杜 亮

北京市海淀区中关村南四街 4 号,8718 信箱,100190
中国科学院 软件研究所 计算机科学国家重点实验室
18500052936
duliang@ios.ac.cn; csliangdu@gmail.com
http://lcs.ios.ac.cn/~duliang

基本信息

姓　　名：	杜亮	性　　别：	男
民　　族：	汉族	籍　　贯：	山西左权
出生年月：	1985 年 4 月	政治面貌：	中共党员

研究方向

领域	数据挖掘、机器学习、大数据分析
问题	聚类分析、特征选择、矩阵分解、文档摘要
应用	推荐系统、排序学习、文本挖掘、社交网络分析等

教育背景

2007 – 2013　**工学博士**, 计算机科学, 中国科学院, 软件研究所, 计算机科学国家重点实验室.

2003 – 2007　**工学学士**, 软件工程, 武汉大学, 国际软件学院.

工作经历

2014 年 7 月 – 至今　助理研究员, 中国科学院, 软件研究所, 北京. 从事网络大数据的表示、度量和语义理解方面的研究。

2013 年 7 月 – 2014 年 7 月　软件工程师, 阿里巴巴集团, 北京. 从事计算广告, 包括大数据分析, 搜索广告排序优化, 定向广告推荐等方面的开发和研究。

2010 年 5 月 – 2011 年 5 月　实习生, 惠普中国研究院, 北京. 研究面向区间数据的低秩非负矩阵分解方法及其在协同过滤中的应用。

项目经历

工业项目 – 阿里巴巴

[4] 基于飞天分布式平台的淘客算法迁移，负责淘客中长尾查询匹配离线任务群从云梯升级到 ODPS 平台。

[3] 基于用户近期历史行为的广告推荐，利用用户在站内的历史行为用户推荐更加精准的广告内容，负责设计并实现线上排序算法。

[2] 潮流单品排序优化，通过引入商品图像特征优化单品排序的美观度，以提升用户访问体验，负责特征抽取和排序分融合。

[1] 爱淘宝综合搜索排序优化，通过搜索打通爱淘宝内部清单/专辑/搭配/店铺等多种集合数据，实现流量的合理分配，负责设计集合类特征抽取和相关性计算。

科研项目

[4] 主持国家自然科学基金青年项目：面向多源大数据的鲁棒聚类模型与算法研究, 编号:61502289, 2016.01-2018.12。

[3] 主持中国科学院软件研究所计算机科学国家重点实验室开放课题：面向大数据的鲁棒特征选择和抽取算法研究, 编号: SYSKF1505, 2015.01-2016.12。

[2] 作为核心人员参与 973 子课题：网络大数据的表示、度量和语义理解，编号:2014CB340301, 2014.1-2018.12。

[1] 硕博连读期间先后受 (1) 互联网环境中文言语信息处理与深度计算的基础理论和方法, 973 子课题，编号 2013CB329305；(2) 通用型回答集程序关键问题研究，NSFC 面上项目；(3) 抽象约束回答集程序关键问题及在语义 Web 中的应用研究，NSFC 面上项目；(4) 非命题化的回答集程序推理算法、系统实现以及典型应用研究，NSFC 面上项目等资助。

主要研究成果

已发表学术论文 20 多篇，其中包括数据挖掘领域顶级国际会议 ACM Knowledge Discovery and Data Mining (KDD 2015), IEEE International Conference on Data Mining (ICDM 2015, 2014, 2013, 2012, 2010)，人工智能领域的顶级国际会议 International Joint Conference on Artificial Intelligence (IJCAI 2015, 2013)，AAAI Conference on Artificial Intelligence (AAAI 2015)，数据工程领域的顶级国际期刊 IEEE Transaction on Knowledge and Data Engineering (TKDE) 等。

Highlights KDD(1), IJCAI(4), AAAI(1), TKDE(1), ICDM(6), SDM(2), CIKM(1).

* indicates corresponding author

[24] Hanmo Wang, **Liang Du***, Peng Zhou, Lei Shi, YuHua Qian and Yi-Dong Shen. Localized Multiple Kernel Experimental Design. in Proceedings of the Fifteenth IEEE International Conference on Data Mining (**ICDM**), pages xxx-xxx, Atlantic City, NJ, USA, November 14–17, 2015, To appear. (Full paper, acceptance rate xxx/xxx = 8.4%).

[23] NanNan Gu, MingYu Fan, Di Wang, LiHao Jia and **Liang Du**. Semi-supervised classification based on affine subspace sparse representation. Science in China-Series F: Information Sciences (中国科学), 2015, To appear.

