲上的喜好相似度，从而给用户推荐和他有相似听歌爱好的其他用户喜欢的歌曲。Last.fm没有使用专家标注，而是主要**利用用户行为计算歌曲的相似度**。

## 1.3 推荐系统在社交网络中的应用

社交网络中的个性化推荐技术主要应用在3个方面：

1、利用用户的社交网络信息对用户进行个性化的物品推荐；

What's On For Xiang: Tuesday, December 13, 2011



The screenshot shows a Facebook interface for a user named 'Xiang'. At the top, there are tabs for 'All', 'TV', 'Movies', 'Web', and 'Music'. The 'Recommended' tab is selected. Below the tabs, there is a recommendation for 'The Big Bang Theory: The Speckerman Recurrence'. The recommendation includes a thumbnail image of the show's main cast members (Sheldon, Leonard, Penny, and Howard) in front of a chalkboard. To the right of the image, the title 'The Big Bang Theory: The Speckerman Recurrence' is displayed in blue. Below the title, there is a 'Why?' section with a brief description: 'Season 5: Episode 11 Date: Dec 08, 2011 Leonard has to face his old high school bully, while Penny thinks she may have been a bully herself.' Below the description, it says 'Available at: Clicker | CBS | VUDU'. At the bottom of the recommendation, it says 'Jocelyn Wang and 12 other friends like The Big Bang Theory'. Below this text is a small horizontal strip of image thumbnails. At the bottom right of the recommendation, there are icons for '+', a share icon, a heart icon, and a dropdown arrow.

