TANVI SAHAY

Brittany Manor Drive, Amherst, MA - 01002
tsahay@cs.umass.edu | | | webpage

Education

May 2018	**University of Massachusetts Amherst, MA**, Master of Science, Computer Science	4.0
June 2016	**Birla Institute of Technology, India**, Bachelor of Engineering	3.0

Skills

Python, Tensorflow, Javascript, D3, Numpy, scikit-learn, nltk, Stanford CoreNLP, SQL, Django, MongoDB, Git, LaTeX

Work Experience

Sep 2017 -Present **Graduate Research | Information Extraction and Synthesis Lab | UMass Amherst**

- Explore machine learning techniques for obtaining interpretable models of researcher expertise to match new scientific works with potential reviewers.
- Explore deep learning models for the purpose of key word/phrase extraction from research papers, to assist in better expertise modeling.

Feb 2017 -Aug 2017 **R&D Intern and Independent Study | Lexalytics Inc.**

- Successfully implemented several NLP baseline and word2vec and deep learning based models to obtain fixed-dimensional distributed representations of phrases for the purpose of clustering semantically-related phrases
- Experimented with KMeans, DBSCAN, Heirarchical and Spectral clustering and several cluster evaluation techniques for obtaining phrase clusters coherent to human evaluators.

Selected Research Projects

Nov 2017 -Dec 2017 **Analysis of RL algorithms for bigram language model MDP**

- Formulated the bigram language model as an environment and evaluated performance of SARSA, Q Learning and Q(λ) on this MDP
- Experimented with four different reward functions and evaluated each experimental model using average expected return of rewards and average probability of sentences generated by the trained models

Sep 2017 -Dec 2017 **Automatic Colorization of Videos**

- Compared the performance of existing image colorization architectures for automatically colorizing video frames using pixel wise rmse and colorization quality measured by human evaluators
- Experimented with an RCNN architecture to account for consistency between consecutive frames and used a re-weighted class-rebalancing loss to avoid desaturated colorization of grayscale images

Feb 2017 -May 2017 **Schema Matching using Machine Learning**

- Engineered custom features to represent schema names and employed Self Organizing Maps and Gaussian Clustering to cluster similar schema names
- Performed within-cluster one-to-one matching using edit distance and introduced the idea of domain-based global dictionary for the purpose of one-to-many schema matching

Oct 2016 -Dec 2016 **Sentence Generation using Fan Theories**

- Employed CoreNLP package for tokenization, relation extraction, PoS tagging and Named Entity Recognition on a database of Fan Theories of Game of Thrones
- Performed noise removal using OpenIE and used Bigram language model, Hidden Markov Model and character-level LSTM for sentence generation.
- Analyzed results based on overall coherence, general fluency and information content with and without considering domain knowledge.

Selected Publications

- A. Aggarwal, T. Sahay, A. Bansal and M. Chandra, "Grid search analysis of nu-SVC for text-dependent speaker-identification," 2015 Annual IEEE India Conference (INDICON), New Delhi, 2015. *Best Paper Award*
- T. Sahay, A. Aggarwal, A. Bansal and M. Chandra, "SVM and ANN: A comparative evaluation," 2015 International Conference on Next Generation Computing Technologies (NGCT), Dehradun, 2015.

Additional Experience

- Grader for the graduate level CS 589 Machine Learning course for Spring 2017, Fall 2017 and Spring 2018
- Volunteered for the official UMass Hackathon HackUMass in Fall 2016.