PRASHA SHRESTHA

shr.prasha@gmail.com ⋄ (205) · 948 · 7087 ⋄ https://prashas.github.io/

EDUCATION

Ph.D. in Computer Science, University of Houston　　　　　　　　Jan 2015 – May 2018
Advisor: Dr. Thamar Solorio, GPA: 4.00
Dissertation Title: Authorship Attribution in Realistic Scenarios

MS in Computer Science, University of Alabama at Birmingham　　　Aug 2012 – Dec 2014
Advisor: Dr. Thamar Solorio, GPA: 4.00

BE Computer Engineering, Tribhuvan University Institute of Engineering　Nov 2005 – May 2010
Final Year Project: Nepali Text to Speech

EXPERIENCE

Post Doctoral Research Associate, Pacific Northwest National Laboratory　　Jun 2018 –Present

- **Virality Forecasting with Graphs**
  A node-aware attention model to predict the virality of repositories on GitHub, tweets on Twitter (multitasking to forecast retweets and mentions together), and videos on YouTube using dynamic network graphs with R^2 >0.7.
- **Information Spread in Online Social Media**
  Analysis of information spread about software vulnerabilities and cybersecurity threats on online social media and the possible factors that drive it.
- **Extensive Evaluation of NLP models**
  Intrinsic and extrinsic evaluation of state-of-the-art named entity recognition, semantic role labeling, and machine comprehension models. Result reproduction and evaluation of generalizability across datasets, error analysis as well as testing performance in downstream tasks.

Research Intern, Pacific Northwest National Laboratory　　　　　　Feb 2018 – May 2018
Mentor: Dr. Svitlana Volkova

- **Anticipatory Analysis with Deep Learning for Graphs**
  Graph embeddings using node2vec and Deepwalk trained on daily network graphs to predict the instability of a country with GDELT data and to predict email exchange volume with ENRON data.

Research Assistant, University of Houston (RiTUAL Lab)　　　　　　Jan 2015 – May 2018

- **Cross-domain representation learning**
  Using domain-colored tokens to learn embeddings for two vastly different domains (Guardian news and tweets) in the same space, for ∼20K documents from 10 authors with 33% accuracy on document level, an improvement of 20% over a bag-of-words baseline.
- **Authorship attribution of short texts**
  Multi-channel CNN model with character n-gram and word sequence inputs for authorship attribution of tweets from 1K users that achieved an accuracy of 38% on tweet level, an improvement of >11% over a character-only baseline.
- **Large scale authorship attribution and verification**
  Authorship attribution via individual verifiers using various lexical and syntactic features for authorship attribution of 1K Amazon reviewers and 500 Yelp reviewers with ∼80% accuracy on a user level.
- **Author profiling on medical forum posts**
  Combined age and gender prediction of ∼84K users from DailyStrength medical forum users using lexical and familial token based features obtaining 88% and 65% accuracy for age and gender prediction respectively.

Research Assistant, University of Alabama at Birmingham (CoRAL Lab)　Aug 2012 – Dec 2014

- **Text alignment in plagiarism detection**
  Using stopword and named entity n-grams matching for alignment of plagiarized passages. Also used TER-p as a similarity metric for text alignment.
- **Plagiarist identification**
  Identification of the original document by measuring similarity between the plagiarized and non-plagiarized parts of two documents.
- **Author profiling on social media posts**
  Efficient implementation of Naive Bayes with tf-idf using MapReduce for author's age and gender identification of English and Spanish documents.
- **Malware family identification**
  Prototype based classification using prominent string sets of ten malware families.

