Rudra Murthy

Research Scientist IBM IRL Bangalore E-mail: rmurthyv@in.ibm.com rudramurthy.iisc@gmail.com Mobile: +91 966 385 4316 URL: http://murthyrudra.github.io

Research Interests

I am broadly interested in the application of Deep Learning to various Natural Language Processing (NLP) tasks. Recently, I have garnered interest in *multilingual assistance to overcome annotated data sparsity in low-resource languages.* Low-resource languages do not have sufficient annotated data, tools, and other resources to successfully train existing machine learning models for any NLP task. I am interested in exploring how language relatedness can be exploited to overcome lack of sufficient labelled data and improve performance of machine learning models for any task in low-resource languages.

Education

PhD - Indian Institute of Technology Bombay

· Guide: Prof. Pushpak Bhattacharyya

· CPI: 8.64 (10)

· Thesis: Multilingual Natural Language Processing in Low-Resource Setting

• **Summary:** My PhD work focuses on borrowing labelled data from one or more languages (known as *multilin-gual learning*) to overcome the issue of annotated data sparsity in low-resource languages. Data sparsity in NLP usually occurs due to a large number of words fall in the long tail of the *Zipfian* curve. For languages with less labelled data, the problem of data sparsity is amplified. Related languages share vocabulary and have similar word order. By borrowing labelled data from related languages we are not only just boosting the frequency of common words but also exposing the models to more patterns enabling the model to perform better. We apply the above intuition to Named Entity Recognition (NER) task. The results validates the above intuition and show that borrowing labelled data from related languages minimizing the language divergence leads to improved performance from multilingual assistance. We establish this on for the task of Machine Translation in an extremely low-resource setting.

M.E. - Indian Institute of Science Bangalore

· Guide: Prof. Shirish K Shevade

· CPI: 5.9 (8)

· Thesis: Learning from Positive and Unlabeled Examples

• **Summary:** In many real life cases, it is easy to collect positive examples. It is difficult to define a negative set, but, instead one can collect large unlabeled data. This unlabeled data contains a mixture of both positive and negative examples. Traditional classifiers cannot be directly used in this setting and requires modification. We explored use of Pairwise Ranking based Logistic Regression model to the problem. The motivation for using logistic regression is to get a confidence score from the system. We hoped that this confidence score can be used for better judgement of the class label. We obtained mixed results by beating the baseline on some datasets and performing closer on some datasets.

B.E. - RNS Institute of Technology, Bangalore

2007 - 2011

· Percentage Marks: 81.08%

2013 - 2020

2011 - 2013

Microsoft India (R&D) Pvt. Ltd.

· Mentor: Niranjan Nayak

• Title: Improving Indic Language to Indic Language Translation Quality

• **Summary:** In many cases, we do not have parallel corpus available between the source and the target languages to train a machine translation system. However, we do have parallel corpus between source - pivot languages and pivot - target languages. As part of my internship, I will be exploring how pivot language can be used to train a translation model from the source language to target language.

IBM IRL Bangalore

May 2016 to July 2016

· Mentor: Prof. Mitesh Khapra

• Title: Multilingual Models for Language Identification in Code-Mixed Data

• **Summary:** We study the performance of deep learning model for language identification task. We experiment with two training strategies. Training individual models for every language pairs involved. Training a single model (multilingual model) for all language pairs jointly. We observe multilingual models to be beneficial for improving the Named Entity tagging performance. However, this comes at a cost as we observed a drop in identification performance for some of the languages.

Publications

Journal

• Rudra Murthy, Mitesh Khapra and Pushpak Bhattacharyya, *Improving NER Tagging Performance in Low-Resource Languages via Multilingual Learning*, ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 2019.

