

# **Product Recommendations Based on Browsing Trajectories**

**Liu Chuang**

**2018-11-13**

# Product Recommendations Based on Browsing Trajectories

- Motivation
- Recommendation Alg
- Innovation
- Processing
- Baselines
- Problems

# 1. Motivation

- 通过浏览序列，发现产品之间的潜在关联性，向用户个性化的定制推荐产品——有相似特质但是不相同的产品，避免持续重复和爆品推荐。
- 通过用户的浏览行为，挖掘用户的爱好和品味
- 对长尾资源的盘活和利用，因为用户对长尾内容通常是陌生的，无法主动搜索，唯有通过推荐的方式，引起用户的注意，发掘出用户的兴趣，帮助用户做出最终的选择。

## 2. Recommendation Alg

- **Content-based**

*Base on the features of the item*

- **Collaborative Filtering-based**

*Base on historical usage data to item*

## 2. Recommendation Alg

- 矩阵分解

### Netflix 比赛——SVD

分解矩阵，将用户和物品映射到  $k$  维空间，得到物品背后暗藏的一些被关注的因素，和 用户在若干因素上的偏好

### 改进方案

- 考虑偏置信息
- 考虑隐式反馈和用户属性
- 考虑时间因素

### 不足

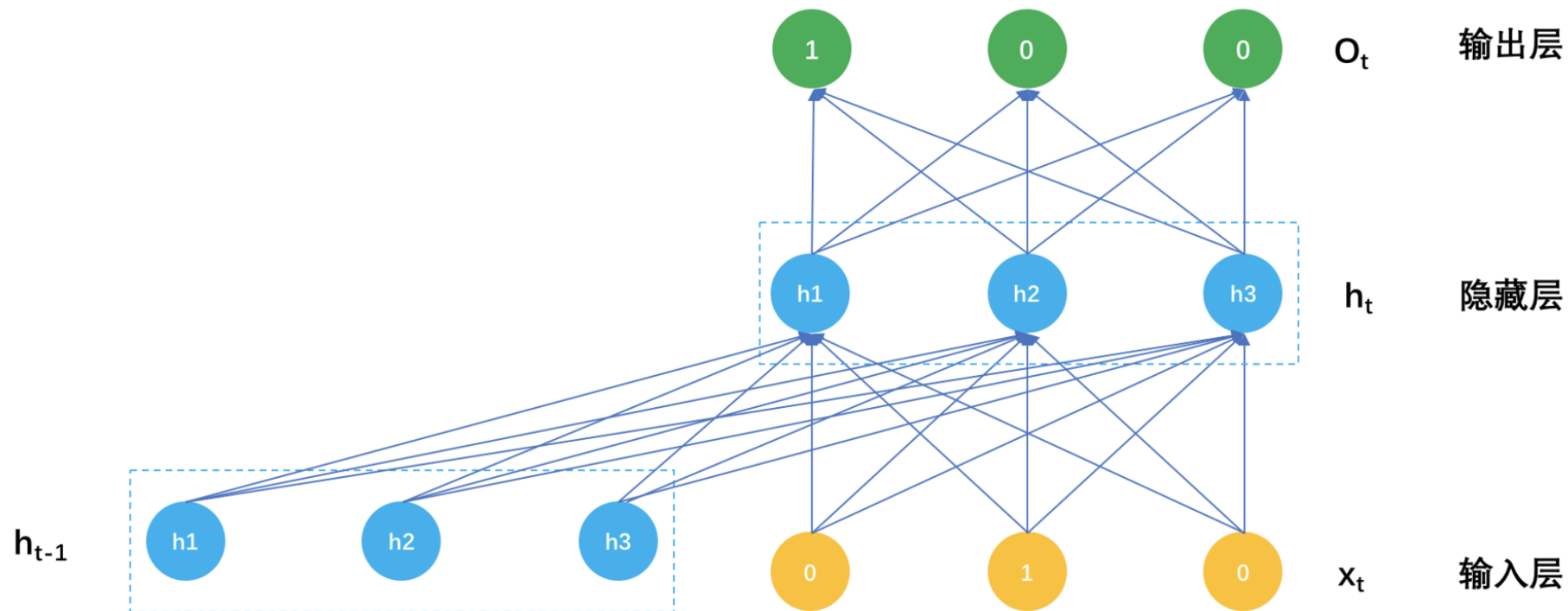
- 主要依靠显示反馈： 评分，收藏，购买