著名的电视剧推荐网站Clicker使用Facebook的Instant Personalization给用户进行个性化视频推荐的界面

## 1.3 推荐系统在社交网络中的应用

### 2、信息流的会话推荐



社交网站本身利用社交网络给用户推荐其他用户在社交网站的会话。每个分享和它的所有评论被称为一个会话，如何给这些会话排序是社交网站研究中的一个重要话题。

Facebook开发了EdgeRank算法对这些会话排序，使用户能够尽量看到熟悉的好友的最新会话。

## 1.3 推荐系统在社交网络中的应用

### 3、给用户推荐好友

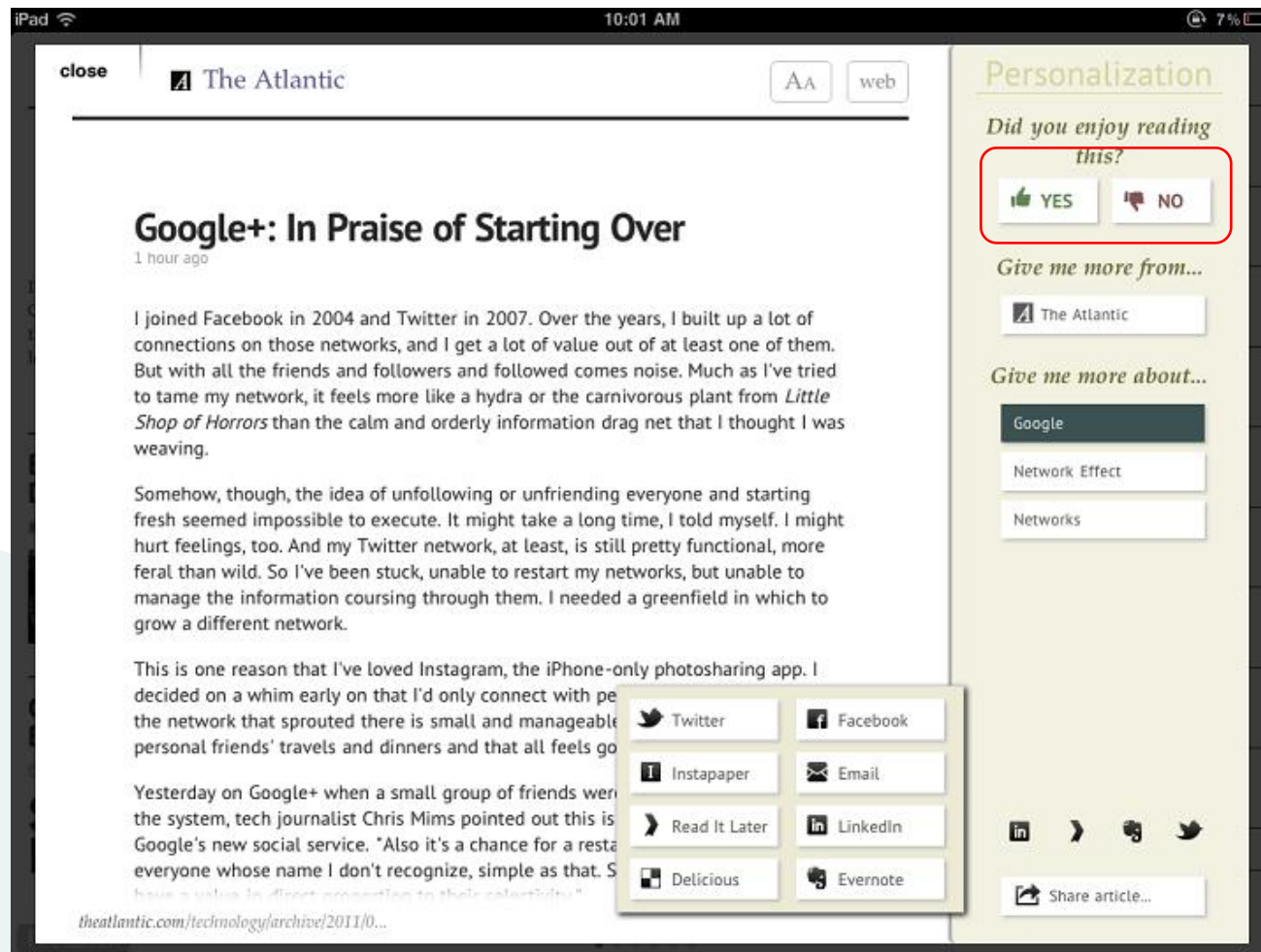


不同社交网站中好友推荐系统的界面（左上为Facebook，左下为Twitter，右上为LinkedIn，右下为新浪微博）



## 1.3 推荐系统在阅读类产品中的应用

个性化阅读工具Zite：收集用户对文章的偏好信息。在每篇文章右侧，Zite都允许用户给出喜欢或不喜欢的反馈，然后通过分析用户的反馈数据不停地更新用户的个性化文章列表。



## 1.3 推荐系统在广告领域中的应用

- 广告是互联网公司生存的根本。很多互联网公司的盈利模式都是基于广告的，而广告CPC、CPM直接决定了很多互联网公司的收入。
- 个性化广告投放目前已经成为了一门独立的学科——计算广告学
- 个性化推荐着重于帮助用户找到可能令他们感兴趣的物品，而广告推荐着重于帮助广告找到可能对它们感兴趣的广告用户，即一个是以用户为核心，而另一个以广告(物品)为核心。
- 目前的个性化广告投放技术主要分为3种。
  - 上下文广告 通过分析用户正在浏览的网页内容，投放和网页内容相关的广告。代表系统是谷歌的AdSense。
  - 搜索广告 通过分析用户在当前会话中的搜索记录，判断用户的搜索目的，投放和用户目的相关的广告。
  - 个性化展示广告 我们经常在很多网站看到大量展示广告（就是那些大的横幅图片），它们是根据用户的兴趣，对不同用户投放不同的展示广告。雅虎是这方面研究的代表。

## 1.3 推荐系统在广告领域中的应用

地区

国家 [?]  
美国 x

☐ 所有地区

☒ 按省 [?]

☐ 按城市 [?]

☐ By Zip Code [?]

California x

人数统计

年龄 [?]  
20 - 任意

☐ Require exact age match [?]

性别 [?]  
☐ 全部 ☒ 男性 ☐ 女性

Interests

Broad Category [?]

Family Status

Interests

Mobile 1

Movie/Film

Music

Retail/Shopping

Sports

☐ Mobile (All)

☐ Android

☒ iPad

☐ iPhone

☐ 其他

☐ RIM/Blackberry

☐ Windows Phone

已选择2个种类 · Show Selected

Estimated Reach [?]

541,800 个人

- 居住在**美国**
- 住在 **California**
- age **20** and older
- 是**男性**
- who are in the category **Cooking** or the category **iPad**

