

Alex Morales | CV

✉ amorale4@illinois.edu • 🌐 amorale4.github.io

My research interests lie in the intersection of Machine Learning, Natural Language Processing and Text Mining. I mainly focus on **Text Mining for Social Media Analysis**, **Community Question Answering** and **Identifying Quality Content in Online Communities**. I have collaborated on interdisciplinary projects that touch societal problems related to health. I am excited about work in areas such as Misinformation, Truth Discovery and User Behavior Modeling.

Education

University of Illinois at Urbana-Champaign

PhD Computer Science, August 2014 - Fall 2020, Advisor: Prof. ChengXiang Zhai

Thesis Topic: *Model-based Feature Construction and Text Representation: A case study of social media analysis*

Committee Members: Jiawei Han, Julia Hockenmaier, Lyle Ungar

University of California, Santa Barbara

M.S. Computer Science, June 2014, Advisor: Prof. Xifeng Yan

Thesis Topic: *Detecting Synthetic Review Spam*

University of California, Santa Barbara

B.S. Computer Science, June 2012

Research Experience

Categorizing Severity in COVID-19 Misinformation

TIMan Group

Categorized the granularity of severity in COVID-19 misinformation using social media posts.

Developed misinformation data collection and annotation scheme, and created a website to streamline this process.

Helped develop several deep-learning model baselines using BERT, HAN, and deFEND.

DAIS @ UIUC

Spring 2020 - Fall 2020

Identifying Inspector Bias via Inspector Report Consistency Analysis

TIMan Group, in collaboration with the Champaign-Urbana Public Health Department

Worked with the CU-PHD to identify inconsistencies and biases in health inspection reports by applying text mining methods.

Developed a text classification method for re-inspection prediction solely based on inspector comments.

Performed inspector experience bias analysis based on temporal scoring patterns over an archive of reports from 2008-2018.

DAIS @ UIUC

Spring 2020 - Fall 2020

Reliable Question Answering in Online Discussion Forums

TIMan Group

Lead the creation of a novel community discussion forum dataset by crawling several ask communities on Reddit.com.

Proposed and developed a user-aspect reliability model using both user topical expertise and comment context information in order to identify trustworthy and reliable comments from users with diverse expertise and perform user-expertise identification.

DAIS @ UIUC

Spring 2018 - Spring 2019

HIV Radar Project

Interdisciplinary research, PI: Prof. Dolores Albarracín

Analyzed social media influences on HIV/STI transmission behaviors by combining Twitter messages and socio-demographic variables for risk factor discovery. Helped develop the Twitter crawlers, maintained the MongoDB server to store the tweets, and developed several related predictive risk factor models using social media data.

Social Action Lab @ UIUC

Fall 2015 - Fall 2020

Automatic Humor Identification in User Reviews

TIMan Group

Proposed and developed model to capture incongruity, unexpectedness, and language differential semantics in reviews from the Yelp Dataset Challenge to identify humorous intent and high quality reviews.

DAIS @ UIUC

Fall 2014 - Spring 2016

Teaching Experience

MOOC - Text Mining and Analysis/Data Mining Capstone

Teaching Assistant @UIUC

Instructor: Prof. ChengXiang Zhai

Helped develop the first offering of the Data Mining Capstone Course offered online for the MCS-DS degree. Created the online quizzes and exams used by Coursera. Developed the data mining capstone project and grading rubrics.

Summer 2015, Summer 2019

CS290D - Advance Data Mining

Teaching Assistant @UCSB

Spring 2014

Instructor: Prof. Xifeng Yan

Sole TA for the first offering of the graduate Data Mining course. Gave guest lectures, and graded students' assignments and exams.

CS165A - Artificial Intelligence

Teaching Assistant @UCSB

Spring 2013

Instructor: Prof. Xifeng Yan

Lead the TAs to ensure the large undergraduate course was managed efficiently.

Work Experience

Adobe Research - Systems Technology Lab

Data Scientist Intern

Summer 2014

Mentor: Dr. Nedim Lipka

Temporal Summarization on Real-Time Social Media Streams. Developed an event summarization prototype system on Twitter using Spark and GraphX.

