MOHAMMAD S. ALODADI

Herndon VA, 20171
+1(443)449-1652
mso1830@gmail.com
https://www.mohammadalodadi.com

PROFILE SUMMARY

Highly analytical and results-oriented data scientist with experience in managing technical teams and meeting objectives within strict timelines. His ability to handle ambiguity and coordinate with cross-organizational teams underlines his practical approach. He can simplify and present complex ideas effectively, and has a strong grasp of data collection and analysis from varied resources. He has a keen interest in Data Science, Machine Learning, Deep Learning, Data Mining, Statistical Data Analysis, and Information Retrieval, with notable work in Natural Language Processing and Large Language Models.

TECHNICAL COMPETENCIES

- Programming and Scripting Languages: Python, R, Bash script
- Machine Learning & Deep Learning Frameworks: Large Language Models (LLMs), Transformers Huggingface, Pytorch, Keras, Scikit-learn, LSTM, GRU, CNN, Distant supervision, Weak supervision, GPT 3.5 API, Transfer learning techniques.
- Natural Language Processing: Word2vec, Doc2vec, FastText, Smooth Inverse Frequency, Averaged Sentence Embeddings, Latent Dirichlet allocation (LDA)
- Data Analytics and Processing: Pandas, Gensim, NumPy, T-SNE, IBM Bluemix personality insight's API, Multiprocessing, Batch processing techniques
- Database and Data Mining Technologies: SQL, PL\SQL, Cassandra, MongoDB, Elasticsearch, Faiss, Pyterrier, PubMed Central® (PMC)
- Automation and Workflow Management: Apache Airflow
- Web Development and API (Beginner): JavaScript, TypeScript, API Development, NextJS
- Cloud Computing Environments: AWS, Google Cloud
- High Performance Computing and Parallel Computing: HPC, SBATCH environments
- Interactive Data Analysis and Visualization: Jupyter notebooks, Streamlit, Gradio, Matplotlib, seaborn, Plotly, Bokeh
- Other Skills: Linux, Social networks content analysis, Topic modelling, Spam detection, Personality traits extraction

HIGHLIGHTS

- Broad skill set and proven success driving diverse technical teams to meet objectives under strict timelines and in dynamic, high-pressure environments.
- Ability to work through ambiguity and drive cross-organization teams to achieve results. Self-assured and confident.
- Exceptional oral presentation skills, including the ability to simplify complex thoughts/ideas, persuasion and excellent interpersonal skills.
- Excellent skills in gathering and analyzing data from heterogeneous resources and performing data ingestion and harmonization.
- Interests:
  Data Science - Machine Learning, Data Mining, Deep Learning, Statistical Data Analysis, Information Retrieval

RELEVANT SKILLS EXPERIENCE

- Developing an End-to-End pipelines for processing, analyzing, and extracting information from text dataset using state of the art NLP/AI and LLMs models.
- Developing state of the art NLP/AI models to identify the human genotype-phenotype-molecular associations with disease model systems data.
- Development of methods for improving NER performance by using Distant supervision and weak supervision.
- Utilizing GPT 3.5 API for building a QA module for medical domain
- Drive the implementation of hierarchical deep learning architectures, including recurrent neural network (RNN) (LSTM, GRU) and convolutional neural network (CNN), to extract information from patient experience data on GPU units using High-Performance Computing (HPC).
- Trained word embedding models (Word2vec, Doc2vec, FastText) on the entire PubMed biomedical literature (34 million abstracts, 30 GB) for the transfer learning task.
- Validation of word embedding using T-SNE dimensionality reduction algorithm and visualization.
- Leveraged transfer learning techniques for sentence embedding encoders using Smooth Inverse Frequency compared to the baseline Averaged Sentence Embeddings.
- Index, vectorize and store the PubMed Central® (PMC) (1.25 million full texts, 87.8 GB) using Elasticsearch and Fakebook's Faiss for efficient similarity search and clustering of dense vectors.
- Effective utilization of IBM Bluemix personality insight's API to extract personality traits for users based on their content on social networks to reconstruct content-based networks. Similarly, we extracted topics using the Latent Dirichlet allocation LDA algorithm to reconstruct the topical-based network.
- Perform extraction of topics from users' content on social networks using Latent Dirichlet allocation LDA algorithm to reconstruct topical-based network. Similarly, we also used LDA to gain insight into Spam detection.