Facebook让广告商选择定向投放的目标用户

## 1.3 推荐系统在其他领域中的应用

### 其他推荐系统的应用场景：

- 基于位置的服务：微信附近的人、美团外卖附近的店
- 个性化邮件：Gmail的优先级邮箱



# 目录

1、什么是推荐系统

3、推荐系统在电商、社交、音乐、视频、阅读、服务等产品中的应用

5、好的推荐系统应该有哪些评测指标和评测维度

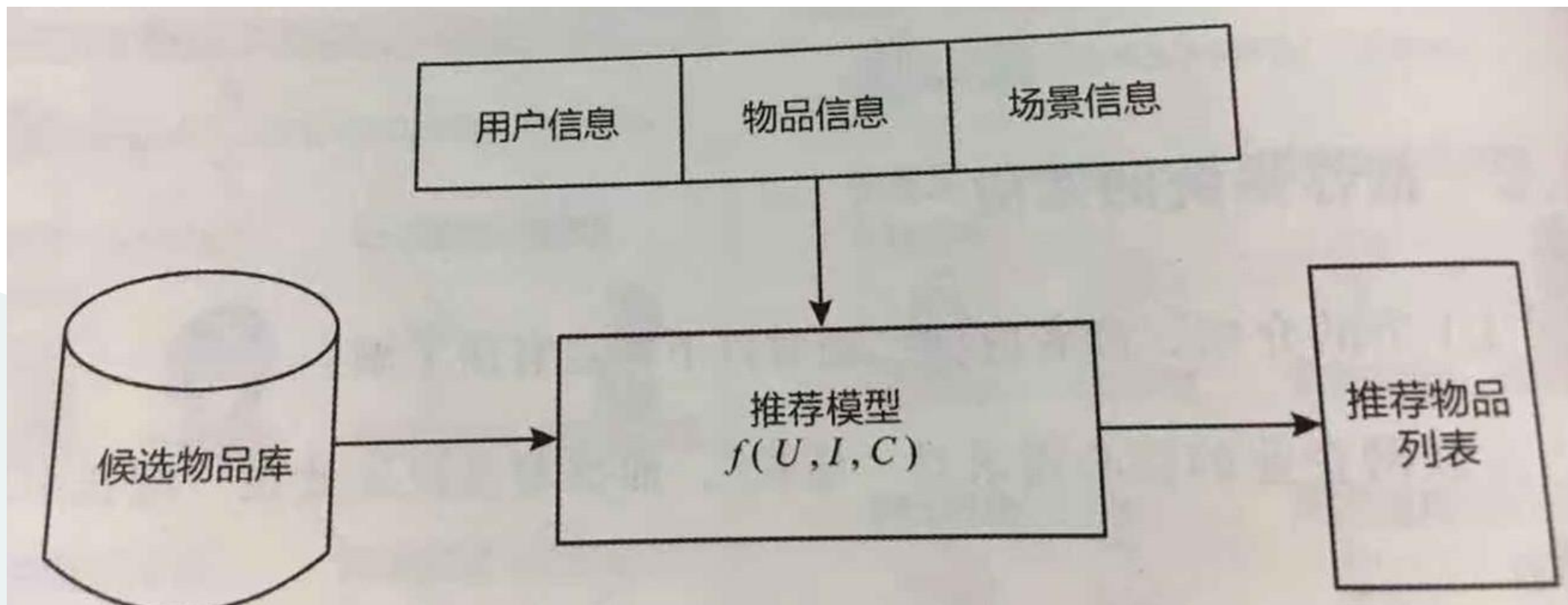
2、推荐系统的作用和意义

4、推荐系统的架构



## 1.4.1 推荐系统的逻辑架构

推荐系统可以形式化的定义为：对于用户 $U$  (user)，在特定场景 $C$  (context)下，针对海量物品 $I$  (Item)，构建一个函数 $f(U, I, C)$ ，预测用户对特定候选物品 $I$  (Item) 的喜好程度，再根据喜好程度对所有候选物品进行排序，生成推荐列表。

