

法律声明

■ 本课件包括演示文稿、示例、代码、题库、视频和声音等内容，北风网和讲师拥有完全知识产权；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或者机构不得盗版、复制、仿造其中的创意和内容，我们保留一切通过法律手段追究违反者的权利。

■ 课程详情请咨询

◆ 微信公众号：北风教育

◆ 官方网址：<http://www.ibeifeng.com/>



人工智能之推荐系统

推荐系统概述

主讲人：Gerry

上海育创网络科技有限公司



课程要求

■ 课上课下 “九字” 真言

- ◆ 认真听，善摘录，勤思考
- ◆ **多温故，乐实践**，再发散

■ 四不原则

- ◆ **不懒散惰性，不迟到早退**
- ◆ **不请假旷课，不拖延作业**

■ 一点注意事项

- ◆ 违反 “四不原则”，不包就业和推荐就业

严格是大爱



寄语



做别人不愿做的事，
做别人不敢做的事，
做别人做不到的事。

课程内容

- 推荐系统概述
- Surprise框架的搭建

推荐系统直观了解



机器学习 平装
周志华
★★★★☆ 142
¥71.00 ✓prime



机器学习实战 平装
哈林顿 (Peter Harrington)
★★★★☆ 158
¥54.30 ✓prime



深度学习(deep learning) 平装
伊恩·古德费洛 (Ian Goodfellow)
★★★★☆ 74
¥126.20 ✓prime



统计学习方法 平装
李航
★★★★☆ 163
¥35.30 ✓prime

经常一起购买的商品



总价: ¥187.60

全部加入购物车

这些商品由不同的卖家进行配送和销售。查看发货和库存信息

- ☑ 本商品: 机器学习 - 弗拉赫 (Peter Flach) 平装 ¥62.30
- ☑ 机器学习 - 周志华 平装 ¥71.00
- ☑ 机器学习实战 - 哈林顿 (Peter Harrington) 平装 ¥54.30

您最近查看的商品和相关推荐

猜您喜欢

第1页, 共5页



图解机器学习
[日]杉山将
★★★★☆ 28
平装
¥40.30 ✓prime



Apple iPhone 8 Plus
(A1864) 64GB 银色 移动联
通电信4G手机
★★★★☆ 16
¥5,987.00



Apple iPhone 8 Plus
(A1864) 64GB 深空灰色 移
动联通电信4G手机
★★★★☆ 16
¥5,994.00



统计学习方法
李航
★★★★☆ 163
平装
¥35.30 ✓prime



Apple 苹果 手机 iPhone X
银色 64G
★★★★☆ 2
¥7,979.00



深度学习(deep learning)
伊恩·古德费洛 (...)
★★★★☆ 74
平装
¥126.20 ✓prime



美国MoKo 苹果iPhone 8
钢化膜 iPhone7弧边钢化玻
璃保护膜 苹果8高清高透 ...
★★★★☆ 40
¥28.00 ✓prime



德国Leicke莱克 无线鼠标
轻薄鼠标 静音鼠标 超薄舒
适 笔记本鼠标 游戏电竞 可
爱无 ...
★★★★☆ 581
¥69.90 ✓prime



Apple iPhone 7 Plus 32G
黑色 移动联通电信4G手机
★★★★☆ 297
¥5,087.00



机器学习
周志华
★★★★☆ 142
平装
¥71.00 ✓prime



推荐系统直观了解

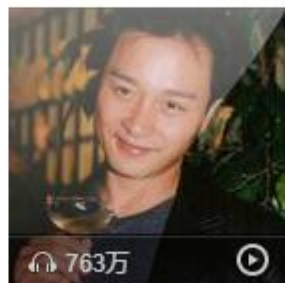
热门推荐 华语 | 流行 | 摇滚 | 民谣 | 电子 更多 →



353万
梦幻流行 | 纵晨遥所
往，长眠雪森



65万
【最爱金曲榜】一听就
有力量的歌



763万
百首老歌：再续当年情



14173人一起听过
电台节目 新世相 | 异地
恋：未来看不到头，现
在看不到你



165万
八十八星座 | 夜空漫游
指南



36093
电台节目 【发刊词】夏
鹏：为什么你应该通过
音乐来学英语



830万
人这一辈子，最怕突然
听懂了一首歌。



858万
电台节目 你把我灌醉 - 冯
提莫

相似歌曲

- 枕边人
李袁杰/筱筱雨沐 ▶ +
- 爱呀
李明霖 ▶ +
- 流浪生
一棵小葱 ▶ +
- 粉红色的回忆
王宝强/刘昊然/肖央/潘粤明... ▶ +
- 比一个年轻人小一点的鹤
丁可 ▶ +

