Mohamed Trabelsi

CONTACT INFORMATION

113 Research Drive (Building C)
Lehigh University
Bethlehem, PA 18015 USA
Phone: +1(610) 570-5833
E-mail: mot218@lehigh.edu
Homepage: https://linkedin.com/in/mohamed-trabelsi-682337107/
Github: https://github.com/medtray
Google Scholar: https://scholar.google.com/citations?user=FBXqOekAAAAJ

RESEARCH INTERESTS

Machine Learning, Information Retrieval, Natural Language Processing, Computer vision

EDUCATION

Lehigh University, Bethlehem, PA, USA **09/2018-Present**

Ph.D. student, Department of Computer Science & Engineering

- Advisor: Jeff Heflin
- Overall GPA: 4.0/4.0

University of Louisville, Louisville, KY, USA **08/2016-05/2018**

Master's student, computer engineering and computer science

- Advisor: Hichem Frigui
- Overall GPA: 3.938/4.0

Tunisia Polytechnic School, Marsa, Tunisia **09/2013-05/2016**

Engineering student

- General Rank: 1/45

Preparatory institution for engineering studies, Sfax, Tunisia **09/2011-05/2013**

Engineering student, Major Mathematics- Physics

- Rank in the national exam for engineering studies: 29/4000

EXPERIENCE

Research Assistant Lehigh University, Bethlehem, PA, USA **01/2022- present**

Intern Nokia Bell Labs, Murray Hill, NJ, USA **06/2021-12/2021**

- Research intern in Statistics and Data Science Research Group.
- Tickets classification: Research project focuses on using log data to classify trouble ticketing system regarding services and other issues in mobile networks.
- Propose a new domain adaptation-based method that transfers knowledge from multiple source domains to a target domain in entity matching task.
- Study the zero-shot learning case on the target domain for entity matching, and demonstrate that our method learns the entity matching task and transfers knowledge to the target domain.

Research Assistant Lehigh University, Bethlehem, PA, USA **08/2020-05/2021**

- Propose a new structure-aware BERT model, called StruBERT, that fuses the textual and structural information of a data table to produce context-aware representations for both textual and tabular content of a data table for table search and matching.

- Propose a new knowledge graph that incorporates both dataset-dependent and dataset-agnostic knowledge from table corpus for table search.
- Propose an ontology to help AI researchers keep track of the scholarly progress of AI related tasks such as natural language processing and computer vision.

Intern Nokia Bell Labs, Murray Hill, NJ, USA **06/2020-08/2020**

- Data science intern in Statistics and Data Science Research Group.
- Generating schema labels automatically for column values of data tables from multiple domains: Web and Log tables.
- Propose a new context-aware semantic labeling method using both the column values and context, which is based on the pre-trained contextualized language model, BERT.

Research Assistant Lehigh University, Bethlehem, PA, USA **06/2019-05/2020**

- Propose a new unsupervised ranking method for dataset search and retrieval using a new schema label embedding.
- Propose a new learning-to-rank architecture that improves dataset search and retrieval.

Research Assistant Lehigh University, Bethlehem, PA, USA **09/2018-05/2019**

- Explainable AI for deep learning models: Knowledge-based Machine Learning for Convolutional Neural Network model.
- Generate less opaque models using human knowledge.

Research Assistant University of Louisville, Louisville, KY, USA **08/2017-05/2018**

- Investigate Multiple Instance Learning (MIL) algorithms for landmine detection.
- Build MIL based algorithms with high performance (low false positive rate).

Research Assistant University of Louisville, Louisville, KY, USA **08/2016-07/2017**

- Propose a new algorithm for Multiple Instance Regression paradigm.
- Achieve high performance for remote sensing application and drug activity prediction using the proposed algorithm.

Intern University of Louisville (Multimedia Research Lab), KY, USA **03/2016-08/2016**

- Human Action Detection and Recognition in Video Data: Improve human action classification by iteratively filtering irrelevant centers of bag-of-words based approach.

Intern ORANGE Telecommunications, Tunis,Tunisia **07/2015-09/2015**

- Implementation of a server for the supervision of network equipment
- Recovery and save of the configurations with a configuration management database.

Publications

StruBERT: Structure-aware BERT for Table Search and Matching **2022**
***M. Trabelsi**, Z. Chen, S. Zhang, B.D. Davison and J. Heflin*
The Web Conference, 2022.

DAME: Domain Adaptation for Matching Entities **2022**
***M. Trabelsi**, J. Heflin and J. Cao*
15th ACM International Conference on Web Search and Data Mining (WSDM)

Neural Ranking Models for Document Retrieval **2021**
***M. Trabelsi**, Z. Chen, B.D. Davison and J. Heflin*
Information Retrieval Journal, Springer.