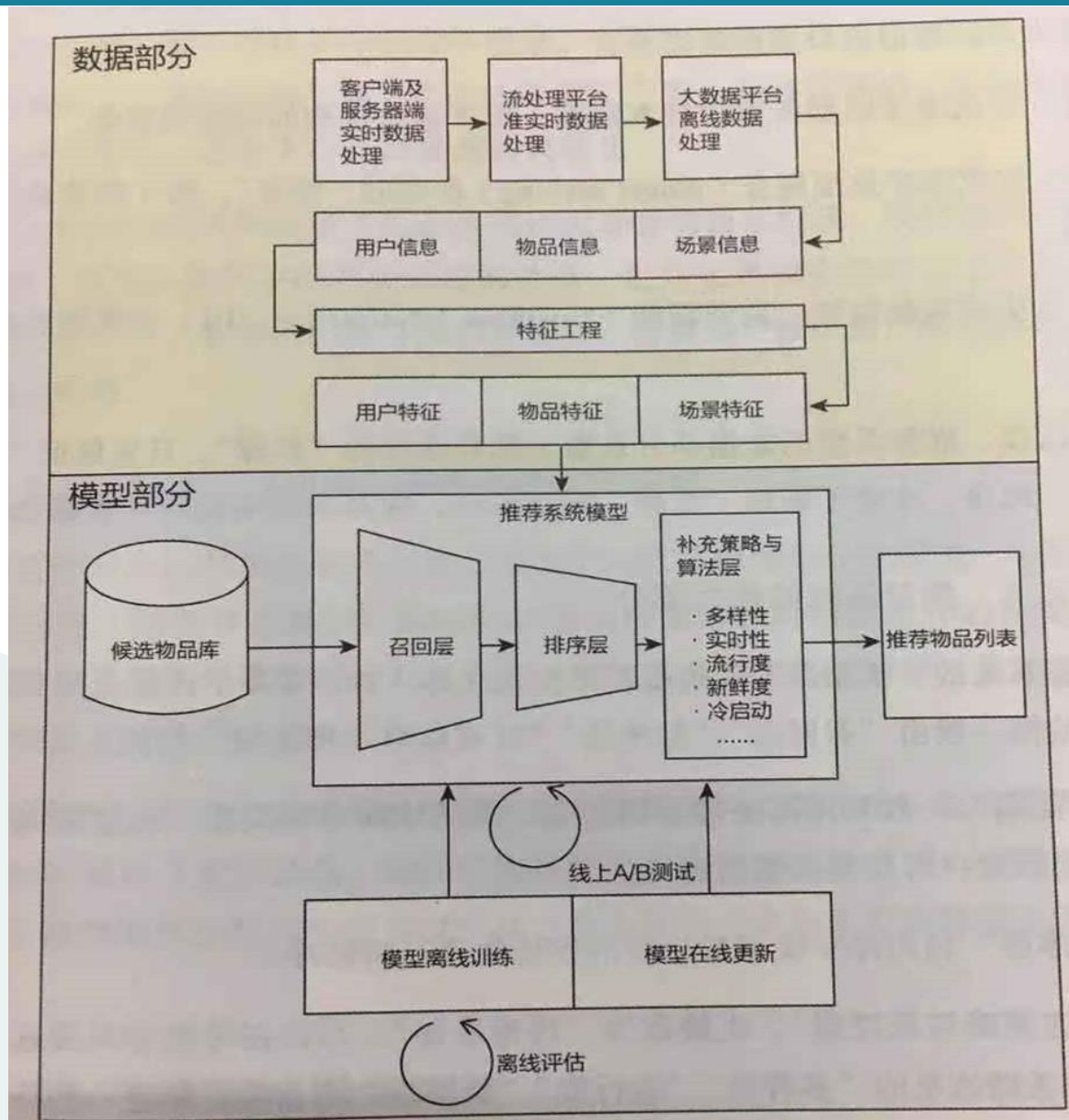


## 1.4.2 推荐系统的技术架构

工程师需要解决的问题是将抽象的概念和模块**具体化**、**工程化**。

- 数据和信息相关问题，即“用户信息”、“物品信息”、“场景信息”分别是什么？如何存储、更新和处理？
- 推荐系统算法和模型相关的问题，即推荐模型如何训练、如何预测、如何达到更好的效果？

## 1.4.2 推荐系统的技术架构

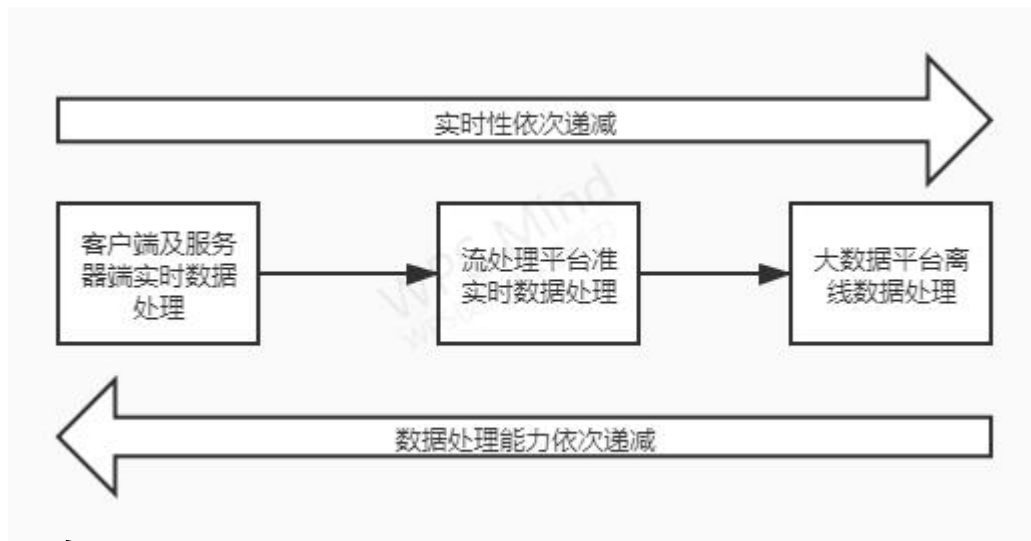


“数据和信息”部分逐渐发展为推荐系统中融合了数据离线批处理、实时流处理的数据流框架。

“算法和模型”部分则发展为推荐系统中集训练 (training)、评估 (evaluation)、部署 (deployment)、线上推断 (online inference) 为一体的模型框架。

### 1.4.3 推荐系统的数据部分

推荐系统的**数据部分**主要负责“用户”“物品”“场景”的**信息收集与处理**。



处理后的数据出口：

- 生成推荐模型所需的样本数据，用于算法模型的训练和评估
- 生成推荐模型服务所需的“特征”，用于线上推断。
- 生成系统监控，商业智能（BI）系统所需的统计型数据。

