

# 通读笔记——*Popularity Prediction for Social Media over Arbitrary Time Horizons*

2112233062 何智鹏

2022 年 12 月 19 日

## 1 论文信息

会议名: VLDB 会议

会议级别: CCF-A 类

会议 track: Volume15,issue9

发表年份: 2021

标题: Popularity Prediction for Social Media over Arbitrary Time Horizons

作者及工作单位:

- Daniel Haimovich (Core Data Science, Meta Platforms)
- Dima Karamshuk (Core Data Science, Meta Platforms)
- Thomas J. Leeper (Core Data Science, Meta Platforms)
- Evgeniy Riabenko (Core Data Science, Meta Platforms)
- Milan Vojnovic (Department of Statistics, LSE)

## 2 关键词

社交媒体 (Social Media)      流行率预测 (Popularity Prediction)      自  
激的霍克斯点过程模型 (self-excited Hawkes point process model)

## 3 概述

作者提出了一个社交媒体项目的人气预测模型, 该模型满足大规模在线平台中出现的一系列设计考虑。

## 4 研究的问题及其意义

研究的问题是在给定内容的初始流行度增长和其他内容特征的信息下，在任意时间范围内预测社会网络内容的流行率。这些预测对于各种应用都是有用的，包括在线社交网络中潜在的有害病毒内容的早期检测。

## 5 前序工作及其优缺点

目前的流行度预测算法是准确的，但大多不能扩展到处理大规模社交媒体内容。

## 6 论文提出的思路和方法，及其优缺点

**思路：**作者提出了一种基于自激霍克斯点过程模型的基于特征的方法，该模型依赖于使用参考时间范围内信息瀑布大小的点预测器和有效增长指数的点预测器，捕捉信息级联传播的时间尺度。

**优点：**高度可扩展、高度的准确性、在任意预测时间范围内预测流行程度。

**缺点：**计算复杂度与预测精度之间还需完善。

## 7 数据集和实验工具

数据库和数据集皆可公开且共享。

**数据集：**有两个。第一个是包含 10 万个公开页面帖子且浏览量在数千亿量级的数据集。第二个是 20 万个随机抽样的公共页面帖子，并且还跟踪了它们在创建后长达 2 个月的转发和浏览量，收集了数十亿个此类事件的时间戳。

**实验工具：**24 Intel Core Processor (Broadwell) CPUs and 114GB of RAM

## 8 实验结果

结果表明，与其他自激点过程模型相比，作者提出的方法可以提供更准确的预测结果，与针对特定预测范围训练的模型相比，具有竞争力。

## 9 我的思考

作者提出了一个社交媒体项目流行率预测模型，该模型满足大规模在线平台中出现的一组设计考虑。这些考虑因素包括为任何给定的预测时间和范

围提供准确的预测，在预测时间具有恒定的时间计算复杂性，并利用静态和时间特征来确保准确。