

基本情况

姓 名： 江 岭

联系电话： 15002887151

邮件地址： linecong@gmail.com

英语能力： CET-6，熟练的读写能力，简单英语听说。

技术能力

- 熟悉 C/C++、python、shell 脚本、Linux/Unix 环境编程；
- 熟悉 mongo、redis；具有 Hadoop, hbase 的使用经验（项目使用较多，但自己只是简单使用）；
- 算法设计和分析、高质量程序设计、分析和解决困难问题、学习钻研能力；
- 熟悉贝叶斯、K 近邻、决策树、逻辑回归、最大熵、SVM、LDA 等机器学习模型和算法；熟悉 UserCF/ItemCF/LFM 等智能推荐模型；基于统计建模解决实际问题；
- 较好的技术指导能力和沟通能力。

教育背景

2004.9—2007.7 北京大学计算机科学与技术系（获理学硕士学位）

2000.9 - 2004.7 电子科技大学计算机科学与工程学院（获工学学士学位）

工作经历

2011.07 至今 盛大创新院搜索主题院 高级研究员

- 盛大内部创业项目：圈乐视频搜索。
- 负责信息抽取、web 数据挖掘方向。
- 成都工程小组负责人；组织成都小组每周的学习讨论会，学习机器学习、数据挖掘相关知识，实现 demo。

2007.07-2011.06 百度网页搜索部 高级研发工程师

- 页面分析、信息抽取，用于包括网页搜索在内的多条产品线。
- 网页相关性排序模型研究（半年）。
- 技术部性能优化大赛冠军；技术部创新奖励；两次获得明星项目奖励。
- 申请一项机器学习相关的发明专利。

2005.07-08, 2006.08 百度搜索应用技术部实习 实习软件工程师

- 开发歌名、人名识别模块。该模块从 05 年 9 月至离开百度时一直在产品线上稳定运行。

典型项目经验

2011-2012 视频搜索推荐项目

- 通用的视频结构化信息提取，去重、聚合、质量控制；
- 利用自适应的方法提取论坛、社区转帖、评论信息；
- 索引归并性能优化；

- 基于统计的方法，实时监控抽取效果异常；
- 基于机器学习方法进行页面分类；
- 实时处理大量新增网页，平均延时在几秒内；
- 利用 selenium 和 phantomjs 处理动态网页，解析 javascript。

2010

国际化和本地化开发

使模块同时适应 gb 和 utf-8 编码 ,利用机器学习方法 ,使模块效果适应于多语言环境。

2009-2010

利用半监督机器学习方法进行网页子功能块解析

无需人工标注的训练数据，利用机器学习算法自动构建模型，用于识别网页中的各个功能块。

- 开发半监督机器学习框架。
- 将该机器学习框架用于多种功能块识别，均能取得良好效果。

2009

网页重复结构识别和应用

基于树匹配算法，利用统计数据建模的方法计算阈值和子树相似度。用于索引页、论坛页识别等。

2009

Swf 解析和应用

对 swf 格式的文件提取链接和文字内容，用于改进网页抓取和检索。

- 开发适用于搜索引擎的 swf 解析模块。
- 将 swf 解析功能用于搜索引擎抓取系统。

2007-2008

CSS 解析、网页视觉解析模块开发

开发用于搜索引擎的极轻量级浏览器，可获取视觉信息用于帮助网页分析。该模块已用于百度多条产品线，被数十个线上模块使用。

- 开发 CSS 解析器。
- 开发高性能、高并发、周期更新、支持数亿数据项的 CSS 数据服务器。
- 开发视觉信息解析和布局算法。

2006-2007

语义 Web Services 自动组装和智能规划算法研究

致力于自动语义 Web 服务组装和执行的研究。设计和实现 Web Services 自动组装平台；设计实现扩展的图规划算法，可用于复杂控制结构的自动组装，并证明其完备性。

2004

基于关键词相关性的网络词句搜索系统

项目总体设计，并负责设计和实现相关词、关键句和同义词查询功能。此系统基于海量网页进行统计，对给定的关键词，可给出其最相关的若干关键词，并找出最具代表性的句子描述它们之间的关系；此系统还可找出给定关键词的同义词。

性格特点

自我驱动、热爱学习和思考、喜欢挑战。曾独立钻研出魔方解法，为计算机科学名著“The Art Of Computer Programming”提出改进意见并被采纳，业余学习机器学习、数学挖掘、矩阵理论等知识并用于解决实际问题。