## 1.4.4 推荐系统的模型部分

- 推荐系统的模型部分一般由“召回层”“排序层”“补充策略和算法层”组成。
  - “召回层”利用高效的召回规则、算法或简单地模型，快速从海量候选集中召回用户可能感兴趣的物品。
  - “排序层”利用排序模型对初筛的候选集进行精排序。
  - “补充策略与算法层”也称为“再排序层”，在将推荐列表返回给用户之前，为兼顾结果的“多样性”“流行度”“新鲜度”等指标，结合一些补充策略和算法对推荐列表进行一定的调整，最终形成用户可见的推荐列表。
- 模型训练：
  - 离线训练：利用全量样本和特征，使模型逼近全局最优点
  - 在线更新：可以准实时的消化新数据样本，更快的反映新的数据变化趋势，满足模型实时性的需求。
- 模型评估：
  - 离线评估
  - 线上A/B Test评估



# 目录

1、什么是推荐系统

3、推荐系统在电商、社交、音乐、视频、阅读、服务等产品中的应用

5、好的推荐系统应该有哪些评测指标和评测维度

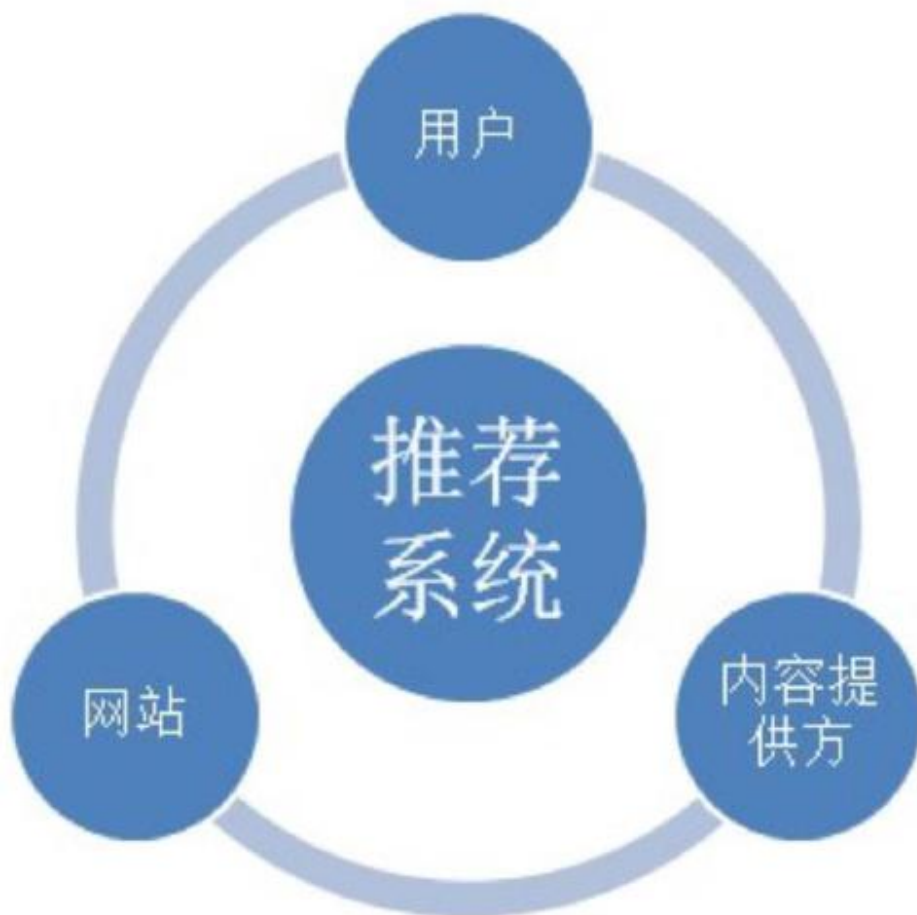
2、推荐系统的作用和意义

4、推荐系统的架构



## 1.5 推荐系统评测

- 如何评价推荐系统？这是推荐系统评测需要解决的首要问题。一个完整的推荐系统一般存在3个参与方：用户、物品提供者和提供推荐系统的网站。



## 1.5 推荐系统评测

- 以图书推荐为例，好的推荐系统需要满足：
  - 首先，推荐系统需要满足用户的需求，给用户推荐那些令他们感兴趣的图书。
  - 其次，推荐系统要让各出版社的书都能够被推荐给对其感兴趣的用户，而不是只推荐几个大型出版社的书。
  - 最后，好的推荐系统设计，能够让推荐系统本身收集到高质量的用户反馈，不断完善推荐的质量，增加用户和网站的交互，提高网站的收入。
- 好的推荐系统不仅仅能够准确预测用户的行为，而且能够扩展用户的视野，帮助用户发现那些他们可能会感兴趣，但却不容易发现的东西。

## 1.5 推荐系统评测

- 在推荐系统中，主要有3种评测推荐效果的实验方法：
  - 即离线实验 (offline experiment)
  - 用户调查 (user study)
  - 在线实验 (online experiment)

