HOANG N.H. VAN

University of Arizona
Phone: (812) 292-9073
Email: vnhh@email.arizona.edu

https://vanh17.github.io
https://github.com/vanh17
linkedin.com/in/hoangnhvan

Education:

Earning **PhD Degree**, Computer Science Department, University of Arizona (AZ). **Research interests:** natural language processing, information extraction, machine learning application for health informatics, and question answering system.
Major: Computer Science, Minors: Linguistics, Informatics CGPA: **4.0/4.0**
Earned **Bachelor of Arts,** Hanover College (IN), May 2017 **Summa Cum Laude**
Majors: Computer Science, Math; Minor: Economics, Physics.
Graduated with Honors in CS and Math. Rank: **1/1000**. CGPA: **4.0/4.0. Dean's List: All Terms**

Related Experience:

Research Assistant, Scalable and Protected Data Group, Oak Ridge National Laboratory (May - Aug 2019)

- Created clean labeled dataset from MIMIC electronical health records with pre-and-post suicidal events with SQL commands.
- Created high suicidal risk vocabulary for unsupervised learning models with TF-IDF and PMI scores.
- Implemented BERT and XLNet models in tensorflow for supervising learning approaches to the newly created gold dataset.
- Created automatic systems to pretrain the state-of-the-art models: XLNet, BERT, ELMo, Glo2Vec on medical health records and other medical datasets, which yields potential improvement on using these models on medical natural language processing tasks.

Graduate Research Assistant, University of Arizona (June 2019 - present)

- **Fake Science Classifier:** implemented and analyzed the performance of state-of-the-art neural networks: XLNet, BERT, and ELMo on climate news articles using tensorflow, and pytorch libraries on UNIX environment. Preliminary result with XLNet is 92.8% accuracy on gold climate data.
- **Medical Text Simplifier:** implemented a text simplification engine and website API with Python to automatically reduced the difficulty of medical text to normal text. The implementation includes both machine learning approaches (BERT, XLNet and other language models) and rule-based approaches.

Graduate Teaching Assistant, Computer Science Department, University of Arizona (Aug - Dec 2018)

- Held office hours, discussion, and midterm review session for students for Data Structure and Algorithm.
- Wrote automatic grading scripts/ testing cases and programming assignment in C++ with Shell scripts.

Research Assistant, University of Kentucky Visualization & Virtual Environments Center (Jun - Sep 2016)
Advisor: Professor. J. Nathan

- Developed a convolutional neural network with Python and Tensorflow to recognize written digits.
- Built from scratch 3 demo-web interfaces for the image recognition models using Python, Flask, and JS.
- Built the automatic program in Python to extract 1800 images for 3D visualization model database. This helped reduce the amount of time spent on collecting data from a month to 2 weeks.

Software Engineer Intern, Environmental Laboratories Inc. (May 2017 – Jan 2018)

- Created an online paying system in C# that allows clients to pay their invoices online. This helped an accounting team increase their invoicing efficiency by 20%.
- Created interface for 6 testing instruments using C# and Visual Basics (automatically processing data and updating results from testing instruments to the company database and websites). This reduced data entering effort by 25% and amount of mistaken data by 50%.
- Implemented automatic transferring system for invoices and payments and created a direct interface between QuickBooks and company financial software. This work helped reduce the amount of time spent on applying payment and reporting company financial status by 50%.

Programming Projects:

Twitter4Social (2018)

- Built a classification system to predict public health information such as diabetes rate at the state-level in the United States. The system was trained on a 5-year long corpora of social media tweets. The model was implemented using Support Vector Machine and Random Forest.
- Analyzed errors on the model performance and the health-related indication of language of food from social media on predicting diabetes risk.
- Our best performance was 80.39% on the diabetes prediction task at the state-level in the United States.

Obedient Project (2016 – 2017)

- Individual Project. Built speech-driven robot using EasyVR shield as speech recognition component and Arduino BOE shield Bot as brain of the robot programmed in Arduino language.
- Obedient was able to recognize my commands from basic to advanced (left, right, turn with numeric degrees, forward, and backward with numeric distance in centimeter), work effectively in an unnoisy environment, and find a way out of the maze with user's spoken instruction.

WorkerScheduleManager on iOS and Android (2015 – 2016)

- Individual project. The app was used by around 50 students and faculties.
- WorkerScheduleManager was an application for student workers at Horner Center, the recreational center at Hanover College, to keep track, trade working shifts, and manage their working hours. It reduced the number of missed shifts by 50-60%, improved the shift exchanging process, and also created a way of tracking the shifts worked at Horner for managerial purposes to the administration staffs which had never existed before.

Activities and Awards:

Distinct Department Award for Best Physics Major (2014 – 2015)
Petticrew Scholarship for Best Computer Science Major (2016-2017)
Distinct Department Awards for Best Computer Science and Mathematics Major (2014, 2015, 2016, 2017)

Richer Grant for Self-developed Researcher, Hanover College (April 2017)

- Topic: Artificial-Intelligence Startup Ecosystem and The Gender Bias Issue in Startup Communities in Asia. Countries: S. Korea, Vietnam, Malaysia, Indonesia
- Met with 20 organizations and attended 5 Startup conferences in the region. Gave presentation at Hanover College and documented all the collected information at Duggan Library, Hanover College.

Member of Computer Science Graduate Council at University of Arizona (August 2018- now)

University of Arizona Computer Science Fellowship for First Year PhD Student (August 2018)

Highlights of Skills:

- Skilled in critical thinking and analysis; possess strong programming, mathematics, and statistics background.
- Programming skills: C++ (proficient), JavaScript(proficient), C# .Net (plenty experience), Java (plenty
- experience), PHP (plenty experience), Ocaml, Visual Basic, Python.
- Skilled in MySQL (proficient), Microsoft SQL server (plenty experience), MongoDB, NoSQL.
- Skilled in Rails, AngularJS, MEAN.JS.
- Skilled in MATLAB. Familiar with Unix/Linux, Mac OS, Windows as well as their command line
- Skilled in HTML and CSS. Have strong algorithms - complexity, problem solving background.

Publications:

Hoang Van, Ahmad Musa, Hang Chen, Mihai Surdeanu, and Stephen Kobourov. *What does the language of food say about us?* In preparation.
Hoang Van, Dongfang Xu, Kris Brown, and Edmon Begoli. *Disambiguation in Concept Extraction in Mental Health.* In preparation.
Hoang Van, Kris Brown, and Edmon Begoli. *Clinical XLNet, ELMo, GloVe Embeddings.* In preparation.