

鲍军威

(+86) 18510919126 | baojunwei001@126.com | 北京



教育经历

- | | |
|--|---------------------|
| 哈尔滨工业大学 | 2008年07月 - 2012年06月 |
| 计算机科学与技术 本科 计算机学院 | 哈尔滨 |
| • GPA : 3.4 / 4.0 (专业前10%) 荣誉/奖项 : 一等奖学金 (2011), 国家奖学金 (2011) | |
| 哈尔滨工业大学 | 2008年07月 - 2012年06月 |
| 金融学双学位 本科 管理学院 | 哈尔滨 |
| • 金融学, 宏观经济学, 微观经济学, 财务学, 银行学 | |
| 哈尔滨工业大学 | 2012年07月 - 2014年06月 |
| 计算机科学与技术 硕士 计算机学院 | 哈尔滨 |
| • 导师: 赵铁军 荣誉/奖项: 国家奖学金 (2013) | |
| 哈尔滨工业大学 | 2014年07月 - 2018年12月 |
| 计算机科学与技术 博士 计算机学院 | 北京 |
| • 导师: 周明 研究方向: 自然语言处理 (NLP), 知识图谱问答 (KBQA), 语义解析 (SP), 文本生成 (NLG) | |

项目经历

- | | |
|--|---------------------|
| 微软亚洲研究院 (MSRA) | 2013年05月 - 2018年12月 |
| 联合培养博士 @ 自然语言计算组(NLC). Supervisor: 周明. Mentor: 段楠. | 北京 |
| • KBQA项目 (13年5月-16年3月)
搭建基于Microsoft Satori/Google Freebase的自动问答系统。通过采用CKY的解码方式, 动态地自底向上地解析 question中每一个片段, 直至整个问题的方式使得在WebQuestions上准确率取得6.1%提升。此项目与Bing搜索引擎多个组合作, 为Bing问答引擎服务。相关论文见[28]。鉴于已有KB-QA数据集主要是single-relation的问题, 我们构建了包含2100条复杂问答对的CompQ数据集。我们提出了基于多限制查询图 (MulCG) 的语义解析方法来解析含有多限制的复杂问题。在WebQuestions上, 我们的方法取得state-of-the-art的结果 (F1提升3%)。相关论文见[25,27]。 | |
| • DocChat项目 (15年1月-15年10月)
参与京东客服系统、微软小冰对话系统的意图检测与查询理解的设计与实现。包括大规模数据的分析整理, 用户查询到基于领域知识的意图表示的解析, 产品测试等。我们提出的DocChat方法在WikiQA数据集上F1取得3.4%提升。相关论文见[22,26]。 | |
| • Table-QA项目 (16年3月-16年10月)
通过精确设计的特征, 研发了基于表格进行自动问答的技术。设计了基于表格行, 列, 单元与question进行多方位匹配的算法。该方法在WikiTableQuestions的子集上ACC和F1均取得了6+%提升。该项目与Bing搜索引擎合作, 旨在帮助Bing进行关于Table的问答。相关论文见[23,24]。 | |
| • Table-to-Text项目 (16年10月-17年5月)
我们设计了Table-to-Sequence神经网络模型, 融合copying机制, 引入local和global的信息帮助解码等关键技术。我们在SimpleQuestions, WikiBIO数据集上都取得大于5个点的BLEU4的提升。我们同时提供WikiTableText数据集来帮助该任务的研究, 在该数据集上我们的T2S方法比baseline提升近10个点的BLEU4。该技术已用于Bing搜索引擎。相关论文见[18,21]。 | |
| • KBQA数据集构建项目 (17年10月-18年6月)
该项目旨在对用户query进行语义理解, 并产生可执行的逻辑表达式。项目基于大规模知识图谱 (Satori/Freebase) 构建了50K+的数据, 包含9大类问题。数据包含了问题, 语义表示, 问题模板, 实体信息, 问题类型, 答案等信息。部分工作已经上线Bing搜索引擎。相关论文见[19]。 | |
| 北京京东世纪贸易有限公司 (JD.com) | 2019年01月 - 至今 |
| 算法工程师 & 研究员 @ JD AI Research. Co-Author: 周伯文, 何晓冬, 吴友政. | 北京 |
| • 京东小智KBQA项目 (19年1月-19年6月)
该项目旨在在京东智能客服场景中对京东商城用户query进行语义理解, 并产生可执行的逻辑表达式从知识图谱中检索 answer。模型方面, 我们采用BERT, 将KBQA任务从传统的匹配任务转化成谓词生成任务。我们的模型落地于京东智能客 | |

