



Amir Yazdavar

Ph.D. Candidate in Computer Science

Kno.e.sis Center

Wright State University

Dayton, Ohio – 45435

(+1) 937-9522304

yazdavar.2@wright.edu

<http://knoesis.org/researchers/amir/>

OBJECTIVES AND RESEARCH INTERESTS

I seek an opportunity where I can further enhance my research in machine learning (incl. deep learning) and semantic web (incl. creation and use of knowledge graphs) and their applications to NLP/NLU and social media analytics. I have particular interest in extraction of subjective information with applications to search, social and biomedical/health applications.

EDUCATION

Ph.D., Computer Science

2015-Present

Wright State University-(Kno.e.sis)

GPA: 4.0/4.0

Thesis Advisor – Prof. Amit Sheth

M.Sc., Computer Science

2012-2013

University Technology Malaysia (UTM), Malaysia

GPA: 3.95/4.0 (2nd place among all M.Sc. students of Computer Science)

Thesis: ‘Fuzzy Based Implicit Sentiment Analysis on Quantitative Sentences in Drug Reviews’

B.S, Computer Science

2006-2011

Shiraz University, Faculty of Computer and Electrical Engineering, Iran

RESEARCH EXPERIENCE

Graduate Research Experience (Ph.D.)

Data Scientist Intern

2017-Summer

Information Sciences Institute (ISI), University of Southern California (USC)

- **EFFECT**: Research/develop a robust big data platform to find traces of early planning activity by malicious actors, from unconventional sources including dark web and social media sites, forum discussions unstructured natural language text, structured data, and public network traffic data, and analyze these data streams to generate warnings of pending cyberattacks.

Keywords: Deep Learning/Machine Learning, Social Media Analysis, Natural Language Processing, Psychological assessment

Research Assistant

2015-Present

Kno.e.sis Center – Wright State University, Dayton, Ohio

Working on several real world projects mainly focused on studying human behavior on the web via Natural Language Understanding, Social Media Analytics utilizing Machine learning (Deep learning) and Knowledge Graph techniques. In particular, my focus is to enhance statistical models via domain semantics and guidance from offline behavioral knowledge to understand user's behavior from unstructured and large-scale Social data.

- **Modeling Social Behavior Depression:** Depression is one of the most common mental disorders in the U.S. and is the leading cause of disability affecting millions of Americans every year. Successful early identification and treatment of depression can lead to many other positive health and behavioral outcomes across the lifespan.

On this *NIH* funded project, we are applying “big data” techniques in collaboration with domain experts from *Weill Cornell Medical College*. Leveraging combinations of online socio-behavioral factors and neighborhood environmental conditions, we are developing state of the art techniques to detect depressive behavior in communities. The result of our study will be utilized for policy designers to facilitate the accessibility of health-care services.

Keywords: Deep Learning/Machine Learning, Social Media Analysis, Natural Language Processing, Knowledge Base Completion, Big Data

- **Twitris 3.0 – Sentiment Analysis for Analyzing Presidential Election:** In this project, we utilize the state of the art deep learning and machine learning techniques to monitor/study user's sentiment during US 2016 presidential election. This was one of the several components that allowed us to correctly predict the election outcome. <http://bit.ly/TElec2016>

- **Gender-Based Violence in 140 Characters or Fewer: A BigData Case Study of Twitter.** Humanitarian and public institutions are increasingly relying on data from social media sites to measure public attitude, and provide timely public engagement.

In this study, we examine Big (Social) Data to analyze public opinion regarding GBV, highlighting the nature of tweeting practices by geographical location and gender. Studying tweeting practices by pragmatic function (such as assertion or belief) and utilizing a data driven approach that combines both supervised and unsupervised machine learning techniques, we first develop a classifier that checks for relevance to GBV topics, and further distinguishes pragmatic function classes using content features such as unigrams, and categories of psychologically meaningful analysis (using LIWC). We reveal public awareness regarding GBV tolerance and suggest opportunities for intervention and assisting both governmental and non-governmental organizations in policy development.

Keywords: Machine Learning, Social Media Analysis, Natural Language Processing, Unsupervised Approach, Supervised approach, Topic Modeling, LIWC, Big Data

Graduate Research Experience (M.Sc.)

Research Assistant

2012-2013

Universiti Teknologi Malaysia (UTM)

Working on understanding patient's sentiment on online social media platforms including medical forums and studying drug effectiveness and side effect in unstructured user generated content by developing fuzzy based statistical models.