[22] **Liang Du**, Yi-Dong Shen. Unsupervised Feature Selection with Adaptive Structure Learning. in Proceedings of the 21th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), pages 209-218, Sydney, Australia, August 10–13, 2015, To appear.

[21] Nannan Gu, Mingyu Fan, **Liang Du**, Dongchun Ren. Efficient Sequential Feature Selection Based on Adaptive Eigenspace Model. Neurocomputing, August 2015, Volume 161, Pages 199–209.

[20] **Liang Du**, Peng Zhou, Lei Shi, Hanmo Wang, Mingyu Fan, Wenjian Wang, Yi-Dong Shen. Robust Multiple Kernel K-means Clustering using $\ell_{2,1}$-norm. in Proceedings of the Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (**IJCAI**), pages 3476-3482, Buenos Aires, Argentinean, July 25-31, 2015. (Oral paper, acceptance rate 575/1996 = 28.8%).

[19] Peng Zhou, **Liang Du***, Lei Shi, Hanmo Wang, Yi-Dong Shen. Robust Multiple Kernel Learning. in Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (**IJCAI**), pages 4105-4111, Buenos Aires, Argentinean, July 25-31, 2015. (Oral paper, acceptance rate 575/1996 = 28.8%).

[18] Peng Zhou, **Liang Du***, Hanmo Wang, Lei Shi, Yi-Dong Shen. Learning a Robust Consensus Matrix for Clustering Ensemble via Kullback-Leibler Divergence Minimization. in Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (**IJCAI**), pages 4112-4118, Buenos Aires, Argentinean, July 25-31, 2015. (Oral paper, acceptance rate 575/1996 = 28.8%).

[17] Peng Zhou, **Liang Du**, Mingyu Fan, Yi-Dong Shen. An LLE based Heterogeneous Metric Learning for Cross-media Retrieval. In Proceedings of the Eleventh SIAM International Conference on Data Mining (**SDM**), pages 64-72, Vancouver, British Columbia, Canda, April 30-May 2, 2015. (Full paper, acceptance rate 72/491 = 14.66%).

[16] Hanmo Wang, **Liang Du**, Peng Zhou, Lei Shi, Yi-Dong Shen. Convex Batch Mode Active Sampling via α-relative Pearson Divergence. in Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (**AAAI**), pages 3045-3051, Austin Texas, USA, January 25–30, 2015.

[15] Liang Wu, **Liang Du**, Bo Liu, Guandong Xu, Yong Ge, Yanjie Fu, Yuanchun Zhou, Jianhui Li and Hui Xiong. Heterogeneous Metric Learning with Content-based Regularization for Software Artifact Retrieval. The 14th IEEE International Conference on Data Mining (**ICDM**), pages 610-619, Shenzhen, China, December 14-17, 2014. (Full paper, acceptance rate 71/727 = 9.7%).

[14] Lei Shi, **Liang Du**, and Yi-Dong Shen. Robust Spectral Learning for Unsupervised Feature Selection. The 14th IEEE International Conference on Data Mining (**ICDM**), pages 977-982, Shenzhen, China, December 14-17, 2014.

[13] **Liang Du**, Zhiyong Shen, Xuan Li, Peng Zhou and Yi-Dong Shen. Local and Global Discriminative Learning for Unsupervised Feature Selection. The 13th IEEE International Conference on Data Mining (**ICDM**), pages 131-140, Dallas, TX, USA, December 7-10, 2013. (Full paper, acceptance rate $94/809 = 11.6\%$).

[12] Jun Deng, **Liang Du**, and Yi-Dong Shen. Heterogeneous Metric Learning for Cross-Modal Multimedia Retrieval, The 14th International Conference on Web Information System Engineering (WISE), pages 43-56, Nanjing, China, October 13-15, 2013.

[11] **Liang Du** and Yi-Dong Shen. Towards robust co-clustering. The 23rd International Joint Conference on Artificial Intelligence (**IJCAI**), pages 1317-1322, Beijing, China, August 3-9, 2013. (Oral paper, acceptance rate $195/1473 = 13.2\%$).

[10] **Liang Du** and Yi-Dong Shen. Joint clustering and feature selection. The 14th International Conference on Web-Age Information Management (WAIM), pages 253-264, Beidaihe, China, June 14-16, 2013.

[9] **Liang Du**, Yi-Dong Shen, Zhiyong Shen, Jianying Wang and Zhiwu Xu. A self-supervised framework for clustering ensemble. The 14th International Conference on Web-Age Information Management (WAIM), pages 253-264, Beidaihe, China, June 14-16, 2013.

