

# 罗 维

工作方向：机器学习、计算广告、自然语言处理

男 | 29 岁 | 电话: (+86) 18911026978 | 邮箱: vividfree@qq.com | 博客: <http://vividfree.github.io/>

## 教育背景

2009.9-2012.7	中科院计算所 - 自然语言处理研究组	工学硕士	学位课成绩: 86.83 (保送)
2005.9-2009.7	华中科技大学 - 软件工程专业	工学学士	学位课成绩: 90.70 (排名 2/209)

## 个人技能

- ◆ 擅长的技术方向包括机器学习、计算广告、自然语言处理、用户画像等;
- ◆ 有良好的软件工程能力和代码风格, 擅长将代码模块化和工具化;
- ◆ 熟悉 C / C++ / Python / Shell 等语言和 TensorFlow / Spark / Hadoop / Hive 等计算工具;
- ◆ 善于从产品、算法、工程三方面综合起来分析业务需求, 有良好的团队合作能力和项目管理经验;
- ◆ 有良好的英语听说读写能力, IELTS(6.0) / CET-6(525) / CET-4(580)。

## 工作经历

- |              |                     |         |
|--------------|---------------------|---------|
| 2016.01 - 至今 | 360 - 商业数据部 - 数据挖掘组 | 高级算法工程师 |
|--------------|---------------------|---------|
- ◆ **Query 推荐服务** 项目 owner & 核心开发
    - ✧ 目标: 支持商业产品 (360DMP<sup>1</sup>、360 商易<sup>2</sup>、360 点睛<sup>3</sup>的账户优化等) 的 query 推荐需求。
    - ✧ Query 相似模型: 基于 DSSM 模型, 底层融合 character level 和 word level 的 embedding, 上层构建神经网络外加 residual connection, 以 cosine 距离和交叉熵损失分别作为距离度量函数和损失函数。该方法在测试集上准确率达到 95%。
    - ✧ Query 推荐服务: 基于上述神经网络, 将全网近一千万 query 转换为 embedding 向量, 基于这批稠密向量使用 product quantization 相似搜索算法组成 query 推荐服务的核心模块。
    - ✧ 正在尝试对 term weight 模型和 query 相似模型做 multi task learning。
  - ◆ **360 DMP 中的 look-alike 定向** 项目 owner & 核心开发
    - ✧ 目标: 对已知精准目标人群进行分析挖掘, 寻找到更多相似的目标受众, 扩大精准营销覆盖范围。
    - ✧ 移动 look-alike 定向: 先融合 word2vec-based、click-based、session-based、attribution-based 等方法挖掘相似 APP, 后基于 360 数亿用户在使用 APP 方面的行为建立 rank 模型, 为每个广告主 APP 计算潜在用户。
    - ✧ PC look-alike 定向: 整合 360 数亿用户在 PC 端各产品上的行为数据, 基于 LR 模型为广告主提交的每个 look-alike 任务计算扩展人群。
    - ✧ 项目副产品: 基于 bagging 思路, 研发大数据量下的 word2vec 的训练与预测工具。在相似 APP 挖掘中, 该工具与 Spark MLlib word2vec 相比, precision@5 从 69% 提高到 81%。
  - ◆ **人口属性预测 (用于 360DMP 中的特征定向)** 项目 owner & 核心开发
    - ✧ 目标: 就 360 全网用户预测人口属性, 并支持 360DMP 对基于人口属性的特征定向的需求。
    - ✧ 移动端覆盖性别、人生阶段、生活达人等维度: 先融合 attribution-based、content-based、word2vec-based、SimRank-based 等方法挖掘相似 APP, 并构建起这些维度的 APP 词典, 后基于 360 数亿用户在使用 APP 方面的行为建立 rank 模型, 为各维度各标签计算其覆盖的人群。
    - ✧ PC 端覆盖性别、年龄、购买力等维度: 整合 360 数亿用户在 PC 端各产品上的行为数据, 基于 Naïve Bayes、ME 等模型为各维度的各标签计算其覆盖的人群。性别、年龄、购买力, 准确率分别为 90.6%, 78.3%, 63.2%, AUC 分别为: 0.948, 0.813, 0.740。

