胡志洋

E-Mail：huzhiyang@ict.ac.cn　　电话：156 5219 3751

求职意向

数据挖掘研发工作；软件开发工作

教育背景

2012.9-2014.7	中国科学院计算技术研究所	计算机系统结构	工学硕士
2008.9-2012.7	华中科技大学	软件工程	工学学士 **Top 2%**

实习经历

2014.1-2014.4　产品评论挖掘　百度在线网络技术有限公司　研发工程师

- *项目介绍：*计算品牌数字资产榜单，挖掘公司品牌在互联网中的重要性（内容量、检索量以及好感度）。
- *负责工作（独立完成）：*
  - 提取百度问答、百度贴吧、垂直论坛中的用户评论，汇总所有品牌的情感词；
  - 挖掘品牌之间的比较关系。
- *主要方法：*
  - 采用**决策树算法**进行产品特征抽取，主要依据候选产品特征与人工选定的句式的 PMI 值；
  - 推断上下文相关的情感词的极性，根据已知极性的情感词、句子间的连接关系；
  - 采用**朴素贝叶斯算法**进行比较句的分类，特征为：融合了比较词、品牌词、产品特征、词性的 **Class Sequence Rules**；
  - 提取品牌之间的比较关系，利用融合了比较词、品牌词、产品特征、词性的 **Label Sequence Rules**。
- *工作成果：*比较句分类准确率：93%，召回率：85%；比较关系提取准确率：90%，召回率:85%；情感词的极性判断准确率：90%。

2013.5-2013.9　广告文本过滤　阿里云计算有限公司　研发工程师

- *项目介绍：*从不同内容源获取网络小说内容，归并章节，去除广告，提供阅读服务。
- *负责工作（独立完成）：*自动化找出海量小说文本中的水印—“广告信息”，并将水印从正文中删除。
- *主要方法：*
  - 采用 **k-shingle 算法**将每句文本转化为向量，计算文本的**最小哈希签名**；
  - 采用 **LSH 算法**对数据降维，筛选出近似近邻文本对，进而计算其相似度；
  - 采用**层次聚类算法**进行聚类，对水印进行过滤。
- *工作成果：*水印识别的准确率：96%，召回率：90%。

项目经历

2013.1-至今　面向服务的未来互联网体系结构（SOFIA）　中科院计算所 973 研究项目　核心开发人员

- *项目介绍：*SOFIA 是一种面向服务的未来网络架构，融合 ICN 与 SDN 思想，服务发现与数据传输分别解耦在服务层与网络层进行处理。
- *负责工作（独立完成）：*
  - 实现 SOFIA 服务层服务标识（类 URL）的路由查找，网络层地址（类 IP）的路由查找，面向服务的内容缓存模块；
  - 实现 IP 网络穿透 SOFIA 网络互联，SOFIA 网络穿透 IP 网络互联。
- *主要方法：*
  - 将服务标识（类 URL）进行分段哈希，利用**字典树**进行前缀匹配查找；
  - 利用**线程池、内容缓冲、select**，进行 IP 流与 SOFIA 流的高效转换；
  - SDN 控制器收集网络拓扑、链路信息，集中计算最优下一跳路由，选择最佳双栈路由器进行 IP 报文与 SOFIA 报文的转换。
- *工作成果：*
  - 服务层路由查找速率：200kpps；
  - IP 网络用户可穿透 SOFIA 网络，并利用 SOFIA 网络缓存，流畅观看 HTTP 高清视频，并在第二届中国未来网络产业高峰论坛中成功演示。

获奖情况

- 2010.10　国家励志奖学金(**Top 5%**)
- 2009.09　华中科技大学“求是”奖学金(**Top 10%**)
- 2011.04　腾讯俱乐部 Android 程序设计大赛三等奖

个人技能

- 熟悉 C/C++，Java，Python，了解 C#
- 熟悉基本数据挖掘理论和方法
- 熟悉基本数据结构和算法
- 熟练使用 Hadoop streaming，Hive
- 英语：通过 CET-6

Zhiyang Hu

Mobile: (+86)1565-219-3751 E-Mail: huzhiyang@ict.ac.cn

Objective

Data Mining Engineer, Software Engineer

Education

2012.9 - 2015.7 Master Institute of Computing Technology, Chinese Academy of Sciences Compute Science

2008.9 - 2012.7 Bachelor Huazhong University of Science & Technology School of Software Engineering **Top2%**

Internship Experience

2014.1 - 2014.4 Feature-based opinion mining of customer reviews of products Baidu Designer & Developer

- *Project Description:* Mine brand reputations on the Internet based on search volume, amount of related web pages, user preference, etc.
- *Personal Responsibilities:*
  - Mine and summarize **feature-based opinion on brands** using customer reviews from Baidu Tieba and some vertical websites;
  - Mine **comparative relations** between brands.
- *Major methods:*
  - Adopt **Decision Tree Classification Model** to extract product features, using PMI values between candidate features and several discriminator phrases;
  - Identify the polarity of the context dependent opinion words, using **intra-sentence/inter-sentence conjunction rules**;
  - Adopt **Naive Bayesian Classification Model** to identify comparative sentences, using **Class Sequence Rules** (combine POS tags, comparative words, domain words);
  - Extract comparative relations using **Label Sequence Rules** (combine POS tags, comparative words, domain words).
- *Results:* Comparative sentence identification, precision: 96%, recall rate: 86%; comparative relations extraction, precision: 90%, recall rate: 85%; opinion words' polarities judgment, precision: 90%.

2013.5 - 2013.9 Remove advertisement from novel Aliyun Designer & Developer

- *Project Description:* Retrieve and integrate novels from different websites to provide reading service.
- *Personal Responsibilities:* Automatically identify advertising information in massive novel contents and remove them.
- *Major methods:*
  - Use **k-shingle algorithm** to convert sentence to 0/1 vector, and assign **min-hash signature** for each sentence;
  - Use **LSH algorithm** to perform dimension reduction to approximate similar sentence pairs and compute their similarity.
  - Adopt **hierarchical clustering algorithm** to cluster sentences and identify advertisement.
- Results: Precision: 96%, and recall rate: 90%.

Project Experience

2013.1 - now Service-Oriented Future Internet Architecture (SOFIA) NSFC 973 Program Main Designer & Developer

- *Project Description:* A service-oriented information centric network (ICN) architecture. It decouples service processing and data transmission into service layer and network layer, respectively.
- *Personal Responsibilities:*
  - Implement routing lookup algorithm in service layer and network layer, and service caching module;
  - Implement SOFIA tunneling through IP network and IP tunneling through SOFIA network.
- *Major methods:*
  - Split service names (*e.g.* URLs) into segments and adopt **longest prefix matching** in name lookup;
  - Use **thread pool, content buffer and select system call** to improve the efficiency of conversion between IP and SOFIA stream;
  - SDN controllers collect network topology and link status from data layer to compute the best next-hop router and the best dual router for the conversion between IP and SOFIA packets.
- *Results:*
  - Lookup speed in service layer: 200kpps per core;
  - IP users can benefit from the caching capacity of SOFIA network and therefore play HD videos more smoothly.
  - The system has been successfully demonstrated on the Second Future Network Development and Innovation Forum.

Awards

- National Scholarship for Encouragement (**top 5%**).
- Seeking Truth Scholarship of Huazhong University of Science and Technology (**top 10%**).
- Third Prize at Tencent Club Android Programming Contest.

Personal skills

- Familiar with C/C++, Java, Python
- Familiar with basic machine learning theory and practice
- Familiar with data structure and algorithm
- Familiar with Hadoop streaming and Hive practice
- English: CET-6