- 忽略操作序列

## 2. Recommendation Alg

- RNN

### Spotify—音乐推荐

时间属性：不但受特征影响，还受上一首歌影响



## 2. Recommendation Alg

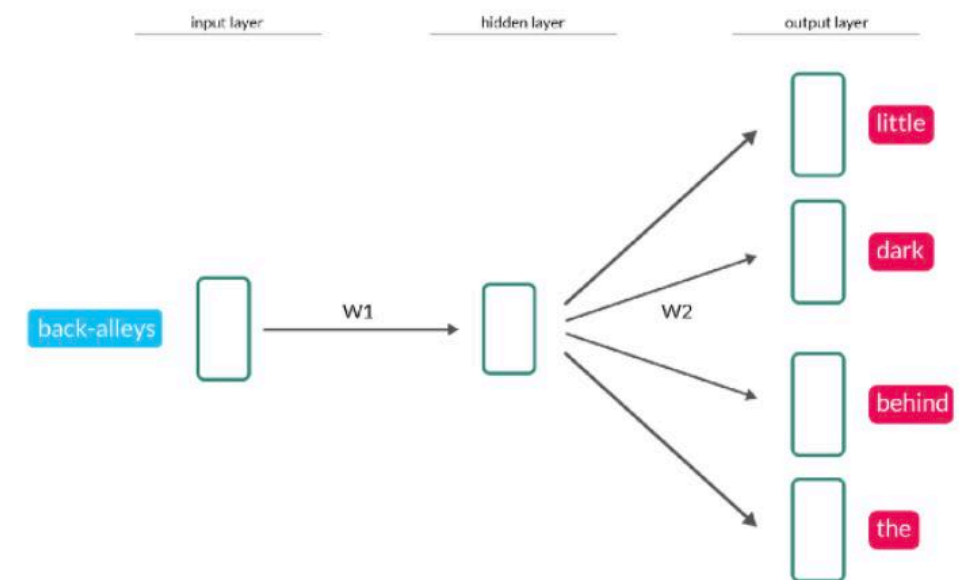
- Word2vec

- Music recommendations at Anghami
- Listing recommendations at Airbnb
- Product recommendations in Yahoo Mail

- Word2vec : 对于每一个词，生成一组向量，作为其在上下文中含义的编码

- 类似于CNN 使用滑动窗口
- 用每一个词 去预测窗口正中央的词 或者相反
- 每次确定  $W1$ ， $W2$ ；移动窗口

- 
- 如果有不同的两个词，出现在相同的上下文之中，我们期望 输出相似=二者矩阵相似
  - 使用中间的权重为词编码



## 2. Recommendation Alg

- Word2vec

Researchers from the Web Search, E-commerce and Marketplace domains have realized that just like one can train word embeddings by treating a sequence of words in a sentence as context, the same can be done for training embeddings of user actions by treating sequence of user actions as context.” - Mihajlo Grbovic, Airbnb