MGNETS: Multi-Graph Neural Networks for Table Search **2021**
*Z. Chen, **M. Trabelsi**, J. Heflin, D. Yin and B.D. Davison*
30th ACM International Conference on Information and Knowledge Management (CIKM).

SeLaB: Semantic Labeling with BERT **2021**
***M. Trabelsi**, J. Cao and J. Heflin*
2021 International Joint Conference on Neural Networks (IJCNN).

A Hybrid Deep Model for Learning to Rank Data Tables **2020**
***M. Trabelsi**, Z. Chen, B.D. Davison and J. Heflin*
2020 IEEE International Conference on Big Data (Big Data), December, 2020.

Relational Graph Embeddings for Table Retrieval **2020**
***M. Trabelsi**, Z. Chen, B.D. Davison and J. Heflin*
2020 IEEE International Conference on Big Data (Big Data), December, 2020.

Semantic Labeling Using a Deep Contextualized Language Model **2020**
M. Trabelsi, J. Cao and J. Heflin
arXiv preprint arXiv:2010.16037, October, 2020.

Towards Knowledge Acquisition of Metadata on AI Progress **2020**
*Z. Chen, **M. Trabelsi**, J. Heflin, Y. Xu and B.D. Davison*
ISWC (Posters & Demos), November, 2020.

Table Search Using a Deep Contextualized Language Model **2020**
*Z. Chen, **M. Trabelsi**, J. Heflin, Y. Xu and B.D. Davison*
In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020), Xi'An, China, July, 2020.

Improved Table Retrieval Using Multiple Context Embeddings for Attributes **2019**
***M. Trabelsi**, J. Heflin, and B.D. Davison*
2019 IEEE International Conference on Big Data (Big Data), December, 2019.

Robust fuzzy clustering for multiple instance regression **2019**
***M. Trabelsi** and Hichem Frigui*
Pattern Recognition, Elsevier.

Fuzzy and Possibilistic Clustering for Multiple Instance Linear Regression **2018**
***M. Trabelsi** and Hichem Frigui*
IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)

Comparison of several single and multiple instance learning methods for detecting buried explosive objects using GPR data **2018**
*Andrew Karem, **M. Trabelsi**, Mahdi Moalla and Hichem Frigui*
Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIII.

Submitted Papers

Semantic Labeling Using a Context-aware BERT Model **2021**
***M. Trabelsi**, J. Cao and J. Heflin*
Journal of Data and Knowledge Engineering, Elsevier.

Teaching

Teaching Assistant
Design and Analysis of Algorithms (CSE 340), Lehigh University, Fall 2020

Selected Projects

COVID-19 Lung Image Segmentation using Adversarial Network

- A Generative Adversarial Network-based model is developed for COVID-19 Lung CT Lesion Segmentation.
- The GAN model is composed of an Autoencoder-based generator and critic network.
- The generator accepts original CT slices and generates the predicted COVID-19 mask, which is compared to the groundtruth mask using the critic network with Multi-scale loss function.
- Our proposed model achieves accurate segmentation results for COVID-19 lung lesions.

Multi-Images one shot prediction attack

- An Autoencoder-based model is developed for multi-images one shot prediction.
- Mixing hidden representations of multiple encoded images, and predicting labels of decoded images to extract the label of each image.
- The Autoencoder is trained by minimizing the mixup loss of the target model (parameters of target model are unknown)
- Reduce computations in the target model, while conserving high accuracy for predictions
- 80% accuracy for one shot prediction of 4 images using CIFAR10 dataset.

Less distorted poisoning attack

- A Less Distorted Poisoning Attack: A Connection Between Evasion and Poisoning Attacks.
- Propose new objective function to iteratively search for poison image and adversarial image using forward-backward splitting.
- create a single attack that is successful and less-detectable than either poisoning or evasion

Hyper-parameters tuning with distributed systems

- Hyper-parameters selection of deep learning models with distributed training.
- Grid search hyper-parameters selection using distributed training and early pruning of a non-promising combination for Estimators of TensorFlow.
- Highly effective method that decreases the time needed to find the best parameters.

Cpprestsdk server for Twitter API

- Design and implementation of server that shows statistics about Twitter users and trends using Cpprestsdk and twitcurl.
- Efficient storage of statistics using NoSQL database MongoDB.

Computer Skills

- Languages: Python, C/C++, Matlab, java, Go, Android
- Machine learning libraries: TensorFlow, PyTorch, Keras
- IDEs: PyCharm, VS Code, Eclipse, MS Visual Studio
- Tools: Hadoop, MapReduce, Spark, AWS, Git, Docker
- Web tools: PHP, HTML, CSS, JavaScript
- Database: MySQL, NoSQL
- Operating Systems: Unix/Linux, Windows.

Languages

English, French, Arabic