Publications

- [1] Arkin Dharawat, Ismini Lourentzou, **Morales, Alex**, and ChengXiang Zhai. Drink bleach or do what now? covid-hera: A dataset for risk-informed health decision making in the presence of covid19 misinformation. *arXiv preprint arXiv:2010.08743*, 2020.
- [2] **Morales, Alex**, Kanika Narang, Hari Sundaram, and ChengXiang Zhai. Crowdqm: Learning aspect-level user reliability and comment trustworthiness in discussion forums. In *The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2020.
- [3] Nupoor Gandhi, **Morales, Alex**, Sally Man-Pui Chan, Dolores Albarracín, and ChengXiang Zhai. Predicting opioid overdose crude rates with text-based twitter features (student abstract). In *the AAAI Conference on Artificial Intelligence. Vol. 33*, 2020.
- [4] **Morales, Alex**, Nupoor Gandhi, Man-pui Sally Chan, Sophie Lohmann, Travis Sanchez, Kathleen A Brady, Lyle Ungar, Dolores Albarracín, and ChengXiang Zhai. Multi-attribute topic feature construction for social media-based prediction. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 1073–1078. IEEE, 2018.
- [5] Sophie Lohmann, Benjamin X White, Zhen Zuo, Man-Pui Sally Chan, **Morales, Alex**, Bo Li, Chengxiang Zhai, and Dolores Albarracín. Hiv messaging on twitter: an analysis of current practice and data-driven recommendations. *Aids*, 32(18):2799–2805, 2018.
- [6] Man-pui Sally Chan, Sophie Lohmann, **Morales, Alex**, Chengxiang Zhai, Lyle Ungar, David R Holtgrave, and Dolores Albarracín. An online risk index for the cross-sectional prediction of new hiv chlamydia, and gonorrhea diagnoses across us counties and across years. *AIDS and behavior*, 22(7):2322–2333, 2018.
- [7] Ismini Lourentzou, **Morales, Alex**, and ChengXiang Zhai. Text-based geolocation prediction of social media users with neural networks. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 696–705. IEEE, 2017.
- [8] **Morales, Alex** and ChengXiang Zhai. Identifying humor in reviews using background text sources. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 492–501, 2017.
- [9] **Morales, Alex** and ChengXiang Zhai. Forecasting research trends by mining literature findings. In *Proceedings of the 11th Annual ACM Richard Tapia Celebration of Diversity of Computing*. ACM, 2017.
- [10] Huan Sun, **Morales, Alex**, and Xifeng Yan. Synthetic review spamming and defense. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1088–1096. ACM, 2013.
- [11] **Morales, Alex**, Huan Sun, and Xifeng Yan. Synthetic review spamming and defense. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 155–156. ACM, 2013.

Awards

FDO-Managed Retail Program Standards Grant, UIUC 2019-2020

Graduate College Distinguished Fellowship, UIUC 2015-2018

Support for Under-Represented Groups in Engineering PhD Fellowship, UIUC 2014-2017

GEM PhD Fellowship, Adobe Fellow, The National GEM Consortium, UIUC 2014

Hispanic College Fund Scholarship, Ford Scholar, UCSB 2011

Hispanic College Fund Scholarship, Google Scholar , UCSB 2010
College of Engineering Scholar - UCSB 2009

Other Activities/Service

CIKM 2020 Applied Research Program Committee Member, CIKM, Summer 2020
Summer Research Experience for Undergraduates Mentor, CS REU @ UIUC, Summer 2020
Illinois Scholars Undergraduate Research Mentor, ISUR @ UIUC, Fall 2017 - Spring 2018
Sloan University Center of Exemplary Mentoring at Illinois, Peer Mentor, UCEM @ UIUC, Fall 2016 - Spring 2017
Mentoring Undergraduates in Science & Engineering, MUSE @ UIUC, Fall 2016 - Spring 2017
CS Graduate Ambassador, CS @ UIUC, Spring 2015
Multicultural Engineering Recruitment for Graduate Education, Graduate Ambassador, MERGE @ UIUC, Spring 2014, 2015, 2016, 2017, 2018

Technical Skills

Programming Languages, Libraries and Tools: Python, Java, Scala, Matlab, C/C++, SQL/mysql, UNIX Shell Scripting, LaTeX, HTML, MongoDB, Spark, GraphX, pandas
Deep Learning and Other packages: pyTorch, torchtext, keras, scikit-learn