音乐：

- 歌曲看作单词，挖掘歌曲的上下文信息

购物：

- 购物的顺序反映产品相关性

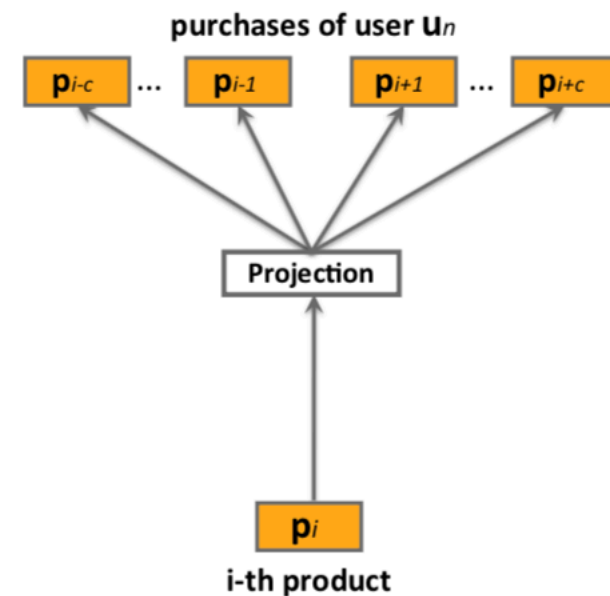
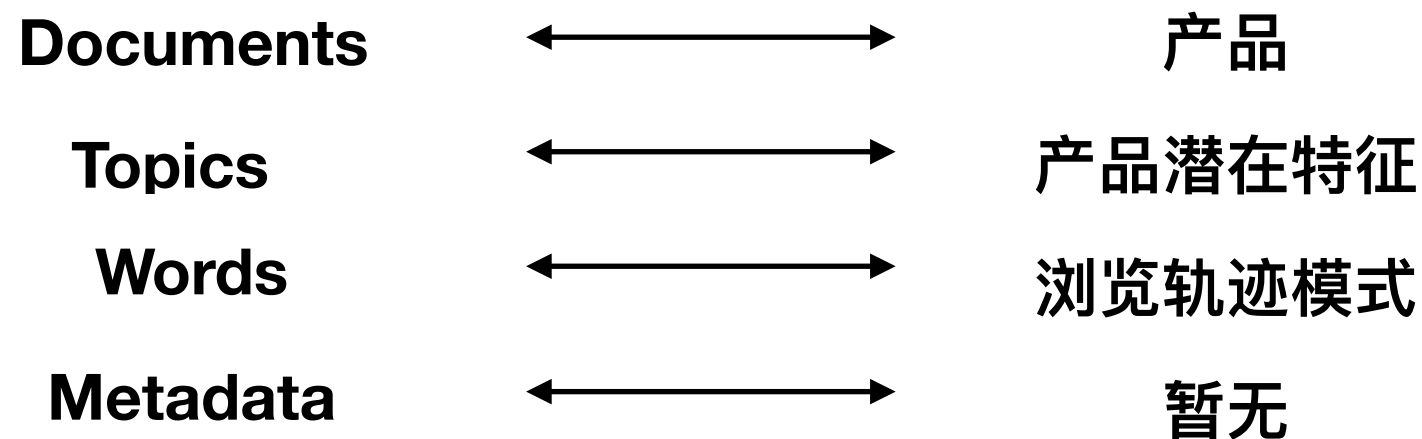


Figure 2: prod2vec skip-gram model



# 3.Innovation

- 考虑到用户的浏览轨迹，会反映产品之间的相关性。使用用户的轨迹，挖掘产品的潜在属性，对产品进行分类，同时分析用户品味，进一步为其进行个性化的推荐。
- 类比于NLP中文档主题挖掘，使用 LDA 对用户轨迹进行分析



## Pros:

- 使用的是用户的隐式反馈信息—相比于显示反馈，更加稠密，更能真实反映用户想法
- 物品的个数是远远小于用户的数量的，而且物品的个数和相似度相对比较稳
- 发现潜在的联系（产品）和兴趣（用户），避免推荐全部是重复和爆品— — 一定程度解决EE问题  
长尾理论作为一种新的经济模式，被成功的应用于网络经济领域。而 对长尾资源的盘活和利用，恰恰是推荐系统所擅长的，因为用户对长尾内 容通常是陌生的，无法主动搜索，唯有通过推荐的方式，引起用户的注意， 发掘出用户的兴趣，帮助用户做出最终的选择。
- 考虑时间序列+ 停留时间 🌟

## 4. Processing

物品编码，输入每个人的访问序列  $a1=(r1,r2,...rn)$

# LDA k-topics

每个物品对应一个k维隐因子向量( $q_1, q_2 \dots q_k$ )

## K-means

## 物品聚类 m类

再逐一计算用户和每个聚类中心的推荐分数，给用户推荐物品聚类

将个人访问序列 映射为-->访问主题序列 (c1,c2...cm)