- Ensure efficient pre-processing of large amounts of data using parallel computing library multiprocessing or using batch processing techniques in Gensim, scikit-learn libraries.

EDUCATION

- **University of Maryland Baltimore County,** Baltimore, MD
  Ph.D. Information Systems ***2020***
  - *Thesis: (Knowledge Discovery Through Linking Multiple Heterogeneous, Unstructured Data Streams: A Case of Clinical Notes Mining)*
  - Advisor: Prof. Vandana Janeja

  M.S. in Information Systems ***2014***
- **King Khalid University,** Aseer Province, Saudi Arabia ***2006***
  B.S in Computer Science and Education

PROJECTS & RESEARCH EXPERIENCE

Entity Relationship Extraction for Rare Disease Phenotype-Genotype-Variants Associations, 2021

- Applied and compared many medical information extraction algorithms from unstructured biomedical literature.
- Fine-tune large biomedical transformers language model to extract information from biomedical literature.
- Built a pipeline to automate manual curation and molecularly characterize rare diseases and render it possible to mine rare disease phenotype-genotype associations.
- Our project is release Ver 1.0 to the public on Feb,2023 on NIH, Rare Disease Day and accompanied by a poster titled: "RARe-SOURCE™: Integrated Bioinformatics Resource for Rare Diseases" **https://raresource.nih.gov**

Semi-supervised Aspect Auto Encoder Extraction from Medical Reviews, 2021

- Developed methods to extract targeted aspects from medical reviews using self-attention based neural networks.
- Developed and implemented sampling method to overcome the issue of unbalanced training dataset using diversity clustering methods to elect candidates sentences from large unlabeled datasets.
- Compared my methods with state-of-the-art text augmentation and text sampling techniques.
- The algorithm would reduce the estimation of the cost of manual labeling from 30k to only 3k for multiple annotators.

Linking Knowledge Discovery in Clinical Note and Massive Biomedical Literature Repositories, 2020

- The aim of this research is to discover knowledge from clinical notes by identifying interlink clinical events. The extracted knowledge supports decision making process and supports the facts with biomedical literature to optimize evidence-based medicine utilization.
- Used NLP, text mining, and data mining methods to extract, analyze, clinical entities' associations from clinical notes in EHR.
- Explored widely used query expansion strategies, local analysis, global analysis, and ontology-based term re-weighting across various search engines.
- Successful development of contextual search index using Approximate Nearest Neighbors to enhance information retrieval precision using aggregation strategy with traditional information retrieval methods.
- Successfully developed an ensemble-based dual embedding information retrieval re-ranking system that increases the precision of retrieval documents.

Clinical Entities Association Rules (CLEAR): Untangling Clinical Notes in Electronic Health Records, 2019

- The focus of this research is to discover patterns of important associations amongst clinical entities, such as treatments, tests, and diagnoses, particularly in clinical notes.
- Developed weighted association rule mining methods to create an end-to-end pipeline to transform clinical notes corpus into a knowledge set.
- The results with labeled data show an accuracy of 98.3% on the relevancy task and 54.4% on the interestingness level.

The similarity in Patient Support Forums Using TF-IDF and Cosine Similarity Metrics (Challenge Finalist list), 2015

- This project was created as a participation in the IEEE ICHI 2015 data science challenge. This project was selected from the finalist list to be presented and demonstrated at the conference.
- The problem of the challenge is to reduce the repetition of posts for patient support forums.
- This paper adopts a standard approach to finding related questions, measuring text similarity by computing the cosine between the questions when each was represented with a bag of words. The significance of this paper lies in the fact that it shows that a simple, straightforward approach can fare well on this task compared to more complicated approaches such as Latent Dirichlet Allocation (LDA), or Latent Semantic Index (LSI).

Content-Based Community Detection Based on Community Network Link Structure, 2016

- Investigated the ability to identify social networks in health-related networks using user's content methods such as topic modeling algorithms and personality traits features.
- Compared content-based networks with the linked-based structured network which is based on community following-followed graph.
- The results indicate that the two network graphs do not necessarily resemble one another due to soft and hard community membership which play a significant role in shaping the network graph.