Software Engineer, Bajra Technologies Pvt. Ltd.   Aug 2011 – May 2012
- Web service for mobile applications
- Website development

Software Engineer, Deerwalk Services Pvt. Ltd.   Mar 2010 – Jul 2011
- Accounting and inventory management for a hospital management system
- MIS for Carolina Medical Homes

TECHNICAL STRENGTHS

Programming Languages	Python, Java, C++
Tools and Libraries	Keras, scikit-learn, nltk, networkx, pandas, matplotlib
Databases	MongoDB, MySQL, SQLite
Cloud Services	Amazon EMR, EC2, S3

AWARDS

First position in University of Houston CS PhD showcase 2017
Recipient of Cullen Travel Grant 2017
Studentship to attend the 2017 European Chapter of the ACL Conference
Third position in the PAN Author Profiling 2014 Shared Task
GHC Scholarship Grant to attend the 2014 Grace Hopper Celebration of Women in Computing
Travel grant to attend the 2014 CRA-Women (CRA-W) Conference
Freeship and fellowships in 6 out of 8 semesters at Tribhuvan University

TRAINING AND CERTIFICATIONS

MILA Deep Learning and Reinforcement Learning Summer School	*Jun 2017*
Lisbon Machine Learning School *(LxMLS 2016)*	*Jul 2016*
Deep Learning Workshop at *MindLab*, *Universidad Nacional de Colombia*	*Jan 2016*
Machine Learning, *www.coursera.org*	*May 2014*

PUBLICATIONS

1. **Prasha Shresha**, Suraj Maharjan, Dustin Arendt and Svitlana Volkova. Forecasting User Behavior from Dynamic Social Interaction Graphs: A Case Study of Twitter, YouTube, and GitHub. 15th International Workshop on Mining and Learning with Graphs (MLG 2019).

2. Suraj Maharjan, Deepthi Mave, **Prasha Shrestha**, Manuel Montes, Fabio A. González and Thamar Solorio. Jointly Learning Author and Annotated Character N-gram Embeddings: A Case Study in Literary Text. In Proceedings of the 2019 Conference on Recent Advances in Natural Language Processing (RANLP 2019) [To Appear]).

3. Suraj Maharjan, **Prasha Shrestha**, Katherine Porterfield, Dustin Arendt and Svitlana Volkova. Towards Anticipatory Analytics: Forecasting Instability Across Countries from Dynamic Knowledge Graphs. 5th Pacific Northwest Regional NLP Workshop (NW-NLP 2018).

4. **Prasha Shrestha**, Sebastian Sierra, Fabio A. González, Paolo Rosso, Manuel Montes-y-Gómez, and Thamar Solorio. Convolutional Neural Networks for Authorship Attribution of Short Texts. 15th edition of the European Chapter of the Association for Computational Linguistics (EACL 2017).

5. **Prasha Shrestha**, Steven Bethard, Ted Pedersen, Nicolas Rey-Villamizar, Farig Sadeque, and Thamar Solorio. Age and Gender Prediction on Health Forum Data. 10th edition of the Language Resources and Evaluation Conference (LREC 2016).

6. **Prasha Shrestha**, Arjun Mukherjee and Thamar Solorio. Large Scale Authorship Attribution of Online Reviews. 17th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2016).

7. **Prasha Shrestha** and Thamar Solorio. Identification of the Original Document by Using Textual Similarities. 16th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2015).

8. **Prasha Shrestha**, Suraj Maharjan, Gabriela Ramírez, Alan Sprague, Thamar Solorio, and Gary Warner. Using String Information for Malware Family Identification. 14th Ibero-American Conference on AI (IBERAMIA 2014).

9. Suraj Maharjan, **Prasha Shrestha**, Thamar Solorio, and Ragib Hasan. A Straightforward Author Profiling Approach in MapReduce. 14th Ibero-American Conference on AI (IBERAMIA 2014).

10. **Prasha Shrestha**, Suraj Maharjan, and Thamar Solorio. Machine Translation Evaluation Metric for Text Alignment. Notebook for PAN at CLEF 2014.

11. Suraj Maharjan, **Prasha Shrestha**, and Thamar Solorio. A Simple Approach to Author Profiling in MapReduce. Notebook for PAN at CLEF 2014.

12. **Prasha Shrestha** and Thamar Solorio. Using a Variety of n-Grams for the Detection of Different Kinds of Plagiarism. Notebook for PAN at CLEF 2013.