Keywords: Machine Learning, Social Media Analysis, Natural Language Processing, Fuzzy Modeling, Supervised approach

TEACHING EXPERIENCE AND MENTORING

- Lecturer in Bahonar Technical and Engineering college, Shiraz, Iran (Fall 2014)
- Mentoring:
 - [Ankita Saxena](#)(Master student) for a project on [studying depressive behavior in social media](#), Fall 2016
 - [Mohammad Saeid Mahdavinejad](#) (Remote Master student) for a project Analyzing user's demographic information on Twitter, Fall 2016
 - [Goonmeet Bajaj](#) (Undergrad Senior) for a project on [Measuring Gender-Based Violence Attitude on Twitter](#), Summer 2016

PUBLICATIONS [\[Google Scholar\]](#)

Journal Articles:

- 2017
 - [1] On the Difficulty of Sentiment Analysis for Dynamic Events: Presidential Election, Monireh Ebrahimi, **Amir Hossein Yazdavar**, Amit Sheth, IEEE Intelligent Systems, Impact Factor:2.34, Accepted
 - [2] Identifying Pragmatic Functions in Social Media Indicative of Gender-Based Violence Beliefs, Tanvi Banerjee, **Amir Hossein Yazdavar**, Hemant Purohit, Andrew Hampton, Valerie L. Shalin, Amit P. Sheth, (In Preparation to submit to Computer in Human Behavior)
- 2016
 - [1] Fuzzy Based Implicit Sentiment Analysis on Quantitative Sentences, **Amir Hossein Yazdavar**, Monireh Ebrahimi, Naomie Salim, Journal of Soft Computing and Decision Support Systems.
 - [2] Recognition of Side Effects as Implicit-Opinion Words in Drug Reviews, Monireh Ebrahimi, **Amir Hossein Yazdavar**, Naomie Salim, Safaa Eltyeb, Online Information Review, Impact Factor: 1.15.
 - [3] Rock strength estimation: a PSO-based BP approach, E.Tonnizam Mohamad, D. Jahed Armaghani, E. Momeni, **Amir Hossein Yazdavar**, Monireh Ebrahimi, Journal of Neural Computing and Applications, Impact Factor: 1.49.
- 2014
 - [1] Transmission of Data with OFDM Technique for Communication Networks Using GHz Frequency Band Soliton Carrier, Iraj Sadegh Amiri, Monireh Ebrahimi, **Amir Hossein**

Yazdavar, S. Ghorbani, S. E. Alavi, Sevia M. Idrus , J. Ali, IET Communications Journal (IEEE), Impact Factor: 0.74.

[2] Analytical Modeling and Simulation of IV Characteristics in Carbon Nanotube based Gas Sensors using ANN and SVR Methods, Elnaz Akbari, Zolkafle Buntat, Aria Enzevae, Monireh Ebrahimi, **Amir Hossein Yazdavar**, Rubiyah Yusof, Chemometrics and Intelligent Laboratory Systems, Impact Factor: 2.32.

Conference Articles:

- 2017

[1] Relatedness-based Multi-Entity Summarization, Kalpa Gunaratna ,**Amir Hossein Yazdavar** ,Krishnaprasad Thirunarayan ,Amit Sheth, and Gong Cheng (Accepted IJCAI 2017).

Representing world knowledge in a machine processable format is important as entities and their descriptions have fueled tremendous growth in knowledge-rich information processing platforms, services, and systems. In this research, we present an approach that can summarize facts about a collection of entities by analyzing their relatedness in preference to summarizing each entity in isolation.

[2] Semi-Supervised Approach to Monitoring Clinical Depressive Symptoms in Social Media, **Amir Hossein Yazdavar**, Hussein S. Al-Olimat, Monireh Ebrahimi, Tanvi Banerjee, Krishnaprasad Thirunarayan, Jyotishman Pathak and Amit Sheth, (Accepted ASONAM 2017).

Using huge set of tweets crawled from users with self-declared depression symptoms, our study uses a novel, semi-supervised LDA-based model to monitor depression symptoms through their expression on Twitter (in terms of word usage patterns and topical preferences) to emulate PHQ-9 depression recognition system. Our screening tool is able to identify clinical depressive symptoms with an accuracy of 68% and precision of 72%.

[3] Word Embedding Based Data Augmentation for Harassment Detection: Lexical Substitution Approach, Monireh Ebrahimi, **Amir Hossein Yazdavar**, Lu Chen, Krishnaprasad Thirunarayan, Amit P. Sheth (In preparation).

All machine learning based approaches used to detect on-line harassment so far rely merely on the labeled training instances. The bottleneck in deep learning is creating sufficiently large and balanced labeled data. To tackle this problem, we developed a novel augmentation approach using the dependency based word embedding model, to replace each word in the training data set with more semantically and syntactically meaning-preserving substitute. We seek to demonstrate better result for RNN, CNN, LSTM and GRU approaches.