服KBQA系统。

- **京东品创-智能营销文案生成**（19年4月-20年11月）

该项目旨在利用电商商品信息，包括商品详情、商品标题、商品知识图谱等，实现自动化的商品营销文案的生成。我们的模型攻克了文案内容一致性、连贯性、流畅性、多样性等难题，覆盖京东商城10大一级品类，包括电脑、手机、家用电器等，文案人工审核通过了超过95%。落地于京东发现好货，智能营销客服等场景。论文见[13,14]。

- **京东预训练语言模型**（20年11月-22年11月）

该项目旨在提高通用及电商、金融等子领域的文本生成质量。我们基于大规模预训练语料，提出基于概率的排列式先知预言模型，该模型在GLGE标准评测上获得了SOTA性能。我们的模型应用于京东金融客服系统中。相关论文见[1,9]。

学术经历

社区贡献与科研项目

2014年08月 - 至今

SPC & PC & 项目成员

北京

- 曾担任IJCAI2021的高级程序委员（SPC）
- 担任ACL，EMNLP，NAACL，ICLR，NeurIPS，AAAI，IJCAI，COLING，TASLP，MM等会议及期刊程序委员(PC)及审稿人。
- 发表多篇论文支撑科技部2030-2020国家项目
- 2022年申报吴文俊奖，获多名院士推荐（共同申报人）

AI竞赛

2019年01月 - 至今

Tech Lead @ JD AI Research

北京

- [任务型对话MotiWoz数据集第一名](#)（2022年10月）论文见[0,5,7]
- [推理式阅读理解DROP竞赛第一名](#)（2022年2月）论文见[6,2,15]
- [推理式混合问答TAT-QA竞赛第一名](#)（2022年2月）论文见[2]
- [对话式阅读理解QuAC竞赛第一名](#)（2021年12月）论文见[12]
- [通用文本生成GLGE竞赛第二名](#)（2021年5月）论文见[1]
- [推理式混合问答HybridQA竞赛第三名](#)（2021年1月，F1指标）论文见[3]

文章【部分】

2011年11月 - 至今

PH.D & Researcher @ MSRA & HIT & JD AI Research

北京

[Homepage](#). [Google Scholar](#). [DBLP](#). [Semantic Scholar](#). [ACL Anthology](#).

[0]. [Mars: Semantic-aware Contrastive Learning for End-to-End Task-Oriented Dialog](#).

Haipeng Sun, **Junwei Bao***, Youzheng Wu, Xiaodong He. Arxiv 2022.

[1]. [P3LM: Probabilistically Permuted Prophet Language Modeling for Generative Pre-Training](#).

Junwei Bao*, et al. **EMNLP 2022(Findings)**.

[2]. [UniRPG: Unified Discrete Reasoning over Table and Text as Program Generation](#).

Yongwei Zhou, **Junwei Bao***, et al. **EMNLP 2022**. *Mentor.

[3]. [MuGER2: Multi-Granularity Evidence Retrieval and Reasoning for Hybrid Question Answering](#).

Yingyao Wang, **Junwei Bao***, et al. **EMNLP 2022(Findings)**. *Mentor.

[4]. [AutoQGS: Auto-Prompt for Low-Resource Knowledge-based Question Generation from SPARQL](#).