推荐系统概述

- 推荐系统是利用网站向客户提供商品信息和建议，帮助用户决定应该购买什么产品，模拟销售人员帮助客户完成购买的过程。而个性化推荐是根据用户的兴趣特点和购买行为，向用户推荐用户感兴趣的信息和商品。
- 推荐系统产生的主要原因是由于随着电子商务的发展，商品数量和种类在快速的增长，客户需要花费大量的时间才能找到自己想购买的商品，而这种时间的消耗无疑会在信息过载问题中使得客户不断流失。
- 为了解决这些问题，个性化推荐系统就应运而生。个性化推荐系统是建立在海量数据挖掘基础上的一种高级商务智能平台，以帮助电子商务网站为客户提供完全个性化的决策支持和信息服务。

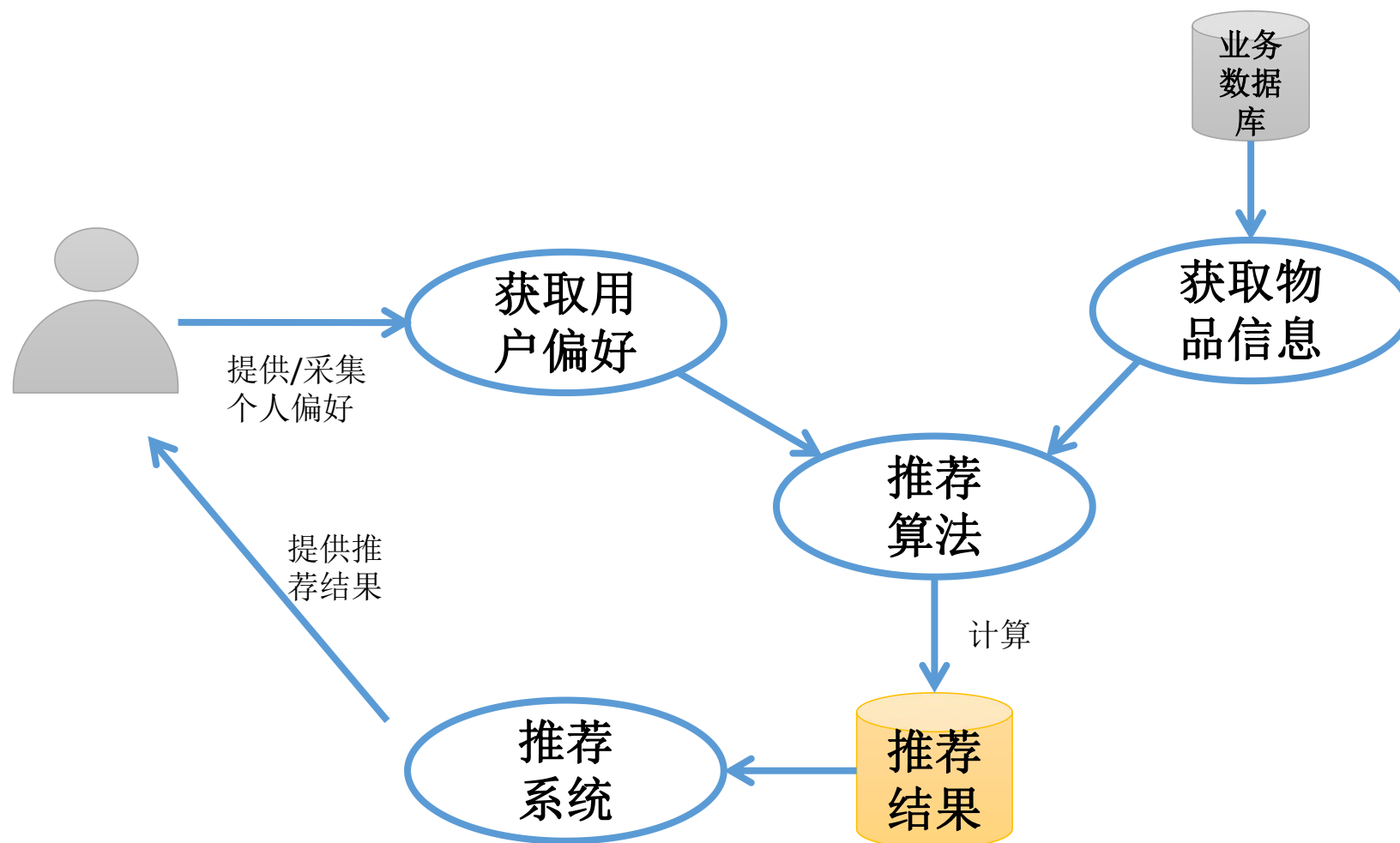
推荐系统概述

- **推荐系统(Recommender System, RS)**是向用户建议有用物品的软件工具和一种技术。常用于多种决策过程，比如购买什么商品、听什么音乐、在网站上浏览什么新闻等等。
- “物品(item)” 是用来表示系统向用户(user)推荐内容的总称，一个推荐系统通常专注于一个特定类型的物品(eg: CD、图书、电器等)，设计的核心是为特定类型的物品提供有用和有效的建议而定制的。
- RS分为**个性化推荐**和**非个性化推荐**两大类；个性化推荐是指基于用户对商品的偏好信息来进行其它商品的推荐，至于非个性化推荐一般是指将热销商品进行推荐(eg: Top10的商品推荐)；这两种不同的推荐方式在某些不同场景下都会存在着比较不错的效益