[4] Mining Adverse Drug Reaction in Social Media via Deep Recurrent Neural Networks and Background Knowledge, **Amir Hossein Yazdavar**, Hussein S. Al-Olimat, Monireh Ebrahimi, Goonmeet Bajaj, Krishnaprasad Thirunarayan, Jyotishman Pathak and Amit Sheth, (In preparation for IEEE Social Computing 2017).

Adverse drug reactions (ADRs) are considered major public health challenge and the sixth leading cause of death worldwide. Unwanted drug effects also have significant economic as

well as clinical costs as they often lead to hospital admissions. Consequently, we introduce DeepADRMine, a semantic annotator tool which extract medical concepts from unstructured text by integrating lexicon-based method (top-down processing) and data-driven method (bottom-up processing). In particular, leveraging available ADR lexicon (FAERS) we first extract explicit and non-descriptive ADRs. Then, to further enhance the semantic understanding of user generated contents, we treat the implicit ADRs (descriptive) detection as a task of sequence labeling.

Invited Talks/Poster presentation

- 2017
 - [1] Understanding clinical depressive symptoms in social media ([NLP seminar/Information science Institute-ISI](#))
 - [2] Identifying Depressive Disorder in the Twitter Population (Poster presentation) Goonmeet Bajaj, **Amir Hossein Yazdavar**, T.K. Prasad, Amit Sheth, Wright Brothers Conference at Wright State University, Dayton, OH, USA. October, 2017.
- 2016
 - [1] Analyzing Clinical Depressive Symptoms in Twitter (Poster Presentation), **Amir Hossein Yazdavar**, Hussein S. Al-Olimat, Tanvi Banerjee, Krishnaprasad Thirunarayan, Jyotishman Pathak, Amit Sheth, Conference on Mental Health Services Research: Harnessing Science to Strengthen the Public Health Impact (MHSR 2016)
 - [2] Analyzing Depression via Social Media(Poster Presentation), **Amir Hossein Yazdavar**, Big Data Surveillance Analytics Conference at Wright State University, Dayton, OH, USA. October, 2016.

PROFESSIONAL ACTIVITIES AND SERVICES

- External Reviewer – ICWSM 2016, IEEE BigData 2016, WWW 2017, IEEE BigData 2017, International Conference on Applications of Natural Language to Information Systems, NLDB 2016
- Recognized Reviewer, Journal of Biomedical Informatics (2014-present) ,Applied Soft Computing

AWARDS AND GRANTS

- Computer Science Graduate Fellowship Award, Wright State University, Monthly stipend and a full tuition fellowship, given to a select handful of the most outstanding incoming computer scientist applicants each year, 2015
- Ranked 2nd place among all M.Sc. students of Computer Science, 2013
- Ranked Top 1% among over 300,000 participants in a nationwide universities entrance exam for undergraduate education, 2006

GRANT WRITING

- Prestigious FRGS grant awarded from Malaysia government for Implicit Opinion Mining Model for Drug Effectiveness and Side Effect Recognition in Medical Reviews. (Vote no: R.J130000.7828.4F373, RM81000), Role- Significant Contribution, 2013

COMPUTER SKILLS

- **Programming Language:** JAVA, PYTHON (Scikit-learn, Pandas), C, MATLAB, MySQL, Elastic Search
- **Deep Learning:** Theano, Keras, TensorFlow
- **Machine Learning Algorithms and Tools:** Neural Network, Particle Swarm Optimization, Genetic Algorithm, FUZZY models, Linear Regression, Logistic Regression, Random Forest, Topic Modeling, LDA, Latent Semantic Analysis
- **Semantic Technologies:** RDF, OWL, Ontology, SPARQL
- **Natural language Processing Tools:** GATE, Lucene, LIWC, NLTK
- **Medical Informatics:** UMLS, MetaMap
- **Scalable Computing:** Map-Reduce, Pig, Hadoop, Spark

SELECTED GRADUATE COURSES

Deep Learning, Advanced Artificial Intelligence, Statistical analysis, Cloud Computing, Advanced Semantic Web, Advanced Database Systems, Machine Learning, Advanced Algorithm Design, Proposal Writing

REFERENCES

- **Prof. Amit Sheth:** Director, Kno.e.sis Center Wright State University, Dayton, OH, 45435 USA. amit@knoesis.org (937) 239-0625
- **Prof. Krishnaprasad Thirunarayan:** Kno.e.sis Center Wright State University, Dayton, OH, 45435 USA. t.k.prasad@wright.edu (937) 775-5109
- **Dr. Jyotishman Pathak:** Division of Health Informatics, Weill Cornell Medical College, New York, NY, USA jyp2001@med.cornell.edu
- **Dr. Kristina Lerman:** Project Leader, USC Information Sciences Institute, Marina del Rey, LA, CA, USA lerman@isi.edu