Maryam Hasan

Email: mhasan@wpi.edu, Web Page: https://maryhasan.github.io

Education

Ph.D. in Computer Science, Worcester Polytechnic Institute (WPI), Worcester, MA, USA, 2012 – 2021
Thesis title: Automatic Emotion Detection in Text Messages using Supervised Learning, GPA: 3.8/4.0

Master of Science in Computer Science, University of Alberta, Edmonton, Canada, 2008 – 2011
Thesis title: Extracting Structured Knowledge from Textual Data in Software Repositories, GPA: 3.7/4.0

Technical Skills

Programming Languages: Python, Java, C++, SQL, PHP
Development tools: IPython Notebook, Eclipse, RStudio, MySQL, MongoDB, NLP tools
Machine learning and Deep learning tools: Scikit-learn, NumPy, Pandas, TensorFlow, PyTorch

Professional Experiences

Research Assistant, Data Science Research Lab, Computer Science department, WPI, 2013 - Present
- Developed a hierarchical clustering model to detect topics on Twitter messages (*using Java and Apache Mahout*)
- Developed machine learning models to classify text messages (*using SVM, Naïve Bayes, Decision Tree*)
- Developed and evaluated transformer neural networks to learn emotion-specific embeddings by fine-tuning language models via transfer learning (*using Python: NumPy, Scikit-learn; TensorFlow: RNN, Bi-LSTM, BERT Transformers*)

Software Engineer Intern, Oracle Company, Marlborough, MA, June - September 2013
- Developed a program to add more information to log messages using *Java, Logback*
- Designed and developed unit test plans and *JUnit* test cases

Collaborating Researcher, InfoLab, CSAIL, Massachusetts Institute of Technology, February 2012- May 2012
- Collaborated in START project to extract information from Web for natural language question answering
- Developed a Web Scraper to extract symbols and synonyms from Persian movie websites (*using PHP*)
- Developed a program to extract information about the extracted symbols from movie websites (*using Scheme*)

Software Engineer, Multimedia Lab, University of Alberta, Canada, January - July 2011
- Collaborated in a team project to develop a web based application for testing high school students
- Developed the user interface using *Java and Google Web Toolkit*, and the server using *PHP and MySQL*

Research Assistant, Service Systems Research Lab, University of Alberta, Edmonton, May 2009- Dec 2010
- Designed and developed a tool to extract structured knowledge from textual data in software repositories (using *Java, XQuery, DB2, XML, Stanford NLP tools: Sentence Splitter, PoS tagger, Stanford Parser*)
- Extracted relations among named entities via Hierarchical Clustering of blogs (*using Java, MySQL, Stanford NLP*)
- Developed a program for feature selection in micro-array gene expression data via Bi-Clustering (*using MatLab*)

Software Engineer, Software Engineering Research Lab, University of Alberta, Edmonton, June 2008-Sep 2008
- Collaborated in Annoki project which is a Wiki tool for researchers
- Developed a graphical web interface for Annoki using *Adobe Flex, PHP and MySQL*

Projects

- **DeepEmotex**: A Deep Transfer Learning framework to Classify Emotion in Text Messages, 2021
    - Developed a deep transfer learning method to learn domain-specific features from context.
    - Fine-tuned pre-trained language models (e.g., BERT and USE) on the target classification task.
    - Trained a baseline neural network model (i.e., Bi-LSTM) and evaluated DeepEmotex models.
    - The proposed DeepEmotex-BERT model outperformed the baseline model by 23%.
    - Implemented using *Python: Scikit-learn, NumPy, Pandas, TensorFlow, PyTorch*
- **EmotexStream**, A two-stage framework to Discover Trends in Public Emotion in Text Streams, 2017.
    - Developed a binary classification model to classify text messages into emotion and no-emotion classes.

  - Developed and evaluated an online method to measure public emotion and detect temporal changes of emotion in a stream of messages during public events.
  - Used Hoeffding's inequality to define an upper bound on the probability that the sum of independent random variables deviates from its expected value. Implemented using *Java*.

- **Emotex**: A Machine Learning approach to Detect Emotion in Text Messages, 2014
  - Collected and processed large corpus of labeled messages for supervised learning of emotions in text.
  - Developed and evaluated machine learning models to classify text messages including Support Vector Machines (SVM), Naïve Bayes and Decision Tree.
  - Implemented using *Python (Scikit-learn, NumPy)* and *Java*

- **LinCoh**, A Feature Selection method using Linear Coherent Bi-Clustering and Sample Set Clustering, 2012.
  - Developed a method to find linear coherent bi-clusters in Gene Expression Microarray data via Beam Searching and Sample Set Clustering. Our method exploits a robust technique to identify conditionally correlated genes, combined with an efficient density-based search for clustering sample sets.
- **NLP4SE**: Analyzing Natural Language Artifacts of the Software Process, 2010
  - Designed and developed a tool to extract structured knowledge from textual data in software repositories
  - Implemented using Java, XQuery, DB2, XML, Stanford NLP tools: PoS tagger, Dependency Parser
  - Extracted relations among named entities via Hierarchical Clustering of blogs (using Java, Stanford NLP)

- **Annoki**, A Wiki tool for researchers, 2011.
  - Collaborated in Annoki project which is a social Wiki tool for researchers
  - Developed a graphical web interface for Annoki. Implemented using Adobe *Spring-Flex, PHP* and *MySQL*

Publications

- **Maryam Hasan**, Elke Rundensteiner, Emmanuel Agu, DeepEmotex: Classifying Emotion in Text Messages using Deep Transfer Learning, IEEE Big Data: Machine Learning on Big Data (IEEE BigData, MLBD), December 2021.
- **Maryam Hasan**, Elke Rundensteiner, Emmanuel Agu, Automatic Emotion Detection in Text Streams by Analyzing Twitter data, *International Journal of Data Science and Analytics*, Springer 2019.
- **Maryam Hasan**, Elke Rundensteiner, Xiangnan Kong, Using Social Sensing to Discover Trends in Public Emotion, In Proceedings of IEEE International Conference on Semantic Computing (IEEE ICSC), 2017.
- **Maryam Hasan**, Elke Rundensteiner, Emmanuel Agu, Using Hashtags as Labels for Supervised Learning of Emotions in Twitter Messages, ACM SIGKDD Workshop on Health Informatics (HI-KDD), August 2014.
- **Maryam Hasan**, Elke Rundensteiner, Emmanuel Agu, EMOTEX: Detecting Emotions in Twitter Messages, In Proceedings of ASE International Conference on Social Computing (SocialCom), May 2014.
- Di Yang, Kaiyu Zhao, **Maryam Hasan**, Hanyuan Lu, Elke Rundensteiner and Matthew Ward, Mining and Linking Patterns across Live Data Streams and Stream Archives, VLDB, October 2013.
- Yi Shi, **Maryam Hasan**, Zhipeng Cai, Guohui Lin and Dale Schuurmans, Linear Coherent Bi-clustering via Beam Searching and Sample Set Clustering, International Journal of Discrete Mathematics, Algorithms and Applications (DMAA). December 2011.
- **Maryam Hasan**, Eleni Stroulia, Denilson Barbosa and Manar Alalfi, Analyzing Natural-Language Artifacts of the Software Process, Early Research Achievement track of the 26th IEEE International Conference on Software Maintenance (ICSM2010), September 2010.
- Yi Shi, **Maryam Hasan**, Zhipeng Cai, Guohui Lin and Dale Schuurmans, Linear coherent bi-cluster discovery via beam detection and sample set clustering, International Conference on Combinatorial Optimization and Applications (COCOA2010), December 2010.