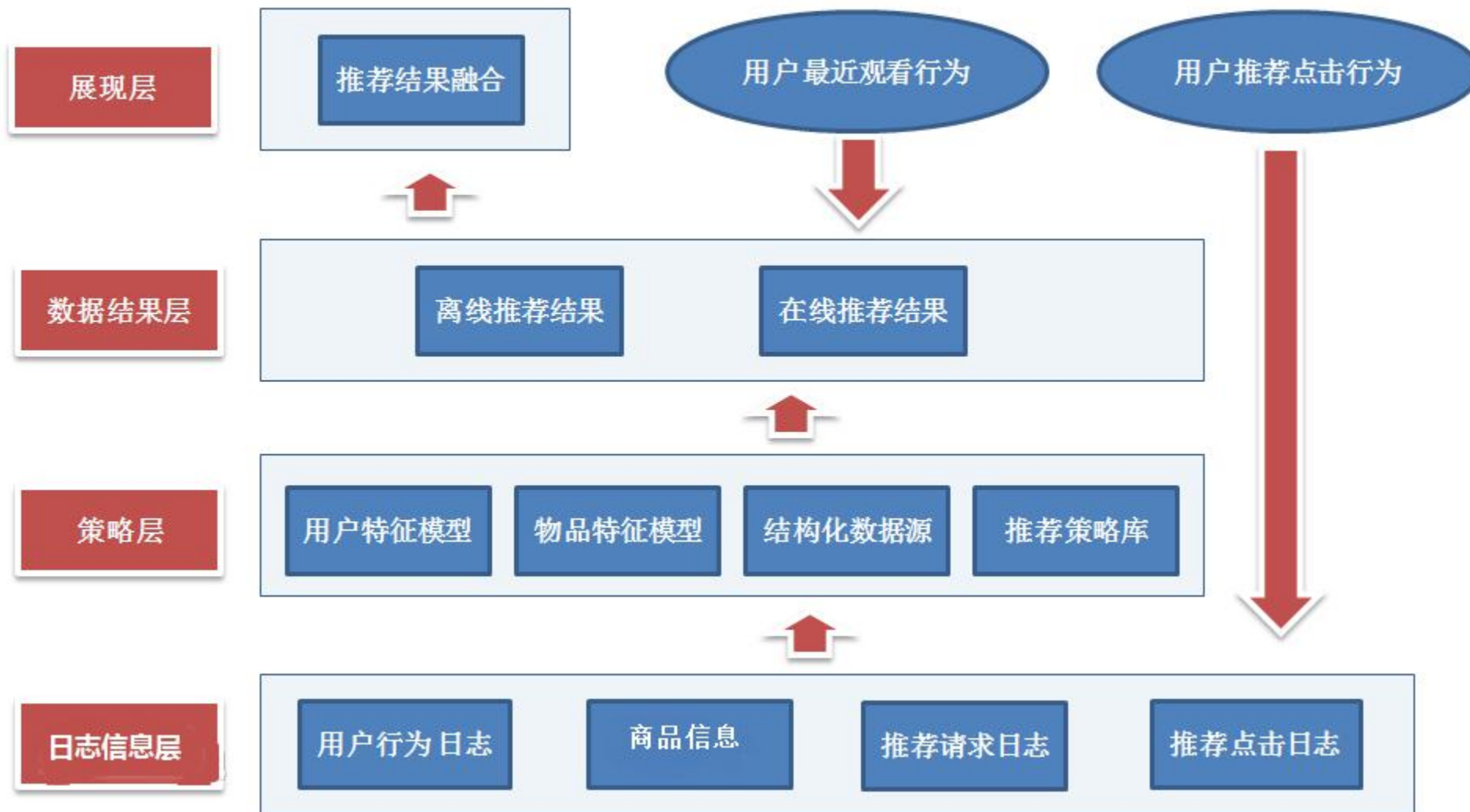
推荐系统的功能

- 增加物品销售的数量
- 出售更多种类的物品
- 增加用户满意度
- 更好的了解用户需求，增加用户忠诚度

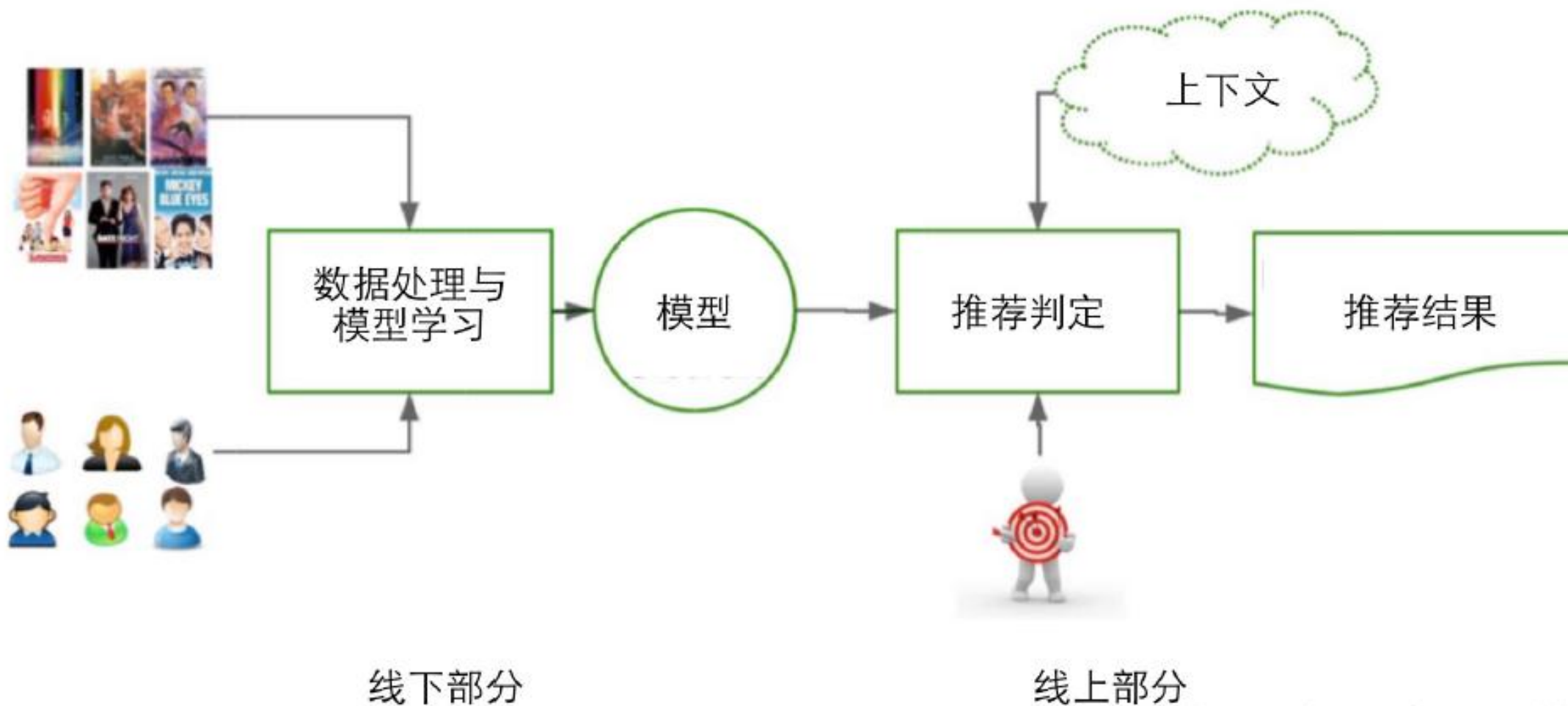
推荐系统通用模型



推荐系统结构



推荐系统结构



推荐系统推荐结果形式

■ 评分预测方式：

- ◆ 推荐的最终结果是根据用户对物品的打分或者物品与物品之间的关联性产生的，最终的推荐商品列表打分比较高或者关联性比较高的商品。

■ TopN预测方式：

- ◆ 最终的推荐商品就是一组商品序列，不会涉及到用户或者商品与推荐结果之间的关联关系的。

推荐系统效果评估

- 可以将推荐系统的运行过程看成一个回归问题或者分类问题，所以可以通过类似的分类/回归相关指标进行衡量，主要包括：
 - ◆ 均方误差(MSE)
 - ◆ 均方根误差(RMSE)
 - ◆ 平均绝对误差(MAE)
 - ◆ 一致序列对比率评分(FCP)
 - ◆ 准确率(Precision)
 - ◆ 召回率(Recall)
 - ◆ ROC曲线 (TPR/FPR)
 - ◆ 覆盖率、多样性、新颖性、惊喜度、实时性、商业目标等

推荐系统效果评估

- MSE(均方误差): Mean Squared Error; 是指**估计值与真实值**之差平方的期望值。

$$MSE = \frac{1}{|R|} \sum_{r_{u,i} \in R} (\hat{r}_{u,i} - r_{u,i})^2$$

- RMSE(均方根误差): Root Mean Squared Error; 是均方误差的算法平方根。