## 1.5 推荐系统评测

- **1.离线实验**的方法一般由如下几个步骤构成

- 通过日志系统获得用户行为数据，并按照一定格式生成一个标准的数据集；
- 将数据集按照一定的规则分成训练集和测试集；
- 在训练集上训练用户兴趣模型，在测试集上进行预测；
- 通过事先定义的离线指标评测算法在测试集上的预测结果。

- **优缺点**

优点	缺点
不需要有对实际系统的控制权	无法计算商业上关心的指标
不需要用户参与实验	离线实验的指标和商业指标存在差距
速度快，可以测试大量算法	

## 1.5 推荐系统评测

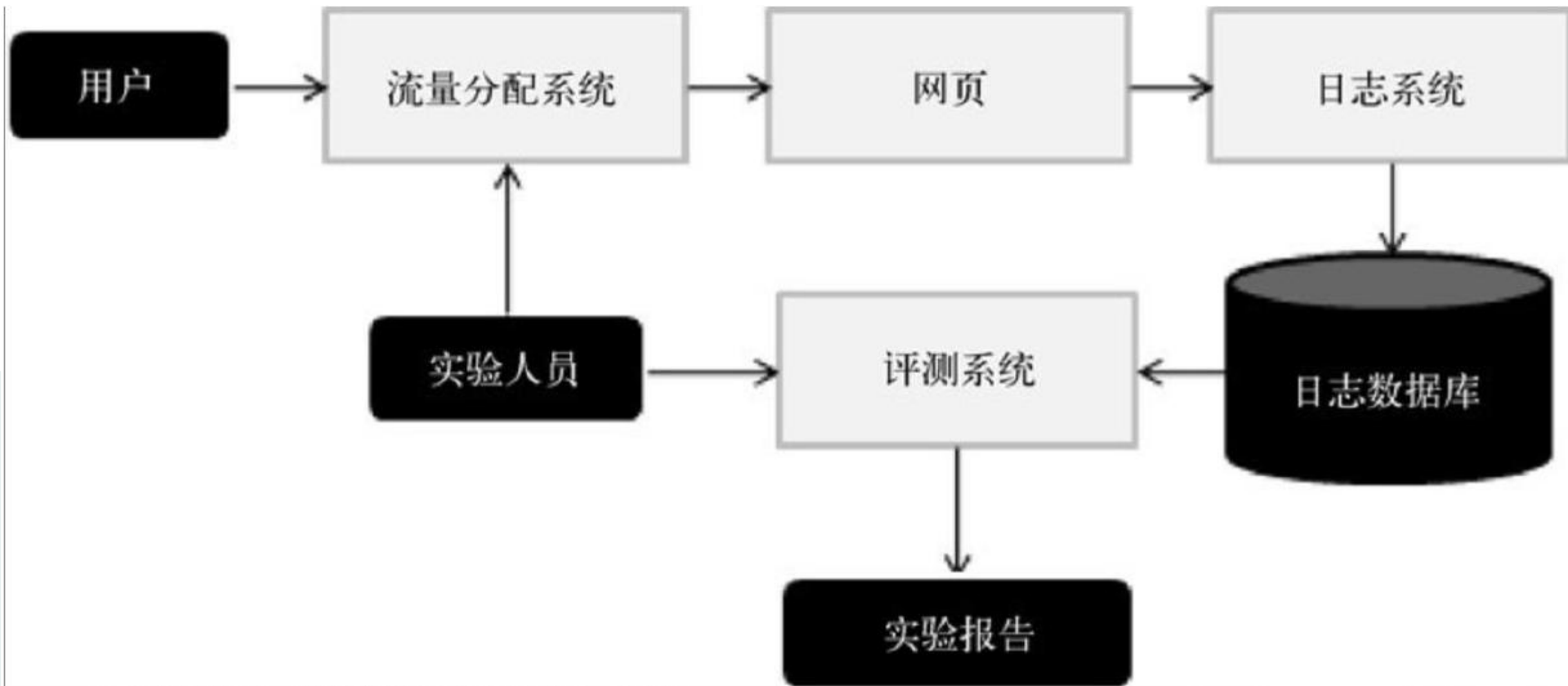
- **2.用户调查**需要有一些真实用户，让他们在需要测试的推荐系统上完成一些任务。在他们完成任务时，我们需要观察和记录他们的行为，并让他们回答一些问题。最后，我们需要通过分析他们的行为和答案了解测试系统的性能。
- **优缺点**
  - 优点是可以获得很多体现用户主观感受的指标，相对在线实验风险很低，出现错误后很容易弥补。
  - 缺点是招募测试用户代价较大，很难组织大规模的测试用户，因此会使测试结果的统计意义不足。此外，在很多时候设计双盲实验非常困难，而且用户在测试环境下的行为和真实环境下的行为可能有所不同，因而在测试环境下收集的测试指标可能在真实环境下无法重现。



