

Debjit Paul

Am Geisenberg 4

66125 Saarbrücken

Germany

+49 17687136243

deb_4u22@yahoo.com

debjitpaul.github.io/debjit.github.io

Education

2014–2017 **MSc. Computer Science**, *Saarland University*, Saarbrücken, Germany, *Grade-1.9 on German Scale*.

Thesis Multitasking Learning With Unreliable Labels

Supervisor Prof.Dr.Dietrich Klakow, Head of Language Spoken System Group

Description Neural networks have shown impressive performance for various classification tasks due to the availability of large datasets. However, a known problem in classification tasks is the presence of unreliable labels. Successive manual annotations can improve the quality of the data set but are expensive to obtain. Hence, most applications rely on the artificial annotation (e.g., Brill tagger for POS tags) of large data. However, the labels generated by the artificial tagger are unreliable or noisy. We can further reduce this noise by also including supervision signals for related classification tasks in this setup. To this aim, we explore extending the single-task based noise-reduction setup to a multi-task case. Hence, we build a Multi-Task Learning-based (MTL) Noisy Label Neural Network. In this work, we establish Chunking as the primary task with POS tag classification as the auxiliary task. We set up the data by adding artificially annotated data with the human annotated Penn Treebank data. In our experiments, we compare the MTL-based setup with and without noise reduction against NLNN on the single task. We show that MTL-based NLNN outperforms the state-of-the-art for Chunking. **Github Page**

2010–2014 **B.Tech in Computer Science**, *GuruNanak Institute of Technology*, Kolkata, India, *Grade-8.73/10*.

Thesis Improved Algorithm for Human and non Human Object Detection

Experience

Jan.2018– **Data Scientist**, *Amplexor International*, Germany.

- Current
 - Working with Neural Network based Machine Translation (NMT) and Language models,
 - Handling Life Science dataset.

Oct.2016– **Research Assistant**, *Saarland University*, Saarland, Germany.

- Nov.2016
 - Department:** Foundation of Exact Algorithm
 - Programming tasks for the organization of the programming challenge **PACE**.
 - The goal is to generate instances to be used in the challenge.
 - Technology:** Python, Shell Scripting **Github Page**

Nov.2015– **Research Assistant**, *Saarland University*, Saarland, Germany.

- Feb.2016
 - Department:** Machine Learning Group
 - Crafted Algorithms using Python to filter data (Arxiv dataset) collection results.
 - Concept utilized are Natural Language Processing, Preprocessing text data.
 - Technology:** Python, Shell Scripting.

April 2015– **Student Research Assistant**, *Saarland University, DFKI*, Saarland, Germany.

- July 2015
 - Department:** Information and Services Systems
 - Developing features in OntoUML.
 - Designing Information Systems based on Conceptual Modeling.
 - Technology:** Java, Protege, OntoUML.

Selected Project

- Topic Machine-Learning-Tools-on-Iris-Dataset
- Description This tool is made to perform Machine Learning classification on Iris Dataset. Machine Learning classification tools are as follows: Support Vector Machine, Logistic Regression, Adaboost, Random forest. **Github Page**
- Topic Pre-Processing and NLP Tagger Tool
- Description Transforming a unstructured text data into CONLL format NLP tagging file. This tool can handle large unstructured data. Then can be used to tag words using artificial taggers such as Senna or Brill Tagger. **Github Page**
- Topic Bidirectional long short-term memory
- Description In recent times Bidirectional long short-term memory (BiLSTMs) networks has proven success for several NLP sequence task such as POS-tagging, NER tagging and Chunking. Implemented a BiLSTMs tagger with word and unicode byte embeddings.

Technical skills

- Language Python, R, Shell Scripting, Java, C, Matlab, MySQL
- Concepts Machine Learning , Statistical Learning, Data Mining
- Tools Pycharm, word2vec, Tensorflow, scikit-learn, Dynet, Stanford CoreNLP, Theano, Eclipse, Matlab, Numpy

Publication

- Title Anup Kumar Thander, Goutam Mandal, Debjit Paul "**Numerical Comparison of multi-step iterative methods for finding roots of non-linear equations, International Journal of Mathematics Trends and Technology "**", Volume4 Issue 8-September 2013 [ISSN: 2231-5373]

Achievements

- Winner of HQ Hackathon 2017, at Trivago

Languages

- English **Proficient**
- German **Basic**
- Bengali & Hindi **Native**

Hobbies

- Playing Football, Cricket and Tennis
- Photography

References

- Available upon request

Declaration

- I hereby declare that all the details furnished above are true to the best of my knowledge and belief