

黄昕

huan233@usc.edu · +1(213)421-2739 · hxin18.github.io

教育背景

南加州大学, 洛杉矶, 美国

2018 年 1 月 – 至今

计算机科学, 硕士研究生

上海交通大学, 上海, 中国

2012 年 9 月 – 2016 年 8 月

信息工程, 学士

专业课程: 数据结构 (85/100), 操作系统 (96/100), 算法 (95/100), 计算机体系结构 (98/100)

技能

- 编程语言: C/C++, Python, HTML/CSS/JavaScript, R, SQL, Java, Matlab
- 平台: Hadoop, Spark
- 语言: 英语 (托福成绩:100), 法语 (TCF:B2)

实习经历

海颐软件

2015 年 7 月 – 2015 年 8 月

软件工程师

- 参与南方电网营销系统的构建 (基于 J2EE)。
- 编写电费管理系统前端网页 (html/css/javascript)。

项目经历

基于学术话题的论文推荐

2017 年 2 月 – 2017 年 11 月

项目组长 (1/4)

导师: 王新兵

- 提出了一个可扩展的生成模型 (能够处理百万量级的学术数据集)。它结合主题模型, 有效地描述论文在不同学术主题中的重要性。
- 该模型在于基于主题模型的引用预测中, 所提出的模型在召回率, 准确率上提高一倍以上。
- 以第一作者的身份在 PAKDD 2018 会议上发表论文 (CCF C 类)

学术大数据中的话题演变挖掘

2015 年 9 月 – 2016 年 8 月

项目组长 (1/3)

导师: 王新兵

- 基于主题模型, 提出了联合生成学术论文的文本, 发表时间和引用网络的生成模型。
- 通过模型构建学术话题演变网络, 并通过 gephi 软件进行数据可视化。

交互式主题模型设计

2015 年 8 月 – 2015 年 9 月

项目组员 (1/3)

导师: 王新兵

- 对 LDA 主题模型进行优化, 设计树形结构记录用户对主题模型结果的反馈, 并将该树形结构作为主题模型的先验超参。
- 使用 Java 实现模型, 并通过 Sparse LDA 的框架提升模型训练速度。
- 在 IEEE 数据集上进行实验, 发现提出模型有效提升主题与文本关联度。

社交网络图压缩算法研究

2016 年 9 月 – 2017 年 2 月

项目组员 (1/3)

导师: Mauro Sozio

- 在 BV-Graph 的框架下, 基于局部敏感哈希算法, 设计算法为图的节点排序以提高压缩率。
- 在压缩社交网络的图文件中, 所提出的算法较现有方法提升约 3%。

发表论文

X.Huang, C.Chen, C.Peng, X.Wu, L.Fu, X.Wang *Topic-sensitive Influential Paper Discovery in Citation Network*. Accepted by PAKDD 2018 (Acceptance Rate: 20.6%)