SAUMYA YASHMOHINI SAHAI

sahai.17@osu.edu +1(607)379-7461

RESEARCH INTERESTS

Broadly interested in applied machine learning, particularly in building and explaining deep neural models with applications in natural language processing/understanding, computational linguistics, social computing and computational brain & behavior.

EDUCATION

The Ohio State University, *Columbus, OH, USA* — *Jun 2019 - Present*
Computer Science and Engineering — *CGPA : 3.8/4.0*

Cornell University, *Ithaca, NY, USA* — *Aug 2016 - May 2019*
Neurobiology and Behavior (minor in Data Science) — *CGPA : 3.9/4.0*

Birla Institute of Technology and Science, *Pilani, India* — *Jul 2011 - Jul 2016*
Computer Science and Physics — *CGPA : 9.2/10.0*

TECHNICAL SKILLS

Proficient	Python, PyTorch, Keras, Scikit-learn, Linux, Git, Amazon Mechanical Turk
Experienced	Java, SQL, R, Matlab, C++, Javascript

PUBLICATIONS

Sahai, Saumya Yashmohini, Oana Balalau, Roxana Horincar *Breaking Down the Invisible Wall of Informal Fallacies in Online Discussions*. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, *to appear* (**ACL 2021**)

Sahai, Saumya Yashmohini*, Dravyansh Sharma* *Predicting and Explaining French Grammatical Gender.* The ACL Special Interest Group on Typology (**NAACL 2021**) (∗ denotes equal contribution)

Sahai, Saumya Yashmohini, Saket Gurukar, Wasiur R. KhudaBukhsh, Srinivasan Parthasarathy, and Grzegorz A. Rempala *A Machine Learning Model for Nowcasting Epidemic Incidence.* arXiv preprint arXiv:2104.02174 (2021), *under review* (**Journal of Mathematical Biosciences**)

Bandyopadhyay, Bortik, Pranav Maneriker, Vedang Patel, **Saumya Yashmohini Sahai**, Ping Zhang, and Srinivasan Parthasarathy *DrugDBEmbed: Semantic Queries on Relational Database using Supervised Column Encodings.* arXiv preprint arXiv:2007.02384 (2020).

Whitehead, Samuel, Troy Shirangi, Theodore Lindsay, **Saumya Yashmohini Sahai**, Erica Ehrhardt, Tsevi Beatus, Nilay Yapici, Michael Dickinson, David Stern, and Itai Cohen *Uncovering the Neural Basis of Flight Control in Fruit Flies.* March Meeting 2018, **American Physical Society**

Snchez-Alcaiz, Juan Antonio, Ana Florencia Silbering, Vincent Croset, Giovanna Zappia, Anantha Krishna Sivasubramaniam, Liliane Abuin, **Saumya Yashmohini Sahai** et al. *An expression atlas of variant ionotropic glutamate receptors identifies a molecular basis of carbonation sensing.* **Nature communications** 9, no. 1 (2018): 1-14.

INDUSTRIAL EXPERIENCE

Machine Learning Engineering Intern *May 2021 - Aug 2021*
Spotify Inc., USA
Mentor: Yakov Kronrod

Working on spoken language understanding with the Voice team.

Product Development Intern *Jan 2016 - Jun 2016*
Adobe Systems, Noida, India
Mentor: Gaurav Bhargava

Developed a plug-in implementing the two-way and three-way difference and merge algorithm, for preserving formatting and design layout of the different classes of text and tables, being reimported into Adobe InDesign® files.

RESEARCH EXPERIENCE

Research Intern in NLP *May 2020 - Aug 2020*
CEDAR Team, Inria, Palaiseau, France
Advisor: Oana Balalau

- Created taxonomy for informal fallacies using classical theories of argumentation and a domain independent dataset for detection of fallacies in discussions on Reddit.
- Developed state-of-the-art deep learning models for detection of fallacies.