[8] Xuan Li, **Liang Du** and Yi-Dong Shen. Update summarization via graph-based sentence ranking. IEEE Transactions on Knowledge and Data Engineering (**TKDE**), May 2013, vol.25, no.5, pages 1162-1174.

[7] Liang Wu, Alvin Chin, Guandong Xu, **Liang Du**, Xia Wang, Kangjian Meng, Yonggang Guo and Yuanchun Zhou. Who Will Follow Your Shop? Exploiting Multiple Information Sources in Finding Followers. Database Systems for Advanced Applications (DASFAA), pages 401-415, Wuhan, China, April 22-25, 2013.

[6] **Liang Du**, Xuan Li and Yi-Dong Shen. Robust nonnegative matrix factorization via half-quadratic minimization. In Proceedings of IEEE 12th International Conference on Data Mining (**ICDM**), pages 201-210, Brussels, Belgium, December 10-13, 2012. (Full paper, acceptance rate $81/756 = 10.7\%$).

[5] **Liang Du**, Xuan Li and Yi-Dong Shen. Cluster ensembles via weighted graph regularized nonnegative matrix factorization. Advanced Data Mining and Applications (ADMA), pages 215-228, Beijing, China, December 18-20, 2011.

[4] **Liang Du**, Xuan Li and Yi-Dong Shen. User graph regularized pairwise matrix factorization for item recommendation. Advanced Data Mining and Applications (ADMA), pages 372-385, Beijing, China, December 18-20, 2011.

[3] Xuan Li, **Liang Du** and Yi-Dong Shen. Graph-based marginal ranking for update summarization. In Proceedings of the Eleventh SIAM International Conference on Data Mining (**SDM**), pages 486-497, Arizona, USA, April 28-30, 2011.

[2] Zhiyong Shen, **Liang Du**, Xukun Shen and Yi-Dong Shen. Interval-valued matrix factorization with applications. In Proceedings of the IEEE 10th International Conference on Data Mining (**ICDM**), pages 1037-1042, Sydney, Australia, December 14-17, 2010.

[1] Xuan Li, Yi-Dong Shen, **Liang Du** and Chen-Yan Xiong. Exploiting novelty, coverage and balance for topic-focused multi-document summarization. In Proceedings of the 19th ACM international conference on Information and knowledge management (**CIKM**), pages 1765-1768, Toronto, Canada, October 26-30, 2010.

学术活动

Journal Reviewer	IEEE Transactions on Knowledge and Data Engineering (TKDE).
	Data Mining and Knowledge Discovery (DAMI).
	Information Sciences
	Neurocomputing.
	International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI).
PC Member	The 20th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2016.
	The workshop, The DAta mining meets Visual Analytics at big data era (DAVA 2015), on The IEEE International Conference on Data Mining (ICDM) 2015.
	The IEEE International Conference on Advanced and Trusted Computing (ATC) 2015.
External Reviewer	The ACM Conference on Knowledge Discovery and Data Mining (KDD) 2010, 2013, 2015.
	The International Joint Conference on Artificial Intelligence (IJCAI) 2011, 2015.

The AAAI Conference on Artificial Intelligence (AAAI) 2012, 2013, 2015, 2016.

The ACM Conference on Information and Knowledge Management (CIKM) 2011.

The International Semantic Web Conference (ISWC) 2010, 2011.

The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2011, 2012, 2013.

The Pacific Rim International Conference on Artificial Intelligence (PRICAI) 2010, 2012, 2014.

The International Conference on Software Engineering and Knowledge Engineering (SEKE) 2010,2011.

专业技能

具有坚实的专业基础知识与技能，有强烈的工作责任心、进取和团队协作精神；

已有多篇高水平论文发表，具备独立从事科研工作的能力，包括科研项目调研申请，论文撰写发表，原型系统开发等；

互联网工业界顶级公司一年的实践经历，具有实际项目的开发管理能力和业务推动能力。

具有良好的沟通能力，较强的英文专业文献阅读、写作及一定的口语交流能力；

团队管理经验

2014 年 7 月 – 至今　参与指导 4 名博士研究生开展多源异构数据鲁棒融合、异构数据语义关系度量、数据采样等方面的研究，其中 8 篇论文已发表，另有多篇论文待发表。

获奖情况

2013 年　中国科学院院长奖

2013 年　北京市优秀毕业生

2013 年　中国科学院大学优秀毕业生

2012 年　IEEE ICDM 2012 Student Travel Award

技能

语言　英语

程序语言　Hadoop, Java, C++, Python, Matlab, LaTeX,

操作系统　Linux, Windows