Advanced ML, IDC

Final projects presentation – Spring 2021

Project structure, requirements, and time line

- A project consist of a subject, an anchor paper and innovative part
  - Anchor paper
    - Understand the paper
    - Find relevant related studies (travel the river of citations)
      - The basic (foundation) work
      - Most recent studies – learn what is interesting, don't reinvent an existing result
    - Implementation
      - Reconstruct (validate) the paper's results
      - Apply the paper's method on a related task – check robustness

  - Innovative part (one of the following)
    - Improve - Improve the existing method
    - Compete - Suggest an alternative approach
    - Inspire - Apply in a different domain / for a different task

- Time line (due by …)
  - Select a project – 17.06.2021
  - Project proposal approval – 01.07.2021
  - Anchor paper missions – 01.08.2021
  - Innovative part & Final report – 05.09.2021

Project proposals - Avrahami

Index	Paper Name	Subject	Link	Innovation ideas	Comments
1	STANCY: Stance Classification Based on Consistency Cues	Stance detection	https://www.aclweb.org/anthology/D19-1675.pdf	1. Improve results with newer BERT models 2. Use the method over other datasets (communities?) to establish a measure over user pairs/whole community	EMNLP 2019
2	Neural Coreference Resolution for Arabic	Coreference Resolution	https://arxiv.org/pdf/2011.00286.pdf https://github.com/juntaoy/aracoref	1. Use advanced BERT models and not only ARABert	Coling 2020
3	Word embeddings quantify 100 years of gender and ethnic stereotypes	Word embeddings + biases	https://www.pnas.org/content/pnas/115/16/E3635.full.pdf Git is available	1. Design a method to test biases in other languages (rather than English) 2. Check the method over different communities datasets	PNAS 2018
4	tBERT: Topic Models and BERT Joining Forces for Semantic Similarity Detection	Word embeddings + topic models	https://www.aclweb.org/anthology/2020.acl-main.630.pdf https://github.com/wuningxi/tBERT	1. Using the method suggested to compare between documents / communities 2. Try it over Hebrew/Arabic 3. Use a more sophisticated topic models (taken from their article)	ACL 2020
5	Sentiment after Translation: A Case-Study on Arabic Social Media Posts	Sentiment analysis + translation	https://www.aclweb.org/anthology/N15-1078.pdf	1. Build an ability to estimate how well the proposed method in the paper will work on a given tweet	NACL 2015
6	Sentiment and Emotion help Sarcasm? A Multi-task Learning Framework for Multi-Modal Sarcasm, Sentiment and Emotion Analysis	Sentiment, Sarcasm	https://www.aclweb.org/anthology/2020.acl-main.401.pdf	1. Test it over another dataset, improve results over another dataset 2. Arabic?	ACL 2020
7	Dialog Based Relation Extraction		https://arxiv.org/pdf/2004.08056.pdf https://github.com/nlpdata/dialogre		2020

Project proposals - Yuval

Index	Paper Name	Subject	Link	Innovation ideas	Comments
1	Grammar Variational Autoencoder.	Variational Autoencoder, Parse Trees, Grammar	https://drive.google.com/file/d/1p-d15gszQTEgtyZmlEdgG0tl0s2V9g0z/view?usp=sharing	1. Use another grammar.	They propose a variational autoencoder which directly encodes from and decodes to parse trees, ensuring the generated outputs are always syntactically valid.
2	A Hierarchical Neural Autoencoder for Paragraphs and Documents.	LSTM auto-encoder	https://drive.google.com/file/d/1LU12M8mM0fWKXWV1vUECm-IQOtXs-IO7/view?usp=sharing	1. Use another language.	Training an LSTM auto-encoder to preserve and reconstruct multi-sentence paragraphs.
3	Deep Variational Information Bottleneck.	Variational Inference	https://arxiv.org/abs/1612.00410	1. Discriminative VAE. 2. Learning smooth representation with side information.	

Project proposals – Shai

Index	Paper Name	Subject	Link	Innovation ideas	Comments
1	Time-series clustering – A decade review • Aghabozorgi, Shirkhorshidi, Wah	Time series clustering	https://wiki.smu.edu.sg/18191isss608g1/img_auth.php/f/fd/Time_Series_Clustering_A_Decade_Review.pdf	1. Compression-based methods 2. Model-based – RNN 3. Hierarchical (multi-resolution) modelling	• Genomic data • Disease Progression • Financial data
2	GP-VAE - Deep Probabilistic Time Series Imputation • Fortuin, Baranchuk, Rätsch, Mandt	Time series – handling missing values	https://arxiv.org/pdf/1907.04155.pdf	1. Incorporate additional generative models to impute 2. Integrate with the paper: "MIWAE - Deep Generative Modelling and Imputation of Incomplete Data Sets" • Mattei, Frellsen	• Need to understand Gaussian Processes
3	Order matters - Sequence to sequence for sets • Vinyals, Bengio, Kudlur	Embeddings of sets	https://arxiv.org/abs/1511.06391	Integrate with graph neural networks (GNN) • Applications in healthcare, bioinformatics and drug discovery	
4	Information Theoretic Analysis of DNN-HMM Acoustic Modeling • Dighe, Asaei, Bourlard	Hybrid HMM-DNN	https://arxiv.org/abs/1709.01144	A non-trivial combination of generative (HMM) and discriminative (DNN) models. Innovation can tackle • Suggest a different approximation to $P(x|q)$ • A different approach – DNN "master", HMM "slave"	• Speech-based Emotion Recognition • Word spotting in handwritten documents
5	WiFi CSI based passive human activity recognition using attention based BLSTM • Chen , Zhang , Jiang , Cao, Cui AttnSense: Multi-level Attention Mechanism For Multimodal Human Activity Recognition • Ma, Li, Zhang, Gao, Lu	Attention-based Human Activity Recognition (HAR)	https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8514811 https://www.ijcai.org/Proceedings/2019/0431.pdf	Relatively new field. Innovation can tackle • Adapt additional advanced NLP methods (transformers?) HAR domain • Improve the models to better match HAR assumptions • Multimodality	HAR Datasets: • Extrasensory dataset • WISDM dataset Clinical data • Sleep-wake: Mesa dataset • Depression: Proprietary

6	Algorithms and Theory for Multiple-Source Adaptation
	• Hoffman, Mohri, Zhang
		Domain Adaptation	https://papers.nips.cc/paper/8046-algorithms-and-theory-for-multiple-source-adaptation.pdf	1. Information theoretic learning
				   • Unsupervised Domain Adaptation
				   • Supervised, DNN
				2. Discriminative Active Learning for Domain Adaptation
				   • Zhou, Shui , Huang, Wang, Chaib-draa	