Guanming Xiong, **Junwei Bao***, et al. **CIKM 2022**. *Corresponding Author.

[5]. [LUNA: Learning Slot-Turn Alignment for Dialogue State Tracking](#).

Jing Zhao, Yifan Wang, **Junwei Bao***, Youzheng Wu, Xiaodong He. **NAACL 2022**. *Corresponding Author.

[6]. [OPERA: Operation-Pivoted Discrete Reasoning over Text](#).

Yongwei Zhou, **Junwei Bao***, et al. **NAACL 2022**. *Mentor.

[7]. [BORT: Back and Denoising Reconstruction for End-to-End Task-Oriented Dialog](#).

Haipeng Sun, **Junwei Bao***, Youzheng Wu, Xiaodong He. **NAACL 2022(Findings)**. *Corresponding Author.

[8]. [Don't Take It Literally: An Edit-Invariant Sequence Loss for Text Generation](#).

Guangyi Liu, Zichao Yang, Tianhua Tao, Xiaodan Liang, **Junwei Bao**, et al. **NAACL 2022**.

[9]. [Fine- and Coarse-Granularity Hybrid Self-Attention for Efficient BERT](#).

- Jing Zhao, Yifan Wang, **Junwei Bao***, Youzheng Wu, Xiaodong He. **ACL 2022**. *Corresponding Author.
- [10]. [A Roadmap for Big model. Chapter15: Application in Text Generation](#).
Haoran Li*, **Junwei Bao***, Yingwei Pan*, Xiaodong He. CoRR 2022. *Equal Contribution.
- [11]. [RevCore: Review-Augmented Conversational Recommendation](#).
Yu Lu, **Junwei Bao***, et al. **ACL 2021(Findings)**. *Corresponding Author.
- [12]. [RoR: Read-over-Read for Long Document Machine Reading Comprehension](#).
Jing Zhao, **Junwei Bao***, et al. **EMNLP 2021(Findings)**. *Corresponding Author.
- [13]. [SGG: Learning to Select, Guide, and Generate for Keyphrase Generation](#).
Jing Zhao, **Junwei Bao***, Youzheng Wu, Xiaodong He, Bowen Zhou. **NAACL 2021**. *Corresponding Author.
- [14]. [CUSTOM: Aspect-Oriented Product Summarization for E-Commerce](#).
Jiahui Liang, **Junwei Bao***, Yifan Wang, Youzheng Wu, Xiaodong He, Bowen Zhou. **NLPCC 2021**.
*Corresponding Author.
- [15]. [EviDR: Evidence-Emphasized Discrete Reasoning for Reasoning Machine Reading Comprehension](#).
Yongwei Zhou, **Junwei Bao***, et al. **NLPCC 2021**. *Mentor.
- [16]. [CUGE: A Chinese Language Understanding and Generation Evaluation Benchmark](#).
Yuan Yao,...,**Junwei Bao**, et al., CoRR 2021.
- [17]. [Learning to Decouple Relations: Few-Shot Relation Classification with EGA and CAT](#).
Yingyao Wang, **Junwei Bao***, et al. **COLING 2020**. *Mentor.
- [18]. [Text Generation from Tables](#).
Junwei Bao, Duyu Tang, Nan Duan, Zhao Yan, Ming Zhou, Tiejun Zhao. **TASLP 2019**.
- [19]. [Weakly Supervised Multi-task Learning for Semantic Parsing](#).
Bo Shao, Yeyun Gong, **Junwei Bao**, Jianshu Ji, Guihong Cao, Xiaola Lin, Nan Duan. **IJCAI 2019**.
- [20]. [Question Generation with Doubly-Adversarial Nets](#).
Junwei Bao, Yeyun Gong, Nan Duan, Ming Zhou, Tiejun Zhao. **TASLP 2018**.
- [21]. [Table2Text: Describing Table Region with Natural Language](#).
Junwei Bao, Duyu Tang, Nan Duan, Zhao Yan, Yuanhua Lv, Ming Zhou, Tiejun Zhao. **AAAI 2018**.
- [22]. [Response selection from unstructured documents for human-computer conversation systems](#).
Zhao Yan, Nan Duan, **Junwei Bao**, Peng Chen, Ming Zhou, Zhoujun Li, Jianshe Zhou. **KBS 2018**.
- [23]. [Content-Based Table Retrieval for Web Queries](#).
Zhao Yan, Duyu Tang, Nan Duan, **Junwei Bao**, Yuanhua Lv, Ming Zhou, Zhoujun Li. CoRR 2017.
- [24]. [An Information Retrieval Based Approach to Table-Based Question Answering](#).
Junwei Bao, Nan Duan, Ming Zhou, Tiejun Zhao. **NLPCC 2017**.
- [25]. [Constraint-Based Question Answering with Knowledge Graph](#).
Junwei Bao, Nan Duan, Zhao Yan, Ming Zhou, Tiejun Zhao. **COLING 2016**.
- [26]. [DocChat: An Information Retrieval Approach for Chatbot Engines Using Unstructured Documents](#).
Zhao Yan, Nan Duan, **Junwei Bao**, Peng Chen, Ming Zhou, Zhoujun Li, Jianshe Zhou. **ACL 2016**.
- [27]. [Answering Questions with Complex Semantic Constraints on Open Knowledge Bases](#).
Pengcheng Yin, Nan Duan, Ben Kao, **Junwei Bao**, Ming Zhou. **CIKM 2015**.
- [28]. [Knowledge-Based Question Answering as Machine Translation](#).
Junwei Bao, Nan Duan, Ming Zhou, Tiejun Zhao. **ACL 2014**.
- [29]. [Query Rewriting Using Statistical Machine Translation](#).
Junwei Bao, De Quan Zheng, Bing Xu, Tiejun Zhao. ICMLC 2013.
- [30]. [技术重构社会供应链：未来科技趋势白皮书](#).
鲍军威 等. 京东技术JDD 2021.
- [31]. [基于序列到序列模型的文本到信息框生成的研究](#).
鲍军威, 周明, 赵铁军. 智能计算机与应用 2019.

