Sandeep Subramanian
5540 Fifth Avenue, Apt 14
Pittsburgh PA, 15232

412-378-8510
sandeeps@andrew.cmu.edu
www.cs.cmu.edu/~ssandeep
GitHub: https://github.com/MaximumEntropy

Education

Carnegie Mellon University / M.S Language Technologies (3.62/4.0)
AUGUST 2014 - PRESENT, Pittsburgh, PA

Relevant Coursework

- Algorithms for Natural Language Processing (11-711)
- Machine Learning (10-701)
- Machine Learning for Text Mining (11-741)
- Large Scale Multimedia Analysis (11-775)
- Independent Study - Computational Biology (11-925)
- Deep Learning (11-785)
- Multimodal Machine Learning (11-776)

VIT University / B.Tech Computer Science and Engineering (8.75/10)
JUNE 2010 - MAY 2014, Vellore, Tamil Nadu, India

Relevant Coursework

- Artificial Intelligence
- Theory of Computation
- Algorithm Design and Analysis
- Data Structures and Algorithms
- Graph Theory
- Applied Probability, Statistics and Reliability
- Linear Algebra

Experience

Swych.it / Co-founder (http://www.swych.it)
JANUARY 2015 - PRESENT, Pittsburgh, PA

Co-founded Swych.it, a language learning application that uses code-switching to teach languages via context.

Contribution

- Developed the entire web-stack and REST APIs
- Contributed to the downstream use of a machine translation system to generate code-switched content

Carnegie Mellon University / Graduate Research Assistant
AUGUST 2014 - PRESENT, Pittsburgh, PA

Awarded a Graduate Research Fellowship that includes full tuition and stipend

Advisor: Dr. Madhavi Ganapathiraju

Language Technologies Institute, School of Computer Science, Carnegie Mellon University Department of Biomedical Informatics, School of Medicine, University of Pittsburgh

Research: Computational Biology, Machine Learning & Natural Language Processing

Invention Labs / Intern (http://www.avazapp.com/)
DECEMBER 2014 - JULY 2014, IIT Madras Research Park, Chennai, Tamil Nadu, India

Mentor: Ajit Narayanan, CEO Invention Labs

Contributed to the development of AvazFreespeech, an Augmentative and Assistive Communication (AAC) application to help kids with autism communicate and learn english. The application converts a semantic graph constructed by kids into syntax using X-bar theory.

Indian Institute of Technology, Madras / Research Intern (NLP)
MAY 2012 - JUNE 2012, Chennai, Tamil Nadu, India

Advisor: Sutanu Chakraborti, Associate Professor, Indian Institute of Technology, Madras

Worked on using information theoretic approaches to empirically prove that sentiment classifiers cannot handle certain linguistic constructs in text such as double negations and zero anaphora.

Books, Publications & Patents

BOOKS :

Smart Cyclone Alerts Over the Indian Subcontinent : New Insights Into Storm Mitigation / *LAP Lambert Academic Publishing*
Satyajit Ghosh, Sandeep Subramanian and Vivek Vidyasagaran
APRIL 2016 (Expected)
Co-authored a book that discusses the following in great detail

- A framework to process large-scale geospatial data
- Using geospatial data with existing open source climate models for weather and extreme event prediction
- Downstream use of weather data for societal good - warning systems for cyclones in India.

PAPERS :

Neural Architectures for Named Entity Recognition (Under Review NAACL 2016)
Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami and Chris Dyer
JANUARY 2016
Named Entity Recognition using character and token-level bidirectional LSTMs with a CRF and Stack-LSTMs .

Towards Extracting Supporting Information for Predicted Protein-Protein Interactions / *IEEE/ACM Transactions on Computational Biology and Bioinformatics*
Adam Roth, Sandeep Subramanian and Madhavi Ganapathiraju
SEPTEMBER 2015

Literature based similarity scores for protein pairs to aid in the interpretation of computationally predicted protein-protein interactions.

Smart Cyclone Alerts over the Indian Subcontinent / *Atmospheric Science Letters*
Satyajit Ghosh, Vivek Vidyasagaran and Sandeep Subramanian
FEBRUARY 2014

Application of weather data from open source climate models for cyclone alerts in rural and coastal India.

Location Specific Tag Weighting for Language Model Based Placing of Images / *MediaEval 2013 Placing Task Workshop*
Sandeep Subramanian, Vivek Vidyasagaran and Krishna Chandramouli
OCTOBER 2013

Estimation of the geographical coordinates (latitude, longitude) of images on Flickr.

Visualizing World University Rankings : A Novel Algorithm / *European Conference on Education (IAFOR)*

Sandeep Subramanian, Vivek Vidyasagaran, Kannabiran Maheswaran and Satyajit Ghosh
JUNE 2013

A data visualization library written in Python and Pygame that provides dynamic and interactive visualizations of data with an application to university ranking data provided by U-Multirank .

PATENTS :

Code-switching as a tool for Language Acquisition / *US Patent (Pending)*
Guillaume Lample, Sandeep Subramanian and Manu Reddy
DECEMBER 2015

A Novel Computational Framework for Extreme Weather Alerts/ *Indian Patent (Pending)*
Sandeep Subramanian, Vivek Vidyasagaran and Satyajit Ghosh
APRIL 2014

Awards

Graduate Research Fellowship / *Carnegie Mellon University*
AUGUST 2014 - PRESENT

- Covers entire tuition and stipend

Special Achievers Award (2014) / *VIT University*

- Awarded for international recognition and media coverage of Bachelor's thesis.

Select Projects

SanDeepLearn - A Deep Learning Library (*https://github.com/MaximumEntropy/SanDeepLearn)*
NOVEMBER 2015 - PRESENT

- A high-level deep learning library built on top of theano that has implementations for MLPs, CNNs, RNNs and LSTMs.
- Includes implementations of related optimization techniques like Adagrad and RMSprop.

Transmembrane Helix Prediction using Recurrent Neural Networks / *Graduate Research Work*
DECEMBER 2015 - PRESENT

- Predicted transmembrane amino acid sequences within a protein sequence using bidirectional RNNs and LSTMs without encoding any prior biological information

Multimodal Sentiment Analysis & Quantifying Negative Correlation Between Modalities in Expressing Sentiment / *Multimodal Machine Learning Class Project*
AUGUST 2015 - PRESENT

- Multimodal Fusion of the audio, visual and textual modalities with Multi-modal Autoencoders
- Recurrent Neural Networks for sentiment classification

Protein-Protein Interaction Prediction by Retrofitting Word Embeddings to Gene Ontology / *Graduate Research Work*
MAY 2015 - PRESENT

- Trained word-embeddings for proteins on PubMed articles and used retrofitted them to Gene Ontology to improve the quality of the vectors
- Obtained statistically significant p-values when comparing similarities of interacting versus random proteins

Violent Scenes Detection in Hollywood Movies / *Large Scale Multimedia Analysis Class Project*
JANUARY 2015 - APRIL 2015

- Multimodal late fusion of MFCCs, Dense Trajectories, SIFT and ConvNet features.
- Exploited the consistent nature of violent scenes with respect to the duration of the movie to achieve a ~10% increase in MAP score.

Skills: Python, C/C++, Numpy/Scipy, Theano, Django, HTML, CSS, Javascript, JQuery, REST