Mahsa Khoshnoodi

RESEARCH INTERESTS

- **Natural Language processing**
- **Computer Vision**
- **Machine learning**
- **Vision-Language Models**
- **Large language Models**

EDUCATION

Amirkabir University of Technology Tehran, Iran
M.sc. in Department of Computer Engineering, Artificial Intelligence 2011–2014

- Supervisor: Prof. Mohammad Mehdi Ebadzadeh
- Thesis title: Improving computational models of Basal Ganglia for planning.
- GPA: **18.06/20 (3.91/4)**

Alzahra University Tehran, Iran
B.Sc. in Computer Engineering 2007–2011

- Supervisor: Dr. Reza Azmi
- Thesis title: Artificial Immune system with negative selection approach for anomaly detection
- GPA: **18.27/20 (3.89/4)**

RESEARCH EXPERIENCE

Predoctoral Fellow April 2023–Present
Fatima Al-Fihri Predoctoral Fellowship

- Proposed a novel metric based on object detection for characterizing object hallucination.
- Conducted research on automatic evaluation metrics for Text-to-Image Models.
- Developed evaluation criteria for assessing consistency and faithfulness in Text-to-Image models.

Artificial Intelligence Research Assistant 2012–2014
Amirkabir University of Technology, Computational Intelligence lab

- Proposed a fuzzy neural network combined with reinforcement learning to model the direct pathway of the Basal Ganglia, which plays a key role in action selection and habit learning in the brain.
- Introduced a recurrent neural network with a novel learning rule for the indirect pathway of the Basal Ganglia, evaluated through sequence learning and planning tasks.
- Implemented the project and conducted a comprehensive assessment.

PUBLICATIONS

Mahsa Khoshnoodi, Fatima Jahara, Michael Saxon, Yujie Lu, Aditya Sharma, William Yang Wang, *"Who Evaluates the Evaluations? Assessing the Faithfulness and Consistency of Text-to-Image Evaluation Metrics with T2IScoreScore"* Project Page (2023). Preprint

WORK EXPERIENCE

Son Corporate Group Oct 2018 – Present
Machine Learning Researcher

Advanced Analysis and Prediction of Stock Market Trends with Deep Learning Models

- Developed and customized a pre-trained large language model (LLM) for financial data in Persian.
- Investigated the role of social media dynamics, particularly sentiment analysis of online discourse, in predicting stock market trends.

- Conducted research in time series forecasting, utilizing Transformer-based models to predict stock prices.
- Applied LSTM models for stock price prediction, incorporating deep learning models to enhance forecasting accuracy.

Scalable Big Data Solution for In-Depth Stock Market Analysis

- Designed and implemented a distributed big data architecture, leveraging technologies such as Hadoop and Spark, to efficiently process and analyze vast volumes of stock market data in near real-time.
- Developed custom data pipelines for data ingestion, cleansing, and feature extraction, ensuring data quality and accuracy for robust financial analysis.
- Implemented machine learning models and statistical techniques for advanced pattern recognition and predictive analytics, enabling informed decision-making in the stock market.

Innovative Instance-Based Anomaly Detection engine for Banking Ecosystems

- Conducted feature engineering by extracting relevant statistical features from banking data, including entropy and standard deviation of transaction amounts within defined time frames.
- Profiled the behaviors of entities within banking ecosystems such as ATMs and bank accounts.
- Designed and implemented a novel anomaly detection algorithm based on distinctive similarity measures to enhance risk assessment and fraud detection.
- Utilized peer group analysis to calculate an aggregated risk score for anomaly behavior.
- Implemented a distributed solution using Apache Spark Streaming and Apache Kafka to enhance the scalability and real-time capabilities of the solution.

Datis Pars Java Solutions Co Dec 2014 – Sep 2018

Machine Learning Enginner

Advanced Customer Behavior Analysis with Hierarchical Clustering of GMMs

- Developed a Hierarchical Clustering model based on Gaussian mixture models to analyze customer behavior in commercial banking.
- Implemented the solution using Scala within a Spark Dataframe.

Anomaly Detection Engine with Bayesian Network

- Designed and developed an anomaly detection model using Bayesian networks and Hidden Markov Models (HMM) to reveal intricate patterns and anomalies in complex datasets.
- Implemented the proposed model in Java, covering data preprocessing, algorithm development, and pipeline integration.
- Ensured efficient and accurate anomaly detection through rigorous testing and optimization, enhancing data quality.

Complex Event Processing Engine for Real-Time Data Insights

- Gathered raw network log data from over 40 different sources, including SYSLOG servers.
- Performed data cleansing with custom Java-based REGEX scripts and enriched data using Redis as an in-memory data structure store
- Designed a Graph-based model for identifying connected components in network data using Gelly (Graph Processing with Apache Flink).
- Devised a multi-task Complex Event Processing (CEP) approach, incorporating predictive analytics based on graph processing for near-real-time data insights and responsive actions applied in the Banking and Insurance sectors.

Advancements in Histogram-Based Outlier Detection

- Developed a novel framework for utilizing an unsupervised distance-based algorithm based on histograms to capture outliers across different time periods.
- Conducted thorough data analysis to uncover patterns and trends within various time segments, leading to refinements and optimizations in outlier detection.

Semantic Search Solutions for Network Data

- Implemented a full-text search application by using Elasticsearch on Network data
- Developed custom data visualization dashboards to provide real-time insights into network activity and search results, improving data analysis capabilities for the team.

PROFESSIONAL DEVELOPMENT TRAINING AND CERTIFICATION

Independent Research-based Course titled "Natural Language Processing" Tehran, Iran
Amirkabir University of Technology Jan 2022

- Implemented multiple applications with Bert and the GPT-3 language model, including Tweet Emotion Recognition.
- Developed a Dialogue System for Restaurant Reservations using hybrid models.
- Conducted research aimed at enhancing the performance of open-domain chatbots through the use of retrieval methods.

Summer School Tehran, Iran
University of Tehran Sep 2018

- Fundamentals of Deep learning and Machine learning with Hands-on sessions

Specialization Certification
Coursera

- Natural Language Processing Specialization
- Deep Learning Specialization

Online Course
Stanford University

- **CS224N**: Natural Language Processing with Deep Learning
- **CS231N**: Deep Learning for Computer Vision
- **CS324**: Large Language Models
- **CS236**: Deep Generative Models
- **CS330**: Deep Multi-task and Meta Learning

University of California, Berkeley

- **CS285**: Deep Reinforcement Learning

TECHNICAL SKILLS

Programming Languages: Python(proficient), Java(Senior), Scala, MatLab, R
Deep learning tools: Pytorch, TensorFlow, Hugging Face Transformers, Diffusion Libraries, NLTK, Scikit-learn
Developer Tools: Git, Docker, Apache Spark, Apache Kafka, VS Code, Jupyter Lab, IntelliJ IDEA
Databse Management: MongoDB, Cassandra, Redis, MySQL, SparkSQL

HONORS AND AWARDS

Predoctoral Fellowship, Fatima Al-Fihri Predoctoral Fellowship 2023
CSDI Proposal Acceptance for Scalable ML Solution in Stock Market Analysis 2019
Proposal Approval for Innovative Anomaly Detection engine in the Iran Banking Sector 2017
3rd Rank in Artificial Intelligence MSc program, Amirkabir University of Technology 2014
Talented Student Office Admission for graduate study, Amirkabir University of Technology 2011
1st Rank in Computer Engineering department, Alzahra University, 2011

LANGUAGES

English : Toefl score: 103, R: 24, L: 27, S: 27, W: 25
Persian : Native

REFERENCES

Further information and Proofs are available upon Request