Kuntal Kumar Pal

Ph.D. Student (Fifth Year), CSE (NLP) Arizona State University, CGPA: 3.97

Profiles

Github:// kuntalkumarpal LinkedIn:// kuntalkumarpal Scholar: Kuntal Kumar Pal Website: kuntalkumarpal.github.io

Education

Ph.D.(Natural Language Processing) Aug 2018 - May 2023 [Expected] Arizona State University, Tempe, USA CGPA: 3.97/4

Master of Technology (CSE) , 2014-2016, National Institute of Technology, Calicut, India CGPA: 8.80/10

Bachelor of Engineering (IT) , 2005-2009, Jadavpur University, Kolkata, India CGPA: 8.53/10

Skills

ML/DL

Pytorch, Pytorch-lightning, Transformers, SpaCy, Numpy, Scikit-Learn, Pandas, OpenCV, Jupyterlab, Google Colab

Languages & DB Python, C/C++, Core Java, SQL

Utilities

Markdown, Git, Vim, JIRA, Sublime Text, Unix, Latex, GDB, PlSql Developer

Coursework

Natural Language Processing Biomedical NLP Machine Learning Data Mining Knowledge Representation Pattern Recognition Information Extraction

Achievements

- Top 20 team, MSFT AI Challenge
- ASU CIDSE Doctoral Fellowship
- Engineering Graduate Fellowship
- Graduate Research Assistantship
- TCS Best Team Award
- Contributor: Al2-Natural Instructions

Research Projects

Generalization via In-Context Instructions (1,600+ NL Tasks) EMNLP, 2022

- Introduced a large benchmark of 1600+ natural language tasks with 70+ distinct tasks.
- Quantified generalization of transformers on unseen task with natural instructions

Bio-Medical NER via Knowledge Guidance and QA ACM JOURNAL (HEALTH), 2021

- Reformulated NER as multi-answer QA task with knowledge contexts on 18 datasets.
- Applied entity-type, definition, examples and their combination as knowledge types.
- Pushed state-of-the-art exact match F1 scores by 1.78-12% for 11 biomedical datasets.

Flow Graph Construction from Cybersecurity Texts FINDINGS ACL, 2021

- Proposed novel NLU task of flow-graph structure generation from public forum texts.
- Introduced large (3154) security procedural text dataset for information extraction.
- Achieved 0.72 PRAUC with good generalization to other domains (0.63, 0.68 PRAUC).

Numerical Reasoning Ability of Text-To-Text ModelFINDINGS EMNLP, 2021• Analyzed T5 model performs 36%-90% less in out-of-domain number range than in-
domain on numeration, magnitude comparison, list-minmax and list-sorting tasks.• Created FeasibilityQA, GPT-3 shows poor action feasibility commonsense reasoning.

Commonsense Reasoning in QA with Implicit KnowledgeAKBC, 2021• Showed that language models with external knowledge can achieve performance
comparable to larger and complex models on 3 commonsense reasoning question an-
swering datasets which involves story understanding, physical and social interactions.• Introduced 3 strategies and 4 methods of incorporating external knowledge with com-
mon framework and improved upto 2% on 2 of 3 such datasets.

OpenbookQA-Careful selection of knowledge(Multistep Reasoning) ACL 2019

- Achieved 11.6% improvement over existing state-of-the-art on OpenBookQA dataset.
 Achieved 72% multiple choice QA accuracy with abductive information retrieval, in-
- formation gain based re-ranking of extracted knowledge and BERT-based reasoning.
- Categorized failed cases to temporal, qualitative, conjunctive and logical reasoning.

Source Code Variable Name Recovery from Decompiled Binaries PREPRINT • Recovered variable names, lost during compilation from 165K C binaries with 84.15% accuracy. Suggested quality names to decompiler generated temporary variables. • Proposed joint varname and origin prediction BERT model using constrained MLM.

Predicting Facebook-Users' Personality from their Status
 Predicted the BIG5 personality traits of Facebook users from their status updates using regression and topic analysis along with other linguistic features.

Experience

2022 Research Intern, Microsoft Knowledge, Reasoning, Language (KRL) Performed an in-depth question tagging behavior analysis on 17 subdomains of StackExchange CQA platform. Developed and evaluated a transformer-based model for tag suggestion which predicts both popular and fine-grained tags.

2016-2018 Engineer I, Cavium Networks

Programmable Network Switch

Developed a proprietary shell-based testing platform for configuring network switches. Worked in a team of 2 on proprietary Xpliant Compiler and delivered to client for configuring switches using forwarding data-plane instructions.

2009-2014 **IT Analyst, Tata Consultancy Services** Data Migration, Insurance Migrated successfully over 2M insurance policy data from 4 UK legacy systems to TCS B α NCs. Reduced database deployment time from 2 days to 30 mins with auto-deployment, delivery and version control tool developed in-house.

1216, E Vista Del Cerro Dr, APT 1129 Tempe, Arizona 85281, United States Mob: +1-4802772474, kkpal@asu.edu