Mass Sentiments Analysis of FCC Comments: Net Neutrality Policy, 2015

- Investigated the geographical sentiment of 800 thousand comments on the net neutrality bill by FCC.
- Built a model that can classify people's opinion whether their submission was with, against, or neutral toward the details of the bill.
- We annotated the opinions presented in the comments and trained a classifier to automate the opinion extraction method.
- We showed that sentiments and opinions are not matched throughout the comments which agree with preserving net neutrality but uses negative sentiment.
- Visualized the different opinions and sentiments on the US map to have better understanding of the location of comments participants.

D3I: Discovery of Drug-Drug Interaction, 2014

- Implemented methods that utilize heterogeneous data sources that will identify possible drug-drug interactions from clinical notes.
- We utilized NLP tools and clinical information extraction systems to create a system that can extract DDIs.
- The pipeline was constructed to unify drugs mentions including various ingredients of drugs using RxNorms API, NDF-RT API, and Drug Bank as DDI descriptions supplementary.
- The proposed system can be used with various data sources such as patient health forums where patients seek first to inquire about alternative illness treatments.

PUBLICATIONS, PATENTS, GRANTS, AND PRESENTATIONS

[1] M. Alodadi, E. Lyons, et..al *"Rare disease variant curation from literature: assessing gaps with creatine transport deficiency in focus",* 2023 (BMC Genomics Journal). JUN 2023

[2] M. Alodadi, E. Lyons, et..al *"Rare Diseases Phenotype-Genotype Associations from Biomedical Literature: Language Model Workflow"* NCI-Frederick 2023 Spring Research Festival, APR 2023 (Poster) **(Winner poster in the category of Informatics)**

[3] M. Alodadi, E. Lyons, et..al *"RARe-SOURCE™: Integrated Bioinformatics Resource for Rare Diseases"* , NIH Rare Disease Day 2023, Feb 2023(Poster)

[4] M. Alodadi, E. Lyons, et..al *"Rare Disease Gene Variant Curation: Creatine Transport Deficiency In Focus"* NCI-Frederick 2023 Spring Research Festival, APR 2022 (Poster) **(Winner poster in the category of Gene Therapy, Genome Editing, and Genetics)**

[5] M. Alodadi, A. Goldstein, "*Opportunities For Ethical Decision Making: A Case Study In K-NN*", 17th annual International Technology, Education and Development Conference (INTED2023), Dec 2022.

[6] M. Alodadi, V. Janeja, *"Clinical Entities Association Rules (CLEAR): Untangling Clinical Notes in Electronic Health Records,"* IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Nov 2019.

[7] M. Alodadi, V. Janeja, *"Linking Knowledge Discovery in Clinical Notes and Massive Biomedical Literature Repositories,"* IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Nov 2019.

[8] M. Alodadi, *"Radiology Clinical Notes Mining Using Weighted Association Rules,"* IEEE International Conference on Healthcare Informatics (ICHI), Aug 2017.

[9] M. Alodadi, V. Janeja, *"Similarity in Patient Support Forums Using TF-IDF and Cosine Similarity Metrics,"* IEEE International Conference on Healthcare Informatics (ICHI), Oct 2015. **(Finalist in IEEE-ICHI Challenge)**

Presentations and Tutorials:

[10] M. Alodadi, V. Janeja *'Anomaly Detection for Keystroke Dynamics using Clustering Algorithms, Hands-on experience'* Dec 2017.

PROFESSIONAL EXPERIENCE

Frederick National Laboratory for Cancer Research (FNLCR), Frederick**,** MD

- Computational Scientist — Sep 2023 – Present
- Bioinformatics Analyst — Feb 2021 – Sep 2023

University of Maryland Baltimore County, Baltimore**,** MD

- Research Associate/general assistant — Oct 2020 – Jan 2021

University of Maryland Baltimore County, Baltimore**,** MD

- Graduate School Dissertation Fellowship Jun 2020 – Aug 2020

University of Maryland Baltimore County, Baltimore**,** MD

- Graduate Teaching Assistant, Aug 2015 – Jun 2020
  - IS410, (undergraduate), Introduction to Database Design, IS676 (Graduate) Information Integration, IS420 (Undergraduate) Database Applications Development, IS733 (Graduate) Data Mining, IS734 (Graduate) Data Analytics for Cyber security