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{|R|} \sum_{r_{u,i} \in R} (\hat{r}_{u,i} - r_{u,i})^2}$$

- MAE(平均绝对误差): Mean Absolute Error; 是指**估计值与真实值**之间绝对差值的期望值。

$$MAE = \frac{1}{|R|} \sum_{r_{u,i} \in R} |\hat{r}_{u,i} - r_{u,i}|$$

推荐系统效果评估

- FCP(一致序列对比率评分): Fraction of Concordant Pairs。计算评分一致的物品对在所有物品对中的占比。

$$n_c^u = \left| \left\{ (i, j) \mid \left(\hat{r}_{u,i} > \hat{r}_{u,j} \ \& \ r_{u,i} > r_{u,j} \right) \text{or} \left(\hat{r}_{u,i} == \hat{r}_{u,j} \ \& \ r_{u,i} == r_{u,j} \right) \right\} \right|$$

$$n_d^u = \left| \left\{ (i, j) \mid \left(\hat{r}_{u,i} - \hat{r}_{u,j} \right) \left(r_{u,i} - r_{u,j} \right) \leq 0 \ \& \ i > j \right\} \right|$$

$$n_c = \sum_u n_c^u \qquad n_d = \sum_d n_d^u$$

$$FCP = \frac{n_c}{n_c + n_d}$$

