Tapas Nayak

CONTACT INFORMATION

Mobile: +91-9564509849
E-mail : tnk02.05@gmail.com
Github: https://github.com/nayakt
LinkedIn: https://www.linkedin.com/in/tapas-nayak-95004a195/
Scholar: https://scholar.google.com/citations?user=4AgZ2VYAAAAJ&hl=en
Website: https://nayakt.github.io/

RESEARCH INTERESTS

I am interested in natural language processing and understanding using machine learning and deep learning approaches. Currently, as a postdoctoral researcher at IIT Kharagpur, I am working on multiple topics such as aspect-sentiment extraction, question-answering in e-commerce, information extraction in material science domain and financial domain. During my PhD, I have focused on relation extraction using deep neural networks from the text for the enrichment of knowledge bases. I would also like to explore other NLP tasks such as open domain question-answering, machine translation, and information extraction in the medical domain. Also, I want to explore the NLP research for Indic languages for the tasks mentioned before.

CURRENT WORK

Post Doctoral Researcher - IIT Kharagpur Sept, 2020 - Present
Advisor: Assoc. Prof. Pawan Goyal

Currently I am working on application of deep neural networks in different fields of natural language processing such as information extraction from material science articles, question-answering based on e-commerce reviews, aspect sentiment extraction, and analyzing documents from financial domain. I have guided 1 MTech student in his masters thesis and currently, I am collaborating with 3 PhD students and 1 MS student.

ACADEMIC BACKGROUND

- **Ph.D. in Computer Science** 2016 – 2020
  Thesis: Deep Neural Networks for Relation Extraction
  Advisor: Prof. Hwee Tou Ng
  Department of Computer Science, School of Computing
  National University of Singapore, Singapore

CGPA: 4.75 / 5

- **Master of Computer Sc. & Engineering** 2014 – 2016
  Thesis: Computer Aided Translation & Automatic Post Editing
  Advisor: Assoc. Prof. Sudip Kumar Naskar
  Department of Computer Science & Engineering
  Jadavpur University, Kolkata, India
  CGPA: 9.67 / 10

- **Bachelor of Computer Sc. & Engineering** 2005 – 2009
  Department of Computer Science & Engineering
  Jadavpur University, Kolkata, India
  CGPA: 8.71 / 10

AWARDS & FELLOWSHIPS

Research Achievement Award, 2020, National University of Singapore, Singapore
NUS Research Scholarship, 2016-2020, MoE, Singapore
Amitava Dey Memorial Gold Medal, 2016, Jadavpur University, India
University Medal, 2016, Jadavpur University, India
GATE Scholarship, 2014-2016, AICTE, India

INDUSTRY EXPERIENCE

Experience: 5 years.

PricewaterhouseCoopers India: 5th Dec, 2012-29th Sept, 2014
Position: Senior Consultant
Location: Salt Lake, Kolkata, India
Projects: I worked as a senior consultant at PwC and led a team of size 3-4 in two different projects. The first project was based on ASP.NET and the second project was based on the C# Windows Presentation Framework (WPF). They were used for the purpose of audits which was the main business of PwC globally. Some of the important features of C# which were used in these tools are GUI, WPF, WCF (Windows Communication Framework), multi-threading, delegates, events, design patterns, etc.

Hewlett-Packard India Software Operations: 5th Aug, 2009 – 9th Nov, 2012
Position: SW Engr Firmware I
Location: Bengaluru, India
Projects: I worked as a software developer here and contributed to build thermal inkjet characteristics tools such as drop weight, drop velocity, and resistor life test. These tools were used to measure the quality of droplets of printer cartridges. Drop weight was used to measure the average weight of a droplet. Drop velocity was used to measure the velocity at which a droplet drops on the paper. Resistor life test was used to determine the lifetime of a cartridge. They have a software and a hardware component. My work was on

the software side to communicate with the hardware. We used C#.Net to build these tools. Some of the important features of C# which are used in these tools are GUI, multi-threading, serial connection, delegates, events, design patterns, etc.

SKILL SET

- **Research Area:** Information extraction, relation extraction, knowledge base population, aspect-sentiment extraction, question-answering, machine translation, natural language understanding.
- **Programming Languages:** C, C++, C#, Java, Python
- **Deep Learning Tools:** Keras, Pytorch, Tensorflow
- **Operating Systems:** Windows, Ubuntu

PUBLICATIONS (* means equal contribution)

1. Rajdeep Mukherjee*, **Tapas Nayak***, Yash Butala, Sourangshu Bhattacharya and Pawan Goyal. PASTE: A Tagging-free Decoding Framework using Pointer Networks for Aspect Sentiment Triplet Extraction. EMNLP 2021.

2. **Tapas Nayak** and Hwee Tou Ng. A Hierarchical Entity Graph Convolutional Network for Relation Extraction across Documents. RANLP 2021.
   Code: https://github.com/nusnlp/MHRE

3. **Tapas Nayak**, Navonil Majumder and Soujanya Poria. Improving Distantly Supervised Relation Extraction with Self-Ensemble Noise Filtering. RANLP 2021.
   Code: https://github.com/nayakt/SENF4DSRE

4. Samson Yu, **Tapas Nayak**, Navonil Majumder, and Soujanya Poria. Aspect Sentiment Triplet Extraction using Reinforcement Learning. CIKM 2021 (Short).
   Code: https://github.com/declare-lab/aste-rl

5. **Tapas Nayak,** Navonil Majumder, Pawan Goyal, and Soujanya Poria. Deep Neural Approaches to Relation Triplets Extraction: A Comprehensive Survey. Cognitive Computation, 2021

6. **Tapas Nayak** and Hwee Tou Ng. Effective Modeling of Encoder-Decoder Architecture for Joint Entity and Relation Extraction. In the Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, New York, USA, 2020.
   Code: https://github.com/nusnlp/PtrNetDecoding4JERE

7. **Tapas Nayak** and Hwee Tou Ng. Effective Attention Modeling for Neural Relation Extraction. In the Proceedings of the SIGNLL Conference on Computational Natural Language Learning (CoNLL), Hong Kong, 2019.