## 1.5 推荐系统评测

- **3.在线实验顾名思义就是在完成离线实验和必要的用户调查后，可以将推荐系统上线做AB测试，将它和旧的算法进行比较。**
- AB测试是一种很常用的在线评测算法的实验方法。它通过一定的规则将用户随机分成几组，并对不同组的用户采用不同的算法，然后通过统计不同组用户的各种不同的评测指标比较不同算法，在完成离线实验和必要的用户调查后，可以将推荐系统上线做AB测试，将它和旧的算法进行比较。
- 优缺点
  - 优点是可以公平获得不同算法实际在线时的性能指标，包括商业上关注的指标。AB测试的缺点主要是周期比较长，必须进行长期的实验才能得到可靠的结果。
  - 缺点是，一个大型网站的AB测试系统的设计也是一项复杂的工程。一个大型网站的架构分前端和后端，从前端展示给用户的界面到最后端的算法，中间往往经过了很多层，这些层往往由不同的团队控制，而且都有可能做AB测试。如果为不同的层分别设计AB测试系统，那么不同的AB测试之间往往会互相干扰。

## 1.5 推荐系统评测



## 1.5 推荐系统评测

- 一般来说，一个新的推荐算法最终上线，需要完成上面三个步骤：
  - 首先，需要通过离线实验证明它在很多离线指标上优于现有的算法。
  - 然后，需要通过用户调查确定它的用户满意度不低于现有的算法。
  - 最后，通过在线的AB测试确定它在我们关心的指标上优于现有的算法。

## 1.5 推荐系统评测

- 这些评测指标可用于评价推荐系统各方面的性能。这些指标有些可以定量计算，有些只能定性描述，有些可以通过离线实验计算，有些需要通过用户调查获得，还有些只能在线评测。接下来列出不同的评测指标：
  - 用户满意度
  - 预测准确度
  - 覆盖率
  - 多样性
  - 新颖性
  - 惊喜度
  - 信任度
  - 实时性
  - 健壮性
  - ...

## 1.5 推荐系统评测

- 用户满意度:用户作为推荐系统的重要参与者，其满意度是评测推荐系统的最重要指标。但是，用户满意度没有办法离线计算，只能通过用户调查或者在线实验获得。



Starhunter 2300

Recommended because you  
have shown interest in Total  
Recall 2070

Are you interested?

yes

no

I've seen it



Brooklyn Rules

Recommended because you  
have shown interest in One  
Eyed King

Are you interested?

yes

no

I've seen it

## 1.5 推荐系统评测

- 用户满意度:用户作为推荐系统的重要参与者，其满意度是评测推荐系统的最重要指标。

但是，用户满意度没有办法离线计算，只能通过用户调查或者在线实验获得。





## 1.5 推荐系统评测

- 预测准确度度量一个推荐系统或者推荐算法预测用户行为的能力。这个指标是最重要的推荐系统离线评测指标，从推荐系统诞生的那一天起，几乎99%与推荐相关的论文都在讨论这个指标。这主要是因为该指标可以通过离线实验计算，方便了很多学术界的研究人员研究推荐算法。
- 计算步骤：
  - 获取离线的数据集，该数据集包含用户的历史行为记录。
  - 将该数据集通过时间分成训练集和测试集。
  - 通过在训练集上建立用户的行为和兴趣模型预测用户在测试集上的行为，并计算预测行为和测试集上实际行为的重合度作为预测准确度。

## 1.5 推荐系统评测

- 离线的推荐算法有不同的研究方向，常见的预测准确度指标有：

- 评分预测



## 1.5 推荐系统评测

- 评分预测的预测准确度一般通过均方根误差（RMSE）和平均绝对误差（MAE）计算。

对于测试集中的一个用户 $u$ 和物品 $i$ ，令 $r_{ui}$ 是用户 $u$ 对物品 $i$ 的实际评分，而 $\hat{r}_{ui}$ 是推荐算法给出的预测评分，那么RMSE的定义为：

$$\text{RMSE} = \frac{\sqrt{\sum_{u,i \in T} (r_{ui} - \hat{r}_{ui})^2}}{|T|}$$

MAE采用绝对值计算预测误差，它的定义为：

$$\text{MAE} = \frac{\sum_{u,i \in T} |r_{ui} - \hat{r}_{ui}|}{|T|}$$

## 1.5 推荐系统评测

- 离线的推荐算法有不同的研究方向，常见的预测准确度指标有：

- TopN 推荐

网站在提供推荐服务时，一般是给用户一个个性化的推荐列表，这种推荐叫做TopN推荐。TopN推荐的预测准确率一般通过准确率（precision）/召回率（recall）度量。