推荐系统效果评估

	被推荐	未被推荐
实际使用/实际访问	真阳性数(tp)	假阴性数(fn)
实际未使用/实际未访问	假阳性数(fp)	真阴性数(tn)

$$recall(TruePositiveRate, TPR) = \frac{\#tp}{\#tp + \#fn}$$

$$precision = \frac{\#tp}{\#tp + \#fp}$$

$$FalsePositiveRate(FPR) = \frac{\#fp}{\#fp + \#tn}$$

MovieLens

- MovieLens是一个专注于个性化电影推荐的项目，定义我们在推荐系统中对数据的格式要求，在常规的推荐系统框架中，基本上所有的模型均要求数据格式必须是服从MovieLens定义的数据格式的。格式为：**用户id、商品id、评分、时间戳**(可选)。
 - ◆ <https://movielens.org/>
 - ◆ <https://grouplens.org/datasets/movielens/>

userId	movieId	rating	timestamp
1	31	2.5	1260759144
2	10	4.0	835355493
3	84236	4.0	1298922130

推荐系统框架

名称	SVDFeature	LibMF	LibFM	Lenskit	GraphLab	Mahout	SparkMLib	EasyRec	Waffles	Myrrix
语言	C++	C++	C++	Java	C++	Java	Java/Scala	Java	C++	C++
分布式	否	否	否	否	是	是	是	否	否	是
学术 界产 品	是	是	是	否	否	否	否	否	否	否

