电商推荐与销量预测相关案例

寒小阳
七月在线　2017年1月21日

Event Recommendation Engine Challenge

Predict what events our users will be interested in based on user actions, event metadata, and demographic information.

背景介绍

We (the competition hosts) are excited to sponsor the Event Recommendation Engine Challenge, which asks you to predict what events our users will be interested in based on events they've responded to in the past, user demographic information, and what events they've seen and clicked on in our app. The insights you discover from this data, and the algorithms the winners create, will allow us to improve our event recommendation algorithm, a core part of our applications and a key element in improving user experience.

This is the first competition launching under the Kaggle Startup Program!

数据信息

Data Files

File Name	Available Formats
event_attendees.csv	.gz (55.03 mb)
events.csv	.gz (161.46 mb)
user_friends.csv	.gz (148.60 mb)
random_benchmark	.csv (146.91 kb)
users	.csv (2.63 mb)
event_popularity_benchmark	.csv (146.91 kb)
public_leaderboard_solution	.csv (6.46 kb)
test	.csv (574.40 kb)
train	.csv (924.23 kb)
event_popularity_benchmark_private_test_only	.csv (51.51 kb)

百度网盘地址如下：

链接: https://pan.baidu.com/s/1bpaHIVl 密码: 3git

问题简单分析

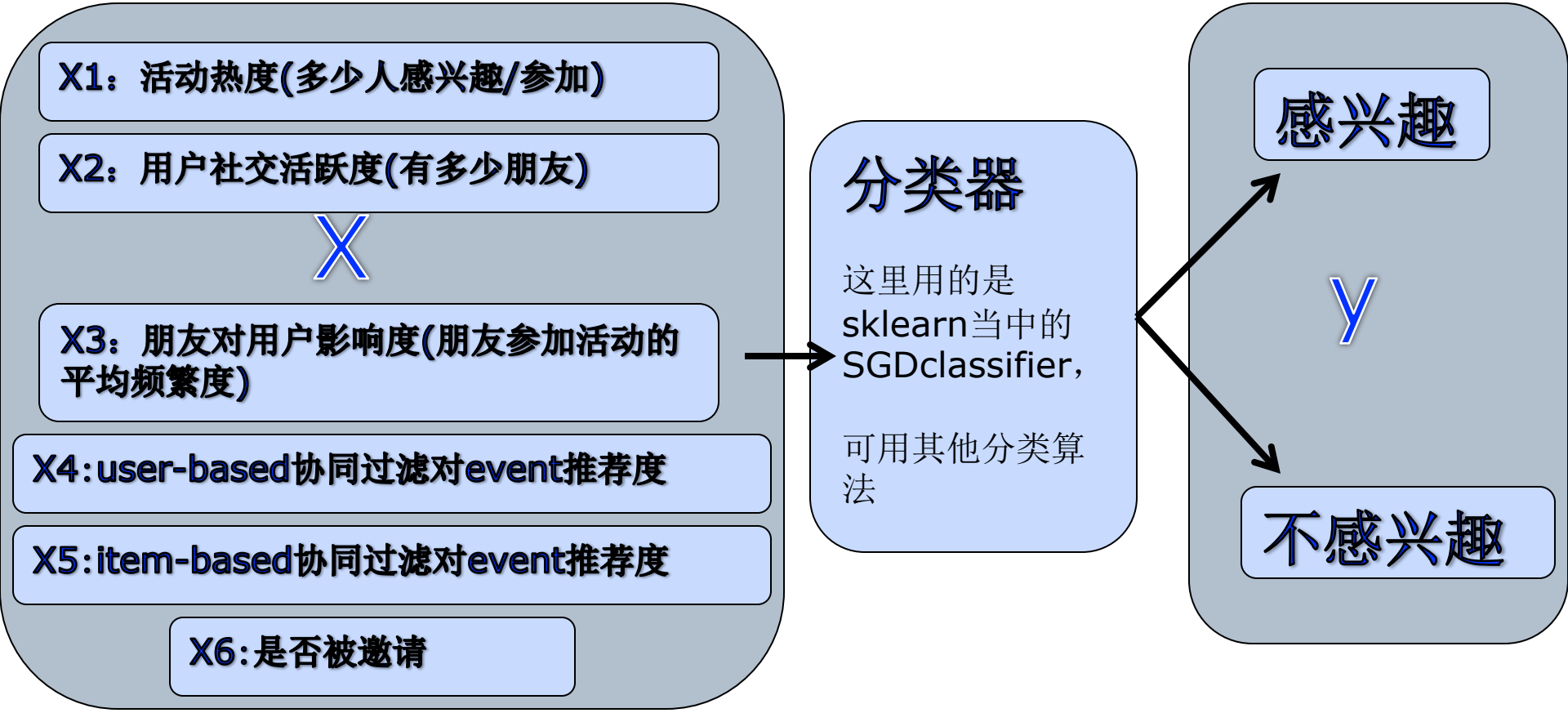
- 首先这是一个推荐的问题
- 我们有下面这样几类数据

1. 用户的历史数据 => 对event是否感兴趣/是否参加
2. 用户社交数据 => 朋友圈
3. event相关的数据 => event相关的信息

- 简单思考

1. 要把更多维度的信息纳入考量
2. 协同过滤是基于 用户-event 历史交互数据
3. 需要把社交数据 和 event相关信息 作为影响最后结果的因素纳入考量
4. 视作分类模型，每一个人 感兴趣/不感兴趣 是target，其他影响结果的是feature
5. 影响结果的feature包括由协同过滤产出的推荐度

工业界标准做法：思路简图



更多细节

请参见
ipython notebook
与
对应注释

第1名的做法

请参见
课程代码与讲解

便利店销量预测问题

数据处理
特征工程
Xgboost loss function修改
详见**ipython notebook**

感谢大家！

恳请大家批评指正！