Van Nguyen Hung Hoang

Computer Science and Linguistics
University of Arizona
Tucson, AZ 85719

Webpage: https://vanh17.github.io/
Email: vnhh@email.arizona.edu
Phone: (812) 292 - 9073

Education

05/2022 Ph.D. Computer Science, University of Arizona
Minor: Linguistics and Informatics
Research Interests: Machine Learning, Artificial Intelligence, and their direct applications to QA, NLP, Healthcare, and Medical document understanding and simplification.
Cumulative Grade Point Average: 4.0/4.0
Advisers: Mihai Surdeanu, Steven Bethard

05/2017 B.S. Computer Science and Mathematics, Hanover College, IN, USA
Minor: Physics and Economics
Summa Cum Laude, Rank: 1/1000, GPA: 4.0/4.0.

Highlights of Skills

Programming Languages	Python (proficient), C++ (proficient), Java(proficient), Scala
NN Libraries	Pytorch, Keras, Tensorflow
Research Skills	Critical thinking and analysis; strong algorithms - complexity background.

Research Experience

05 – 09/2019 **Research Scientist Intern,** Oak Ridge National Laboratory
Implemented Bert and XLNet pretraining pipelines for suicide risk detection from 10 billions electrical health records collected by U.S Department of Veteran Affairs.

Implemented unsupervised NER techniques to improve suicide risk detection on public MIMIC-III electronic health record dataset.

2019 – 2020 **Research Assistant, Text Simplification,** University of Arizona
Implemented XLNet and Bert pipelines for text simplification.

Invented language model auto-completion pipelines for simplifying text in medical documents.

Our model outperforms baseline by 30%.

2018 – 2019 **Research Assistant, Twitter for Social,** University of Arizona
Built a classification system to predict public health information such as diabetes and obesity rate at the state-level in the United States. The system was trained on a 5-year long corpora of social media tweets.

Our best performance was 80.39% on the diabetes prediction task at the state-level in the United States

05 – 08/2019 **Research Assistant, Fake Science,** University of Arizona
Implemented XLNet pipelines for detecting fake news. The pipelines outperforms current BERT-related approaches by 5% on the internal annotated science dataset.

Experienced with domain adaptation techniques, which results in another 2% improvement on the performance.

06 – 09/2016 **Deep Learning Research Assistant,** University of Kentucky
Developed a convolutional neural network with Python to recognize written digits.

Built the program to automatically extract 1800 images for 3D visualization model database, reducing the data collecting time from a month to 2 weeks.

05 – 12/2017 **Lead Software Engineer,** Environmental Laboratory Inc.
Created an online payment system that allows clients to pay their invoices online, improving accounting department efficiency by 20%.

Integrated 6 testing instruments with the company database and website, reducing data entering time by 25% and amount of mistakes by 50%.

Implemented automatic transferring system for invoices and payments, improving efficiency in invoicing by 50%.

08 – 12/2018 **Teaching Assistant,** University of Arizona
Held office hours, discussion sessions, and exam review sessions for students.

Assisted professor in teaching computer labs and grading assignments.

Honors, Awards, and Rankings

Recipient of the Computer Science Department Graduate Fellowship Grant, UArizona, 2018
Recipient of Distinct Department Award for Best Computer Science Graduate, HC, 2017
Petticrew Scholarship for Best Computer Science Major, HC, 2015-2017
Distinct Department Award for Best Mathematics Graduate, HC, 2017
Recipient of Distinct Department Award for Best Physics Sophomore, HC, 2014
Recipient of Richer Grant for Self-developed Researcher, HC, 2017

Services

Member of Computer Science Graduate Council, UArizona (2018- now)
Coordinator for NLP reading group, UArizona 2019-2020.

Currently working on

1. Deep neural network models for simplifying text in medical/healthcare documents.
2. Suicide risk and health-related problem detection from the language of electrical health records and social media.
3. Knowledge aggregation in QA, focusing on coherence and human interpretability for robust and scalable QA systems.
4. Graph Neural Networks and their applications to NLP

Publications

1. Hoang Van, Ahmad Musa, Hang Chen, Stephen Kobourov, and Mihai Surdeanu. "What does the language of foods say about us?." *In Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis* (LOUHI 2019).
2. Hoang Van, Ahmad Musa, Mihai Surdeanu, and Stephen Kobourov. "On indirect effects of Covid-19" *In Preparation.*
3. Hoang Van, Gondy Leroy, and David Kauchak. "Autocomplete for Medical Text Simplification." (In Preparation).
4. Hoang Van and Steven Bethard. "Predicting affective content in tweets with Pretrained Language Models" (In Preparation)
5. Hoang Van, Dongfang Xu, Kris Brown, and Edmon Begoli. "Sucide risk detection in electrical health records" (In Preparation)