- 除了常用的这十种推荐框架外，还有很多其它的框架，比如: [Surprise](#)(python)、[Recommendable](#)(Ruby)、[Oryx](#)(Java)、[Recommenderlab](#)(R)等等，这些框架除了使用的编程语言不同以外，实现的算法大同小异，主要是：SVD、ALS、CF及其改进算法等等。

Surprise

- Surprise (Simple Python Recommendation System Engine) 是scikit系列的一个库。支持多种推荐算法，是python语言使用实现推荐算法的首选方式。
 - ◆ 官网: <http://surpriselib.com/>
 - ◆ 文档: <http://surprise.readthedocs.io/en/stable/>
 - ◆ 源码: <https://github.com/NicolasHug/Surprise/>
- 安装方式(依赖numpy服务):
 - ◆ pip install scikit-surprise 或者 conda install -c conda-forge scikit-surprise

```
C:\Users\ibf>pip install scikit-surprise
Collecting scikit-surprise
  Downloading scikit-surprise-1.0.5.tar.gz (2.4MB)
    100% |#####| 2.4MB 7.9kB/s
Requirement already satisfied: numpy>=1.11.2 in c:\anaconda3\lib\site-packages (from scikit-surprise)
Requirement already satisfied: six>=1.10.0 in c:\anaconda3\lib\site-packages (from scikit-surprise)
Collecting joblib>=0.11 (from scikit-surprise)
  Using cached joblib-0.11-py2.py3-none-any.whl
Building wheels for collected packages: scikit-surprise
  Running setup.py bdist_wheel for scikit-surprise ... done
  Stored in directory: C:\Users\ibf\AppData\Local\pip\Cache\wheels\99\99\d8\5a5468fbc316324616dee7cb2ab0c4fe9671441b96130b67e87
Successfully built scikit-surprise
Installing collected packages: joblib, scikit-surprise
Successfully installed joblib-0.11 scikit-surprise-1.0.5
```

```
C:\Users\ibf>python
Python 3.6.0 |Anaconda 4.3.1 (64-bit)| (default, Dec 23 2016, 11:57:41) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import surprise
>>>
```

Surprise

- Surprise支持基本的常用推荐算法：
 - ◆ 基础算法/baseline algorithms
 - ◆ 协同过滤算法(基于近邻算法)/neighborhood methods
 - ◆ 矩阵分解算法/matrix factorization-based(SVD, SVD++, NMF)

Surprise支持的算法

算法	描述
random_pred.NormalPredictor	基于统计的随机预测打分，假定用户的打分分布是服从高斯分布的。
baseline_only.BaselineOnly	基于统计的基准线预测打分。
knns.KNNBasic	基本的协同过滤算法。
knns.KNNWithMeans	基本协同过滤算法的变种，考虑每个用户的平均评分。
knns.KNNWithZScore	基本协同过滤算法的变种，考虑每个用户评分的归一化操作。
knns.KNNBaseline	基本协同过滤算法的变种，考虑基于统计的基准线评分。
matrix_factorization.SVD	SVD矩阵分解算法。
matrix_factorization.SVDpp	SVD++矩阵分解算法。
matrix_factorization.NMF	一种基于非负矩阵分解的协同过滤算法。
slope_one.SlopeOne	SlopeOne协同过滤算法。
co_clustering.CoClustering	一种基于协同聚类的协同过滤算法。

http://surprise.readthedocs.io/en/stable/prediction_algorithms_package.html

Surprise支持的相似度度量指标

相似度度量标准	描述
cosine	计算所有用户或者所有物品之间的余弦相似度。
msd	计算所有用户或者所有物品之间的平均平方差相似度。
pearson	计算所有用户或者所有物品之间的皮尔森相似度。
pearson_baseline	使用基准线方式计算皮尔森相似度。

<http://surprise.readthedocs.io/en/stable/similarities.html>

Surprise支持的评估准则

评估准则	描述
rmse	Root Mean Squared Error，计算均方根误差。
mae	Mean Absolute Error，计算平均绝对误差。
fcp	Fraction of Concordant Pairs.

<http://surprise.readthedocs.io/en/stable/accuracy.html>



THANK YOU

上海育创网络科技有限公司