<sup>1</sup> 360 DMP 主页 <http://dmp.360.cn/>

<sup>2</sup> 360 商易 主页 <http://shangyi.360.cn/>

<sup>3</sup> 360 点睛 主页 <http://e.360.cn/>

2013.12 - 2015.12

360 - 商业产品事业部 - 数据挖掘组

算法工程师

核心成员

◆ 联盟广告项目高商词触发与排序

- ✧ 高商词触发：负责 content targeting 和 query targeting 策略。
- ✧ 负责高商词排序：先在流量、地域、时间三维上研发带回退策略的一跳统计 CTR 和二跳统计 CPM，后融合一跳 CTR 和二跳 CPM 得到高商词的 rank\_score。
- ✧ 策略效果：大站的 CTR1 提升超过 180%，CPM1 提升超过 120%。
- ✧ 项目副产品：基于 MapReduce 的大规模稀疏矩阵相乘。

◆ 展示广告项目广告触发

独立负责

- ✧ 目的：旧版广告引擎直接建立从 tag 到 ad 的倒排索引，该方案存在倒排索引平均长度长，在检索时需要硬剪枝，造成 CTR 等投放效果不好。为此建立从 conjunction 到 ad 的倒排索引。
- ✧ 具体策略：conjunction 包括流量维度、定向维度（包括兴趣定向、host 重定向、关键词重定向）、广告粒度维度（包括广告、广告组、广告主、广告类目）。针对<ad, conjunction> pair 先在这三维上研发带回退策略的统计 CTR 模块，然后结合 bid price 计算<ad, conjunction> pair 的 rank\_score。
- ✧ 策略效果：CTR1 提升 30%，CPM1 提升 10%。

2012.7 - 2013.11

人民搜索 - 舆情部 - 社交网络挖掘小组

算法工程师

- ◆ 为监控微博舆情，独立负责了地域识别、情感正负面分析、识别博主所属行业等策略模块以及流式处理微博数据等工程模块；并负责设计 social business 产品，其功能包含社会化媒体管理、社交网络监听、社交账号分析和潜在客户挖掘等功能。

2011.6 - 2011.8

Hulu 北京 - 视频检索组

研发工程师（实习）

- ◆ 优化 query understanding 模块和拼写错误纠正模块。错误纠正的准确率从 75%提高至 78%。

科研经历

2011.9 - 2012.5 词语对齐快速增量式训练方法(863 项目“多语言互联网语言翻译关键技术”子课题) 独立完成

- ◆ 词语对齐是翻译模型构建流程的瓶颈，对基于无监督学习的词语对齐问题，当有新语料时，应用迭代训练收敛速度更快的 online EM 算法，以替换通常所用的 batch EM 算法的方法，实现增量式训练。
- ◆ 实验效果：模型训练速度提高 5 倍，词语对齐和机器翻译的质量与 batch EM 方式的效果持平。

2010.5 - 2010.9

2010 年国际口语机器翻译评测 (IWSLT 2010)

核心成员

- ◆ 研发汉英、英汉双方向基于规则的命名实体翻译系统；优化基于层次短语模型的统计机器翻译解码器。
- ◆ 评测成绩：总共 8 项评价指标中 6 项第一和 2 项第二（全球 11 支科研队伍参赛）。

论文 & 专利

- ◆ 在 360 工作期间，以第一作者身份提交 16 篇专利申请，其中法律状态为“有权-授权”的有 1 篇，状态为“审中-实审”的有 3 篇，状态为“审中-公开”的有 5 篇。（截止 2018 年 1 月）
- ◆ 罗维. 词语对齐的快速增量式训练方法研究. 北京大学学报（自然科学版）, 2013, 49(1).
- ◆ Hao Xiong, Jun Xie, Hui Yu, Kai Liu, Wei Luo, Haitao Mi, Yang Liu, Yajuan Lv and Qun Liu. The ICT Statistical Machine Translation Systems for the IWSLT 2010. In Proceedings of IWSLT 2010.
- ◆ 罗维, 吉宗诚, 吕雅娟, 刘群. 一种改进词语对齐的新方法. 第五届全国青年计算语言学研讨会 (YWCL2010), 武汉. 2010 年 10 月.

荣誉 & 奖励

- ◆ 在 360 工作期间：2017 年度公司优秀个人、2016 年第一季度季度之星；2017 年全年绩效位于 top 5%、多次绩效评估位于 top 20%。
- ◆ 在中科院计算所读书期间：连续 3 年获得中国科学院研究生院三好学生。
- ◆ 在华中科技大学读书期间：湖北省 2009 年优秀学士学位论文(2009.7)、华中科技大学 2009 届优秀毕业生、1 次国家奖学金、2 次校三好学生、1 次校级特优生、1 次校级优秀学生干部。