## 加和平均一用户兴趣特点

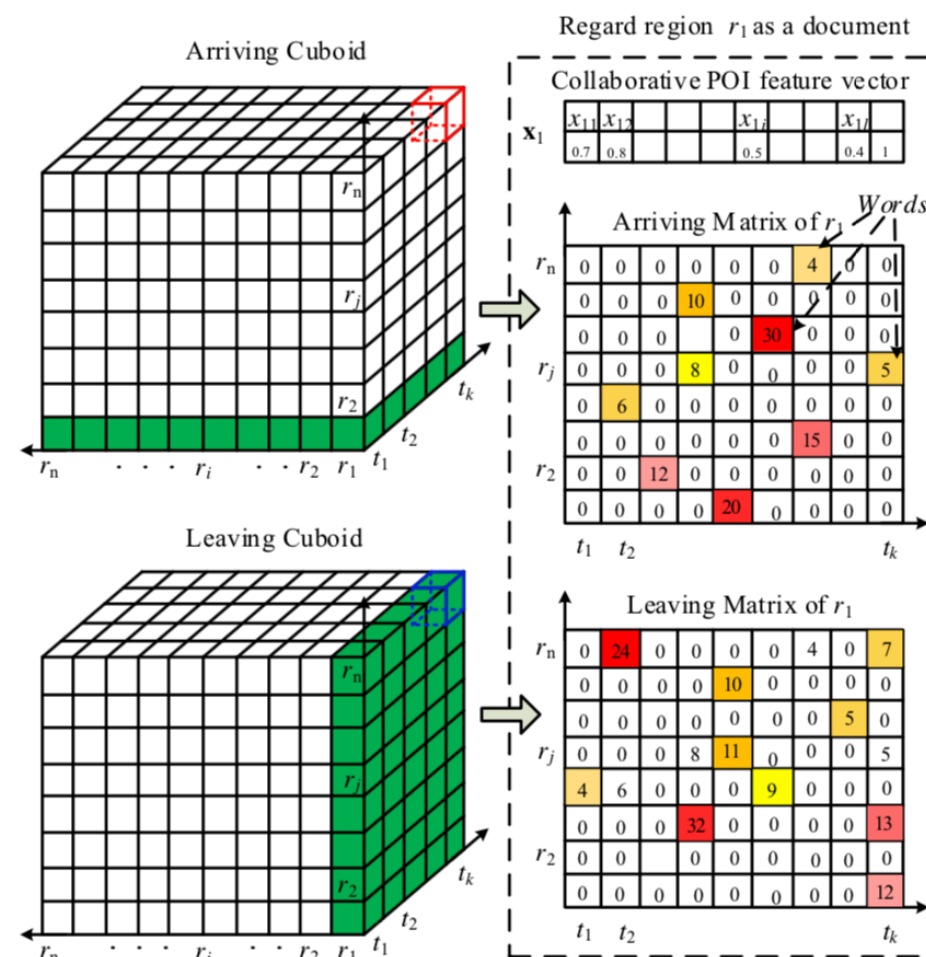
## 频繁序列挖掘

## 深度学习

## 按照聚类中 rank 推荐物品

## 推荐的物品聚类

## T-SNF 压缩 二维可视



# 5. Baselines

## 开源工具

- 矩阵分解

用户访问商品次数的矩阵—提取特征

**Lightfm**

Python 单机多线程

**Kgraph**

KNN 相似度计算

- Word2vec

不考虑时间属性

**Gensim**

**Tensorflow-word2vec(Google)**

- LDA

不考虑时间属性

**FastText(Facebook)**

Python 单机多线程

- RNN

**Tensorflow**

## 6. Problems

### 1. 推荐的可解释性

2. 人的mobility 出发和 离开 最终都有目的地，然而购物， 序列不连贯会终止跳出，下次打开搜索新的物品

### 3.冷启动问题 💡

### 4.EE问题 💡

每个人对重复的定义不一样。有人觉得这篇讲皇马和巴萨的文章，昨天已经看过类似内容，今天还说这两个队那就是重复。但对于一个重度球迷而言，尤其是巴萨的球迷，恨不得所有报道都看一遍

### 5. 时间序列—意义

不考虑具体的什么时间访问，只关心停留时间==POI

### 6. Rank

价格，用户兴趣特点，EE，