Graduate Research Associate *Jun 2019 - Apr 2021*
Data Mining Research Lab, The Ohio State University, Columbus, OH
Advisor: Srinivasan Parthasarathy

- Generated informative representation for raw drug description with the goal of classifying the type of interaction it has with another drug, using LSTM based-neural network architectures in PyTorch framework.
- Developed a random forest model for nowcasting the daily infection count for COVID-19, which outperformed the current state-of-the-art hierarchical Bayesian model.
- Working towards quantifying distances between opinions using state-of-the-art models for natural language understanding.

Graduate Research Assistant *Jun 2017 - May 2019*
Department of Neurobiology and Behavior, Cornell University, Ithaca, NY
Advisor: Nilay Yapici
Modeled the optimal strategies underlying foraging decisions in *Drosophila melanogaster* under varying conditions of internal states and external stimuli using reinforcement learning models. Wrote end-to-end pipeline for data analysis in R and Python.

Rotating Graduate Student *Jan 2017 - May 2017*
Department of Physics, Cornell University, Ithaca, NY
Advisor: Itai Cohen
Investigated the neuronal basis of flight stabilization under perturbation in *Drosophila melanogaster*. Carried out experiments and analysed data under a control theory-inspired model. Also revised the 3-D *Drosophila* hull generation code for better wing reconstruction.

Undergraduate Researcher *Jan 2015 - May 2015*
Department of Computer Science and Information Systems, BITS Pilani, India
Advisor: Yashvardhan Sharma
Restructured the existing JAVA based text summarizer to work for non-English languages. Created corpora of reviews in non-English languages and predicted their sentiments.

Winter Research Scholar *May 2015 - Jul 2015*
Queensland Brain Institute, The University of Queensland, Brisbane, Australia
Advisor: Geoffrey Goodhill
Carried out the project 'Mathematical Analysis of Neural Coding in Zebrafish Tectum'. Worked towards the development of mathematical models to explain spontaneous activity taking place in the tectum of zebrafish in the absence of any visual stimuli.

Summer Research Fellow *May 2014 - Jul 2014*
Indian Institute of Science, Bangalore, India
Advisor: Vijay B. Shenoy
Worked on the project 'Will a Superfluid Deplete at Zero Temperature in Presence of Interactions?'. The simulations were performed in Mathematica.

Summer Research Intern *May 2013 - Jul 2013*
National Centre for Antarctic and Ocean Research, Goa, India
Advisor: Alvarinho J. Luis
Carried out the project 'A polar geospatial information extraction system using spectral characteristics of high resolution optical remote sensing data and bathymetric extraction' using MATLAB.

OTHER PROJECTS

Stance-based summarization of online debates *Aug 2020 - Dec 2020*
Generate stance-based extractive summaries of debates, by encoding discourse structure in addition to the content.

Discourse act classification of online forum conversations *Aug 2020 - Dec 2020*
Classify discourse acts of written conversational text by using BERT based models.

HONOURS AND AWARDS

Global Gateway Graduate Student Research Abroad Grant, The Ohio State University	2020
University Fellowship, The Graduate School, The Ohio State University	2019 - 2020
Mong Junior Fellow, Cornell Neurotech, Cornell University	2018 - 2019
Cornell Fellowship, Department of Neurobiology and Behavior, Cornell University	2016 - 2017
Best Academic Performer for the Batch of 2011, Physics Society, BITS Pilani	2015
Summer Research Fellowship, Indian Academy of Sciences	2014
INSPIRE Fellowship, Department of Science and Technology, Govt. of India	2011-2016

LANGUAGES

English:	Full professional proficiency
Hindi:	Native language

COURSEWORK

Speech and Language Processing
Social Media and Text Analytics
Bayesian Data Analysis
Algorithms
Cutting-Edge Topics in NLP
Ethics and fairness in AI
Machine learning and Data Mining
Computer Architecture

TEACHING EXPERIENCE

Graduate Teaching Assistant *Jan 2018 - May 2018*

Introduction to Neuroscience, Cornell University, Ithaca, NY

Duties included leading discussion sections, coordinating lecture sessions, designing preliminary and final examinations and grading assignments and exams for this introductory neuroscience course for sophomores.