

# 李 晓 东

大连理工大学软件学院  
大连市经济技术开发区, 116620  
☎ +86 13426475724  
✉ lxd.dlut@gmail.com  
📄 [stephenlee.github.com](https://github.com/stephenlee)

## 教育背景

- 2011– 至今 硕士, 大连理工大学软件学院, 将于 2014 毕业, 研究方向: 异构网络挖掘.  
感兴趣方向: 机器学习, 社交网络, 数据分析
- 2007–2011 本科, 大连理工大学软件学院.

## 个人经历

### 获奖情况

- 2011– 研究生入学一等奖.
- 2009–2010 学习类二等奖学金, top 15%.
- 2007–2008 学习类二等奖学金, top 15%.

### 实习

- 2013.6.18– now 数据分析, 人人游戏, 中国北京, 移动广告平台的反作弊分析, 微博数据抓取, 微博主题分析、好友可视化、影响力分析, 新闻内容抓取, 新闻内容summary.  
Hadoop, Python, ElasticSearch
- 2010.6– 2010.9 Java工程师, 阿里巴巴有限公司, 中国杭州, 设计并实现了一个监控通知系统, 包括页面抓取, 关键字匹配, 邮件/旺旺提醒, 账户管理等模块.  
Java实现

### 竞赛

- 2013.9.1– 2013.10.31 **Kaggle**, *StumbleUpon Evergreen Classification Challenge*, top 25%,  
<http://www.kaggle.com/users/26002/stephen-lee>.

## 项目经历

- 2011–2012 **topic\_model**, 分析新浪微博微博主的博文主题, Python, gensim, 课程项目.  
代码: [https://github.com/stephenLee/topic\\_model](https://github.com/stephenLee/topic_model)
- 2012–至今 **sobot**, 基于新浪微博和人人网的机器人, 实现了定时从Reddit上抓取热门的机器学习话题然后发送到微博和人人上. 微博主页: <http://weibo.com/reddit4ml>, Python, Google App Engine, 个人项目.  
代码: <https://github.com/datahacking/sobot>
- 2012–至今 **visu**, 可视化个人的人人网好友, python, d3.js, 个人项目.  
代码: <https://github.com/stephenLee/RenrenRecipies>

---

## 学术研究

2013 **Link Prediction in Heterogeneous Networks Based on Tensor Factorization**, 基于张量对异构网络建模,通过对真实的异构网络数据实验说明了使用CP张量分解能够更好的挖掘异构网络的结构, 提高了链路预测的准确性, Journal of Engineering Science and Technology Review, *ET*.

---

## 语言技能

中文简体 母语  
英语 CET-6

---

## 计算机技能

编程语言 C(熟悉), C++, Java  
脚本 Python(熟悉), Sql, R, Bash Shell, L<sup>A</sup>T<sub>E</sub>X  
框架和库 科学库(了解numpy, scipy, matplotlib, pandas, scikit-learn), web框架(Flask, webapp2(Google App engine)), 图形库(了解d3.js)  
大数据 Hadoop(python streaming)  
工具 Emacs(常用), vim, Eclipse, git  
操作系统 Linux(Ubuntu Desktop)

---

## 其他兴趣

我也热爱开源文化, 听音乐, 游戏和运动.