

反作弊系统与技术

反作弊相关内容简介

徐恩

2016 年 4 月 18 日



背景与现状

微博现有的数据分布不是很乐观:

微博数据分布的段子

斯诺登称美国已窃取百万计中国人微博内容。其数据分析结果如下: 35% 是人生哲理. 25% 是商业广告. 30% 是冷笑话、贷款、复制手机卡监听小三等垃圾信息, 奥巴马绝望地问: 那剩下的 10% 一定是重要内容了? 美情报局长: 报告总统, 剩下的 10% 是黄段子。



作弊的数据分布

在微博数据质量里面有一部分数据是刻意为之的作弊数据，其占总体微博量的**万分之一的量级**，但是这些作弊微博**集中在一些搜索热门领域**的内容中, eg. 普通热门搜索词、垂直领域的热门搜索词。



什么是作弊

作弊指的是恶意利用 优化手段 以期 获得某种利益 的行为。

作弊的表现形式

- 1 内容作弊: 主题无关与图文无关
- 2 行为作弊: 恶意导流与恶意转赞



作弊手段与技术

- 1 初级篇: 堆砌热搜词到垃圾内容包含热搜词, 再到随机文本包含热搜词, 最终为相关热搜词内容无关图片



作弊手段与技术

- ❶ 初级篇: 堆砌热搜词到垃圾内容包含热搜词, 再到随机文本包含热搜词, 最终为相关热搜词内容无关图片
- ❷ 进阶篇: 少量用户快速大量发作弊内容到大量稀疏作弊



作弊手段与技术

- ❶ 初级篇: 堆砌热搜词到垃圾内容包含热搜词, 再到随机文本包含热搜词, 最终为相关热搜词内容无关图片
- ❷ 进阶篇: 少量用户快速大量发作弊内容到大量稀疏作弊
- ❸ 中级篇: 控制大量群体稀疏发作弊内容, 到控制部分群体作弊、部分社会化操作



作弊手段与技术

- ❶ 初级篇: 堆砌热搜词到垃圾内容包含热搜词, 再到随机文本包含热搜词, 最终为相关热搜词内容无关图片
- ❷ 进阶篇: 少量用户快速大量发作弊内容到大量稀疏作弊
- ❸ 中级篇: 控制大量群体稀疏发作弊内容, 到控制部分群体作弊、部分社会化操作
- ❹ 终极篇: 引入大量优质异常用户发相关作弊内容, 分割作弊群体部分转, 伪装部分正常, 化作弊群体模拟社会化



反作弊的主要业务

根据作弊内容的分布反作弊在垂直业务层次上分成了热搜词作弊、营销关键词作弊、色情作弊三个大类；



反作弊业务与搜索

反作弊是微博搜索的防火墙也是一个微博数据入侵检测系统，与微博搜索的其他业务之间的关系如下：



反作弊的基础原理

基本理论方法 1,2,3



反作弊系统介绍

基本理论方法 1,2,3



反作弊系统页面部分

页面部分是反作弊系统的传感器部分, 肩负着数据采集以及特征分析的功能,



反作弊系统用户部分

用户部分是反作弊系统的核心分析部分，主要负责单个用户的动态和静态行为分析，



反作弊系统群体部分

群体部分是反作弊系统的用户社会行为扩充部分，负责分析用户群体的行为分析



反作弊系统资源属性部分

资源属性部分是反作弊系统的关联检测部分，负责分析作弊用户所使用的资源性质



历史问题回顾



未来发展规划

