Tapas Nayak

CONTACT INFORMATION

Mobile: +65-9806-3758
E-mail : tnk02.05@gmail.com, nayakt@u.nus.edu
Github: https://github.com/nayakt
LinkedIn: https://www.linkedin.com/in/tapas-nayak-95004a195/
Website: https://nayakt.github.io/
Scholar: https://scholar.google.com/citations?user=4AgZ2VYAAAAJ&hl=en

RESEARCH INTERESTS

Language is the best medium for interaction between humans and computers. The understanding of human language is an important step towards achieving the goal of general artificial intelligence. Any kind of computerized assistance tools such as robots, home assistants, or drones, they need to process and understand the human language properly to make the correct decisions. This thought process got me into the research of natural language processing (NLP) and understanding (NLU). With the advances made in deep learning, the field of NLP and NLU has also progressed a lot. I have worked in this field for the last four years during my Ph.D. and contributed in a limited way. My major research focus was on the extraction of relational facts from the text for the enrichment of knowledge bases (KBs). Since KBs are very useful to understand the human language, I want to continue to contribute more here.

Along with this, I want to focus on neural machine translation too. There are so many languages spoken in the World, and in my own country India, there are so many languages, so many dialects spoken by more than a billion people. Many of these languages are very low resource languages where there is not enough data available to use deep learning. For universal human-computer interaction, we need to develop tools that can understand many different languages with different dialects and can also work with low resource languages. I want to work on how neural machine translation models can be used for the Indian languages so that the interaction between the people and computers becomes easy.

CURRENT WORK

- **Ph.D. in Computer Science** — Aug, 2016 – Aug, 2020
  CGPA: 4.75 out of 5 [Thesis Submitted]
  Department of Computer Science
  School of Computing
  National University of Singapore
  Singapore

PUBLICATIONS

1. **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

2. **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

3. Santanu Pal, Sudip Naskar, Marcos Zampieri, **Tapas Nayak**, 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.

4. **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.

5. 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.

6. **Tapas Nayak**, Sudip Kumar Naskar, Santanu Pal, Marcos Zampieri, Mihaela Vela, 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.

UNDER REVIEW

1. **Tapas Nayak** and Hwee Tou Ng. A Hierarchical Entity Graph Convolutional Network for Relation Extraction across Documents.

2. **Tapas Nayak** and Hwee Tou Ng. The Zero Cardinality Challenge in Distantly Supervised Relation Extraction.

3. **Tapas Nayak** and Hwee Tou Ng. Improving Distantly Supervised Relation Extraction with Self-Ensemble Noise Filtering.

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, we explore a pipeline approach where we assume that two

entities are given and we need to find the relation between them, or that no relation exists between them. In our second work, we explore a joint extraction approach for entities and relations. We 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 our third work, we explore 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 given a question and a set of answers, rank the 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.

Bachelors Final Year Project: I worked on a single document summarization project as my undergraduate project. We used the frequency of words, the title of the document, the structure of the document to summarize it. The summarization model was generic and could be used for multi-document summarization as well.

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
CS4248: Natural Language Processing	Prof. Hwee Tou Ng	Sem 1 AY 19/20

AWARDS

Research Achievement Award, 2020, National University of Singapore, Singapore
Amitava Dey Memorial Gold Medal, 2016, Jadavpur University, India
University Medal, 2016, Jadavpur University, 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 audit 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

- **Programming Languages:** C, C++, C#, Java, Python

- **Deep Learning Tools:** Keras, Pytorch, Tensorflow
- **Operating Systems:** Windows, Ubuntu

ACADEMIC BACKGROUND

- **Master of Computer Sc. & Engineering** 2014 – 16
  Department of Computer Science & Engineering
  Jadavpur University
  Kolkata, India
  CGPA: 9.67 out of 10

- **Bachelor of Computer Sc. & Engineering** 2005 – 09
  Department of Computer Science & Engineering
  Jadavpur University
  Kolkata, India
  CGPA: 8.71 out of 10

REFERENCES

1) 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

2) Dr. Sudip Kumar Naskar
   Associate Professor,
   Computer Science & Engineering Department,
   Jadavpur University, India.
   Email: sudip.naskar@gmail.com / sudip.naskar@cse.jdvu.ac.in