

# 李言迪

(+86) 18401596880 | liyandi007@163.com | yandili | Yandi | yandi.space

北京, 中国

## 工作经历

- |  |           |                   |
|--|-----------|-------------------|
| 新浪微博   | 北京, 中国    | 11/2014 - 至今      |
| <ul style="list-style-type: none"><li>机器学习算法工程师<ul style="list-style-type: none"><li>负责用户画像及兴趣预测模型</li><li>短期兴趣实时化改造, 从日级别提升到10分钟内</li></ul></li></ul>       |           |                   |
| Yahoo! 研究院   | 巴塞罗那, 西班牙 | 02/2014 - 07/2014 |
| <ul style="list-style-type: none"><li>实习研究员<ul style="list-style-type: none"><li>第一作者发表文章: Applying space syntax to online mapping tools</li></ul></li></ul> |           |                   |

## 教育经历

- |   |                    |                   |
|---|--------------------|-------------------|
| 加泰罗尼亚理工(UPC) & 里昂二大   | 巴塞罗那, 西班牙 & 里昂, 法国 | 09/2012 - 09/2014 |
| <ul style="list-style-type: none"><li>硕士, 数据挖掘与知识管理专业 (DMKM)<ul style="list-style-type: none"><li>Erasmus Mundus奖学金</li><li>课程包括: Kernel-based Learning, Statistical NLP, Advanced Statistical Modeling, Bioinformatics</li></ul></li></ul> |                    |                   |
| 浙江大学  | 杭州, 中国             | 09/2008 - 06/2012 |
| <ul style="list-style-type: none"><li>本科, 数学与应用数学<ul style="list-style-type: none"><li>GPA: 3.87/4.0, 学院排名5%</li></ul></li></ul>  |                    |                   |
| 加州大学洛杉矶分校 (UCLA)  | 洛杉矶, 美国            | 08/2009 - 09/2009 |
| <ul style="list-style-type: none"><li>暑期学校, 金融、英文写作</li></ul>   |                    |                   |

## 论文研究

- [P1] **Apply Space Syntax to Online Mapping Tools**, *WSDM 2017*
- Yandi Li**, Nicola Barbieri(Tumblr), Daniele Quercia(Bell Labs)
  - 研究生毕设, 基于时间、天气信息, 推荐有趣导航路线的系统
  - 技术点: Factorization Machine, BPR, PostgreSQL, PostGIS, QGIS, A\*, Space Syntax
- [P2] **Chebyshev-Tao方法的不同实现及其比较**
- 本科毕设, [论文部分], [文献综述], [答辩]

## 工作项目

- |   |             |
|---|-------------|
| 微博用户建模  | 05/2022- 至今 |
| <ul style="list-style-type: none"><li>使用深度模型为兴趣建模<ul style="list-style-type: none"><li>降噪VAE训练方式, Caser结构, 目标做人群归一化, 意外度</li><li>将多种用户特征/行为日志嵌入统一空间 (包括标签序列, 关注列表, app列表等)</li></ul></li><li>丰富全站用户画像</li><li>技术点: Fuxi-CTR, GBDT, PyTorch, Ollama, Huggingface</li></ul> |             |

- 微博用户兴趣** 技术负责人 05/2019 - 05/2022
- 长短期兴趣计算
    - 短期兴趣实时化改造：更新时间从天级别逐步提升到10分钟内, 接入3倍数量的行为日志
    - 难点：兴趣计算中灵敏度，马太效应，粒度细化等问题（[专利CN115827966A](#)）
  - 广告商业兴趣重构：统计模型 --> 监督式算法（GBDT）的改造，提高定向投放转化率
  - 算法迭代效率：建设兴趣计算的离线和在线评估，效果展示，调试等模块
  - 技术点：Flink, Hivemall, lightgbm, grafana, streamlit, clickhouse

- 微博图片后推荐** 技术负责人 08/2017 - 04/2019
- 推荐场景：图片右滑展示相关图片，完成全流程0-1打通
  - 使用对比学习训练，生成图片语义向量
  - 技术点：asyncio, sanic, PyTorch, Spark, faiss

- 图片特征挖掘** 08/2016 - 10/2017
- 图片分类和人脸识别（明星和政治人物）
  - 作弊识别: [仿冒明星认证的头像](#)
  - 智能图片裁剪：为图片选取最适合的位置裁剪缩略图
  - 技术点：Tensorflow, Keras, Docker, CNN, faiss, nginx, Saliency object detection, MTCNN

- 微博头条推荐物料库** 03/2015 - 11/2016
- 长文本分类模型：基于标题，摘要，正文，作者等多维度，构建textCNN，多模型Stacking
  - [标题党识别模型](#)：Bi-LSTM模型
  - 相似图片排重系统
  - 技术点：Keras, scikit-learn, Elasticsearch, MySQL, phash

## 个人项目

- 
- KDD Cup(Authorship Disambiguation)** 里昂，法国 03/2013 - 06/2013
- 对微软学术论文50M数据库中19M作者名字进行甄别，目的是将文章与其真实作者相联系，[\[report\]](#)
  - 多语言文本预处理处理，文本匹配度，LDA主题提取, 随机森林
  - 主要技术: R, PostgreSQL, Python,  $\LaTeX$

- Yet Another Datalog Interpreter** 里昂，法国 09/2012 - 06/2013
- 主要技术：Datalog, Ocaml, SQL, [\[report\]](#)

- 最优救援搜索路径** 杭州，中国 2010 - 2012
- 主要技术：Graph theory, Hamilton path, [\[report\]](#)

## 技能和其他

- 
- **编程语言**: Python, SQL, R, Matlab, Java
  - **工具框架**: PyTorch, Keras, scikit-learn, Flask, ELK, Flink, Docker
  - **爱好**: 网球🎾, 乒乓🏓, 太极🧘, 滑雪🎿, 修理🔧