

# 数据库系统概论新技术篇

## 社交网络大数据分析 ——社交媒体初探 (3)

赵鑫

中国人民大学信息学院

2017年5月

# 典型任务与解决方法

数据质量清洗

用户舆情分析

用户画像构建

用户兴趣学习



# 典型任务与解决方法（续）

## ❖ 用户画像构建

- 直接抽取

- 知识推断

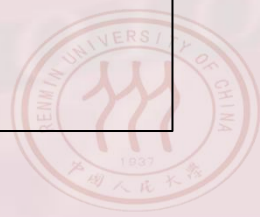
- 隐含表示

直接抽取:

社交媒体上拥有丰富的用户信息，可以直接抽取，但是可能存在缺失与虚假信息。

常用技术:

- 1、数据归一化技术
- 2、属性补全技术
- 3、用户链指技术



# 典型任务与解决方法（续）

## ❖ 用户画像构建

- 直接抽取
- 知识推断
- 隐含表示

知识推断:

结合用户文本信息与知识规则，进行推断。

“这款手机送给我老婆” => 性别 = 男

“IT公司经常加班.....” => 工作 = IT

常用技术:

- 1、属性信息抽取
- 2、知识图谱推断



# 典型任务与解决方法（续）

## ❖ 用户画像构建

- 直接抽取
- 知识推断
- 隐含表示

隐含表示:

除了显式特征外，还可以使用隐含表示的用户画像。

常用技术:

- 1、利用社交网络嵌入表示技术
- 2、针对特定任务，使用深度学习模型来学习用户隐含表示



# 典型任务与解决方法

数据质量清洗

用户舆情分析

用户画像构建

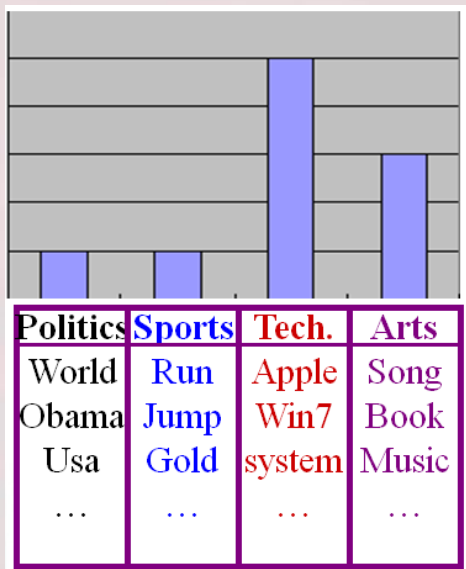
用户兴趣学习



# 典型任务与解决方法（续）

## ❖ 用户兴趣学习：话题兴趣

### ■ 短文本主题模型



常用的改良技术手段

聚合用户微博作为一个文档

短文本局部主题相关性

引入背景模型

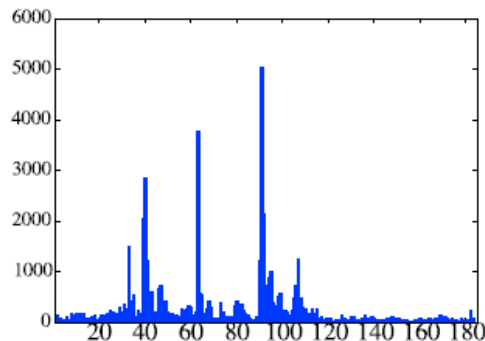


# 典型任务与解决方法（续）

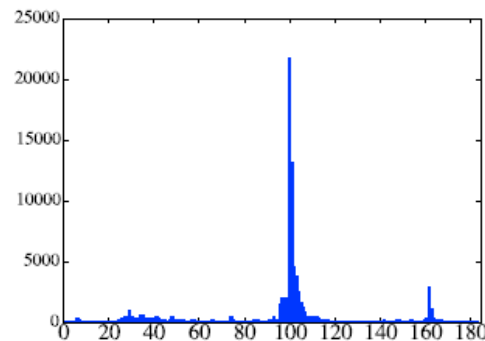
## ❖ 用户兴趣学习：话题兴趣

### ■ 突发热点检测模型 (burst detection)

- 热点话题会催生社交媒体上特定内容的 **burst**



(a) "Earthquake" in Twitter.



(b) "Nobel Prize" in Twitter.

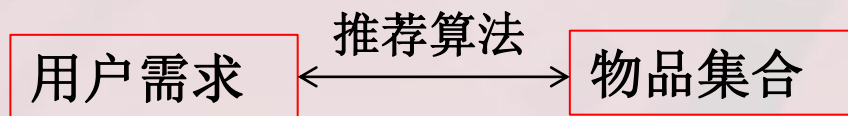




# 典型任务与解决方法（续）

## ❖ 用户兴趣学习：通用推荐问题的刻画

- 如何满足一位社交用户的多种喜好，如微博、产品、音乐、文章等



# 典型任务与解决方法（续）

## ❖ 用户兴趣学习：通用推荐技术

### ■ 传统算法

- 流行度
- $K$ 近邻

### ■ 分析

- 容易实现、效果稳定
- 能够利用的信息种类少，不容易拓展



# 典型任务与解决方法（续）

## ❖ 用户兴趣学习：通用推荐技术

### ■ 基于矩阵分解的技术

$$\frac{1}{2} \|R - U^T V\|_F^2,$$

### ■ 变种

- 引入文本内容特征
- 引入关系特征
- 引入时间特征
- . . .



# 典型任务与解决方法（续）

## ❖ 用户兴趣学习：通用推荐技术

### ■ 基于深度模型的推荐技术

- 使用深度网络刻画用户与物品间交互行为
  - 例如，基于**DNN**实现的矩阵分解方法
- 对于附加信息实现深度变换与映射
  - 例如，通过自动编码器模型对于文本内容进行变换
- 捕捉更为复杂的行为模式
  - 例如，使用**RNN**来刻画序列行为特征



# 讲述提纲

社交媒体概述

常用数据处理技术

常见任务与解决方法

未来研究展望

小结



# 未来研究展望

多网站信息源的有效融合

数据与算法的“充分打通”

应用对研究的“充分驱动”



# 总结

了解常用社交媒体数据处理技术

初步了解常见社交媒体任务的解决方法