Conference

- Rudra Murthy, Anoop Kunchukuttan and Pushpak Bhattacharyya, *Addressing word-order Divergence in Multilingual Neural Machine Translation for extremely Low Resource Languages*, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, USA, 2-7 June, 2019.
- Rudra Murthy, Anoop Kunchukuttan and Pushpak Bhattacharyya, *Judicious Selection of Training Data in Assisting Language for Multilingual Neural NER*, Association for Computational Linguistics (ACL), Melbourne, Australia, 15-20 July, 2018.
- Joe Cheri Ross, **Rudra Murthy**, Kaustuv Kanti Ganguli, Pushpak Bhattacharyya, *Identifying Raga Similarity in Hindustani Classical Music through Distributed Representation of Raga Names*, 13th International Symposium on Computer Music Multidisciplinary Research, 2017, **(ISCMMR)**, Matosinhos, Porto, 25-28 September, 2017.
- Rudra Murthy and Pushpak Bhattacharyya, *A Deep Learning Solution to Named Entity Recognition*, International Conference on Computational Linguistics and Intelligent Text Processing (CICLing), Konya, Turkey, 3-9 April, 2016.
- Sudha Bhingardive, Dhirendra Singh, **Rudra Murthy**, Pushpak Bhattacharyya, *Using Word Embeddings for Bilingual Unsupervised WSD*, Proceedings of the 12th International Conference on Natural Language Processing (ICON), Trivandrum, India, December, 2015.
- Sudha Bhingardive, Dhirendra Singh, Rudra Murthy, Hanumant Redkar, Pushpak Bhattacharyya, Unsupervised most frequent sense detection using word embeddings, Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Denver, Colorado, USA, 31 May - 5 June, 2015.

Symposium

• Jyotsana Khatri, **Rudra Murthy** and Pushpak Bhattacharyya, **A Study of Efficacy of Cross-lingual Word Embeddings for Indian Languages**, Proceedings of the ACM India Joint International Conference on Data Science and Management of Data (Young Researchers' Symposium) (**CODS-COMAD**), 2020.

pre-print

• Tamali Banerjee, **Rudra Murthy V**, Pushpak Bhattacharyya, Ordering Matters: Word Ordering Aware Unsupervised NMT, arXiv, 2019

Tutorials

- Introduction to Deep Learning: Sudha Bhingardive, Rudra Murthy, Kevin Patel, Prerana Singhal, and Pushpak Bhattacharyya, 12th International Conference on Natural Language Processing (ICON), Trivandrum, India, December, 2015.
- Applications of Deep Learning to NLP: Rudra Murthy, Kevin Patel, Mohammed Shad Akhtar, and Pushpak Bhattacharyya, 14th International Conference on Natural Language Processing (ICON), Kolata, India, December, 2017.

Academic Activities

Industry Events Participation

- Amazon Research Days Bangalore on 28th September 2018
- ACM-MSR Academic Research Summit 2018 at IIIT Hyderabad on 24th-25th January 2018

Academic Responsibilities

- Journal Reviewer: ACM Transactions on Asian and Low-Resource Language Information Processing (TAL-LIP), Journal by Sadhana - Academy Proceedings in Engineering Science,
- Conference Reviewer: LREC (2020), NAACL (2019), LREC (2018), COLING (2016)

Training Sessions

- Conducted a training session on Deep Learning at LGSoft Bangalore in June 2017
- Conducted a training session on NLP and Deep Learning at BNP Paribas Mumbai on 9th January 2015
- Conducted Lab Session on WSD related activities at IASNLP 2014, IIIT Hyderabad from 1st July 2014 to 15th July 2014

Tools Created

Source Code on Github: https://github.com/murthyrudra

Relevant Course Work

Data Mining, Linear Algebra, Natural Language Processing, Foundation of Machine Learning

Technical Skills

- Programming Language: C/C++, Java, Python.
- Platform: Linux.
- Deep Learning Framework: Torch, PyTorch
- Tools: LATEX, Beamer

References

- Prof. Pushpak Bhattacharyya
- Prof. Mitesh Khapra

Personal Details

E-mail: rmurthyv@in.ibm.com rudramurthy.iisc@gmail.com

I, hereby declare that all the information furnished above is true to the best of my knowledge.

Date: July 29, 2020 Place: Bengaluru

Rudra Murthy V.