$$\text{Recall} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|}$$

召回率定义

$$\text{Precision} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|}$$

准确率定义

## 1.5 推荐系统评测

- 覆盖率（coverage）描述一个推荐系统对物品长尾的发掘能力。覆盖率有不同的定义方法，最简单的定义为推荐系统能够推荐出来的物品占总物品集合的比例。假设系统的用户集合为 $U$ ，推荐系统给每个用户推荐一个长度为 $N$ 的物品列表 $R(u)$ 。那么推荐系统的覆盖率可以通过下面的公式计算：

$$\text{Coverage} = \frac{|\bigcup_{u \in U} R(u)|}{|I|}$$

## 1.5 推荐系统评测

- 多样性：为了满足用户广泛的兴趣，推荐列表需要能够覆盖用户不同的兴趣领域，即推荐结果需要具有多样性。
- 多样性描述了推荐列表中物品两两之间的不相似性。因此，多样性和相似性是对应的。

假设  $s(i, j) \in [0, 1]$  定义了物品  $i$  和  $j$  之间的相似度，那么用户  $u$  的推荐列表  $R(u)$  的多样性定义如下：

$$\text{Diversity} = 1 - \frac{\sum_{i, j \in R(u), i \neq j} s(i, j)}{\frac{1}{2} |R(u)| (|R(u)| - 1)}$$

而推荐系统的整体多样性可以定义为所有用户推荐列表多样性的平均值：

$$\text{Diversity} = \frac{1}{|U|} \sum_{u \in U} \text{Diversity}(R(u))$$



## 1.5 推荐系统评测

- 新颖的推荐是指给用户推荐那些他们以前没有听说过的物品。  
在一个网站中实现新颖性的最简单办法是，把那些用户之前在网站中对其有过行为的物品从推荐列表中过滤掉。

## 1.5 推荐系统评测

- 惊喜度 (serendipity) 是最近这几年推荐系统领域最热门的话题。
- 惊喜度与新颖性有什么区别是首先需要弄清楚的问题。
- 如果推荐结果和用户的历史兴趣不相似, 但却让用户觉得满意, 那么就可以说推荐结果的惊喜度很高。

## 1.5 推荐系统评测

- 实时性：在很多网站中，因为物品（新闻、微博等）具有很强的时效性，所以需要在物品还具有时效性时就将它们推荐给用户。比如，给用户推荐昨天的新闻显然不如给用户推荐今天的新闻。因此，在这些网站中，推荐系统的实时性就显得至关重要。

## 1.5 推荐系统评测

### ● 信任度

- 如果用户信任推荐系统，那就会增加用户和推荐系统的交互。
- 度量推荐系统的信任度只能通过问卷调查的方式。

### 提高推荐系统的信任度主要有两种方法：

- 首先需要增加推荐系统的透明度（transparency），而增加推荐系统透明度的主要办法是提供推荐解释。
- 其次是考虑用户的社交网络信息，利用用户的好友信息给用户做推荐，并且用好友进行推荐解释。

## 1.5 推荐系统评测

- 健壮性：任何一个能带来利益的算法系统都会被人攻击，这方面最典型的例子就是搜索引擎。搜索引擎的作弊和反作弊斗争异常激烈，这是因为如果能让自己的商品成为热门搜索词的第一个搜索果，会带来极大的商业利益。推荐系统目前也遇到了同样的作弊问题，而健壮性（即robust,鲁棒性）指标衡量了一个推荐系统抗击作弊的能力。

## 1.5 推荐系统评测

- 商业目标：很多时候，网站评测推荐系统更加注重网站的商业目标是否达成，而商业目标和网站的盈利模式是息息相关的。一般来说，最本质的商业目标就是平均一个用户给公司带来的盈利。不过这种指标不是很难计算，只是计算一次需要比较大的代价。因此，很多公司会根据自己的盈利模式设计不同的商业目标。



## 1.5 推荐系统评测

- 本节提到了很多指标，其中有些指标可以离线计算，有些只能在线获得。但是，离线指标很多，在线指标也很多，那么如何优化离线指标来提高在线指标是推荐系统研究的重要问题。关于这个问题，目前仍然没有什么定论，只是不同系统的研究人员有不同的感性认识。

	离线实验	问卷调查	在线实验
用户满意度	×		○
预测准确度	√		×
覆盖率	√		√
多样性	○		○
新颖性	○		○
惊喜度	×		×

## 1.5 推荐系统评测

- 推荐系统还需要通过不同维度来进行评测：
  - 用户维度 主要包括用户的人口统计学信息、活跃度以及是不是新用户等。
  - 物品维度 包括物品的属性信息、流行度、平均分以及是不是新加入的物品等。
  - 时间维度 包括季节，是工作日还是周末，是白天还是晚上等。

1、什么是推荐系统

3、推荐系统在电商、社交、音乐、视频、阅读、服务等产品中的应用

5、好的推荐系统应该有哪些评测指标和评测维度

2、推荐系统的作用和意义

4、推荐系统的架构





Thank you!