Code: https://github.com/nusnlp/MFA4RE

8. Santanu Pal, Sudip Naskar, Marcos Zampieri, **Tapas Nayak**, and Josef van Genabith. CATaLog Online: A Web-based CAT Tool for Distributed Translation with Data Capture for APE and Translation Process Research. In the Proceedings of The 26th International Conference on Computational Linguistics (COLING): System Demonstrations, Osaka, Japan, 2017.

9. **Tapas Nayak**, Santanu Pal, Sudip Kumar Naskar, Sivaji Bandyopadhyay, and Josef van Genabith. 2016. Beyond Translation Memories: Generating Translation Suggestions based on Parsing and POS Tagging. In the Proceedings of the 2nd Workshop on Natural Language Processing for Translation Memories (NLP4TM), Portorož, Slovenia, 2016.

10. Santanu Pal, Marcos Zampieri, Mihaela Vela, **Tapas Nayak**, and Sudip Kumar Naskar, Josef van Genabith. 2016. CATaLog Online: Porting a Post-editing Tool to the Web. In the Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC), 2016.

11. **Tapas Nayak**, Sudip Kumar Naskar, Santanu Pal, Marcos Zampieri, Mihaela Vela, and Josef van Genabith. CATaLog: New Approaches to TM and Post Editing Interfaces. In the Proceedings of the Workshop on Natural Language Processing for Translation Memories (NLP4TM), Hissar, Bulgaria, 2015.

SCIENTIFIC RESPONSIBILITIES

PC Member for Computational Linguistics Journal (CL), ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), Information Fusion Journal, NeuroComputing Journal, Natural Language Engineering Journal, ICON 2020, AAAI 2021, IJCAI 2021, EMNLP 2021, CIKM 2021, SustaiNLP 2021, AAAI 2022.

RESEARCH PROJECTS

Phd Thesis: During my Ph.D., I have worked in the area of relation extraction from text for knowledge base enrichment. I have explored different deep neural network models for this task. First, I have explored a pipeline approach where two entities are given and we need to find the relation between them, or that no relation exists between them. In our second work, I have explored a joint extraction approach for entities and relations. I have proposed deep neural models for this task and achieve significantly improved performance with both approaches when evaluated on publicly available relation extraction datasets. In my third work, I have explored a new multi-hop relation extraction task, where we use multiple documents to find relation tuples. This can help to extract a higher number of relations from knowledge bases than sentence-level relation extraction.

PhD Course Projects:

1. Implemented a composition and decomposition-based similarity and dissimilarity neural network model for an answer sentence selection task. The task is about ranking the answers for a given question from a set of answers. Code is available at https://github.com/nayakt/Answer-Sentence-Selection

2. Implemented a neural model for Community Question Answering task in SemEval 2016. The task is about, given a new question, rank the answers from the past question-answers thread. More details about the task can be found about the task at http://alt.qcri.org/semeval2016/task3/. Code is available at https://github.com/nayakt/CQA
3. Implemented a query focus document retrieval system. It can index a large set of documents and then retrieve the ranked documents based on a query. This can be used in small organizations to search for documents. Code is available at https://github.com/nayakt/DocumentRetriever
4. Implemented a useful system at NUS for automatic evaluation of programming assignments (neural models and non-neural models both). This is extensively used in the CS4248 (Natural Language Processing) module.

Masters Thesis: During my masters, I worked on the computer-aided translation and automatic post-editing. First, I built a computer-aided translation tool based on translation memory. This tool could be used by the human translator to modify the machine translation output to make them correct. Next, I improved the suggestions of the translation memory by fusion of mismatched parts of the input sentence using POS tags and Parse tree.

TEACHING EXPERIENCE

During the course of my PhD, I have worked as a teaching assistant for following modules at School of Computing, National University of Singapore.

Module	Lecturer	Sem/AY
CS1020: Data Structures and Algorithm I	Prof. Tan Sun Teck	Sem 2 AY 16/17
CS4248: Natural Language Processing	Prof. Hwee Tou Ng	Sem 1 AY 17/18
CS1020E: Data Structures and Algorithm I	Prof. Kok-Lim Low	Sem 2 AY 17/18
CS4248: Natural Language Processing	Prof. Hwee Tou Ng	Sem 1 AY 18/19
CS2040C: Data Structures and Algorithm	Prof. Gary Tan	Sem 1 AY 18/19

REFERENCES

1) Dr. Sudip Kumar Naskar
   Associate Professor
   Department of Computer Science & Engineering
   Jadavpur University, India.
   Email: sudip.naskar@gmail.com

2) Dr. Pawan Goyal
   Associate Professor
   Department of Computer Science & Engineering
   IIT Kharagpur
   Email: pawang.iitk@gmail.com

3) Dr. Soujanya Poria
   Assistant Professor
   Singapore University of Technology and Design, Singapore
   Email: soujanya.poria@gmail.com

4) Prof. Hwee Tou Ng
   Provost's Chair Professor
   Department of Computer Science
   School of Computing
   National University of Singapore
   Email: nght@comp.nus.edu.sg