Department of Education, Asser Province, Saudi Arabia

- Computer Science Instructor Jul 2006 - Sept 2011
  Computer Literacy course and Advanced Computing Topics (High School, Grade 10-12th)
- Computer Science Assistant Sept 2005- May 2006
  Computer Literacy Course (High School, Grade 10-12th)

AWARDS AND SCHOLARSHIPS

- 2023 National Center for Advancing Translational Sciences (NCATS) Director's Award.
  - For the conception and development of "RARe-SOURCE™ ", an innovative bioinformatics data platform to facilitate therapeutic discovery and advance translational science more quickly for rare diseases.
- 2023 Frederick National Laboratory for Cancer Research Outstanding Achievement Award – Scientific.
  - For the team's commitment to improving the diagnosis and treatment of rare diseases by developing an integrated bioinformatics resource for rare disease research, showcasing originality, impact, and dedication beyond normal job expectations.
- NCI-Frederick 2023 Spring Research Festival Outstanding Poster in the category of Informatics with a poster titled, "Rare Diseases Phenotype-Genotype Associations from Biomedical Literature: Language Model Workflow" Apr 2023
- Frederick National Laboratory for Cancer Research travel grant to attend SIG KDD2022-DC
- Frederick National Laboratory for Cancer Research travel grant to attend IEEE BIBM 2021
- UMBC Summer 2020 Dissertation Fellowship Jun 2020 – Aug 2020
- UMBC Graduate Students Association Travel grant Nov 2019
  For presenting two papers at IEEE BIBM 2019
- UMBC Department of Information Systems Travel grant Nov 2019
  For presenting two papers at IEEE BIBM 2019
- NSF Workshop travel grant Jun 2019
  For attending as a student volunteer organizer for an Invitation-only workshop on Including Ethics in Data Science Pedagogy EDSP 2019- Alexandria, VA
- UMBC Graduate School Doctoral Candidacy Awarded May 2019
- IEEE ICHI Travel award Aug 2017
  For attending and presenting my work on clinical notes, mining
- UMBC Department of Information Systems Travel grant Aug 2017
  For presenting my research and attending Doctoral Students Consortium at IEEE ICHI 2017
- IEEE ICHI Analytics challenge Finalist Oct 2015
  For presenting my paper and attending Doctoral Students Consortium at IEEE ICHI 2015
- Saudi Arabia Scholarship for PhD Degree Aug 2014
  Full-tuition scholarship with a stipend for PhD studies. One of 300 awardees in 2014

LEADERSHIP AND VOLUNTEER

- Poster + Booth presenter at the NIH Rare Disease Day 2023.
- Organizing a hands-on showcase workshop at Frederick National Laboratory for Cancer Research 2022
- Peer-mentor for Leidos Biomedical Research Inc. newly hired employees on March 2022.
- Serve on the Selection Committee to help the Office of Data Science Strategy review IC proposals for the 2022 Data and Technology Advancement (DATA) National Service Scholar Program under the NIH (2021)
- Research affiliate at Mdata Lab, Information Systems Department, UMBC.
- Co-mentoring multiple Research Experience for Undergraduates (REU) Program for data science area at University of Maryland Baltimore country.2021
- (Volunteer organizer) at the 2020 Data Science Leadership Summit
- IEEE 32nd International Conference on Tools with Artificial Intelligence ICTAI program committee member(2019, 2023)
- University of Maryland, Baltimore County, Graduate Student Representative in Faculty Promotion and Review Committee, 2019– 2020
- Engaged actively in hiring and mentoring a few graduate students who are employed through grants in the DINAMIC lab
- Technical mentoring multiple Cyber scholar undergraduate students.
- Student volunteer at NSF funded Workshop (invitation-only scholars' event) on Including Ethics in Data Science Pedagogy, 2019.
- Student volunteer at IEEE International Conference on Healthcare Informatics, 2017.
- Peer Reviewer volunteer for Encyclopedia of GIS 2nd edition, 2015.
- Student volunteer at IEEE International Conference on Healthcare Informatics, 2015.
- Student volunteer at IEEE Intelligence Security Informatics Conference, 2015

RELATED COURSEWORK

- Data Mining, Systems, and Information Integration, Computational Methods, Advanced Quantitative Methods, Cyber security Analytics, Social Media Analytics, Decision Making Support Systems.