[32]. [基于知识的自动问答与问题生成的研究](#).

鲍军威. 导师：周明. 哈尔滨工业大学. 博士学位论文 2018.

[33]. [基于知识的自动问答研究](#).

鲍军威. 导师：赵铁军. 哈尔滨工业大学. 硕士学位论文 2014.

发名专利

2019年01月 - 至今

Research Scientist @ JD AI Research

北京

[1] ZL202110437033.2 《一种高效实现多样性自动文案写作的方法》 京东集团-京东云与AI事业部：王逸凡,鲍军威（专利局审查阶段JDZL2020N6714）

[2] ZL202111244062.3 《一种基于ELECTRA与自评测的机器阅读理解算法》 京东集团-京东云与AI事业部：赵靖,鲍军威（专利局审查阶段JDZL2020N5735）

[3] ZL202110085890.0 《一种基于BERT聚类 and 动态选择机》 京东集团-京东云与AI事业部：梁嘉辉,鲍军威,吴友政（专利局审查阶段JDZL2020N5034）

[4] ZL202110117687.7 《一种在解码器端融入图注意力》 京东集团-京东云与AI事业部：梁嘉辉,鲍军威,吴友政（专利局审查阶段JDZL2020N5462）

[5] ZL202110188846.2 《一种基于图卷积神经网络建模》 京东集团-京东云与AI事业部：梁嘉辉,鲍军威（专利局审查阶段JDZL2020N6140）

[6] ZL202110476637.8 《一种联合抽取式和生成式摘》 京东集团-京东云与AI事业部：梁嘉辉,鲍军威（专利局审查阶段JDZL2020N6802）