Dow Meeting 05/10 Progress Update

Andrew Simon, Shijie Zhang, Meenal Rawlani

Contents

- Progress update
    - Unified models under one downstream task
    - Incorporated new models into ensemble
    - Added concatenated ensemble and will begin meta embedding
    - Updated results for each model, expanded extrinsic metrics
- Discussion
    - Loss of testing accuracy, need for fine tuning models
    - Restructuring models to train on chemistry and engineering abstracts
    - Establishing GPU pipeline to quickly train new datasets on contextual models
- Looking Ahead
    - Model fine tuning to increase testing accuracy
    - Hyperparameter optimization of meta embedding
    - (if necessary) intrinsic evaluations

Unified Tensorflow Model & New Models

- Previously, models were trained on a variety of ML algorithms
  - Sklearn, pytorch, tensorflow, etc…
- Embeddings now pass through a uniform tensorflow model
  - tensor dimensions (labels, padded sentence length, embedding dimensions)
- Model becomes more generalized, resulting in accuracy loss
- Looking to improve model accuracy over next two weeks
- Incorporated Bag of Words and TF-IDF as low level models for the ensemble
- Added CoVe and FastText to incorporate more contextual embedding algorithms

Ensemble embeddings

- With a uniform ML model and embedding structure, the concatenated ensemble was able to be implemented as an independent NLP embedding model
  - Took average of the embedding vectors, looking to increase to a fourth degree tensor in the future
- Despite literature review, concatenated ensemble underperformed when compared to other methods

Results

	BoW	TF-IDF	W2V	GLoVE	BeRT	ELMo*	Fast Text	CoVe	Concat*
Acc			0.645	0.638	0.502	0.56			0.570
Acc.(lit)			~0.65	~0.750	~0.750	~0.850			?
Recall			0.586	0.578	1.000	0.273			0.563
Precision			0.666	0.659	0.5017	0.50			0.551
F1			0.312	0.308	0.334	0.177			0.278
AUC			0.762	.6872	0.500	0.6153			0.506

Results (n=500 samples)

	BoW	TF-IDF	W2V	GLoVE	BeRT	ELMo	Fast Text	CoVe	Concat
Acc			0.58	0.53	0.52				0.570
Acc.(lit)									?
Recall			0.417	0.354	0				0.563
Precision			0.589	0.548	0				0.551
F1			0.244	0.215	0				0.278
AUC			0.519	0.499	0.500				0.506

Discussion

- Accuracy of models decreased transitioning to a uniform model
  - Facilitates the need for further hyperparameter tuning
- Contextual embeddings are costly
  - Required over one hour of CPU time to extract 500 pre trained embeddings for sentiment analysis
  - Consider transitioning to lower level models depending on results
- Now have access to multi gpu support, training and embedding for contextual models should be < 1 hour
- Access to new dataset containing over 1 million chemistry abstracts, train models on this dataset in the future
- Concatenated ensemble will be costly to train and embed, recommend removing it in favor of meta embedding and singular algorithm selection

Looking Forward

- Rigorous review of tensorflow methods in NLP, looking to increase testing accuracy to comparable literature values
- Further optimization of meta embedding, pursuing more advanced algorithm selection methods
- Possibly Incorporate intrinsic methods, however literature and experimental results look promising as is
- Finalization of Github Repository
- Incorporation of unit tests for concatenated and meta ensembles
- Tensorflow projected models to identify key differences models not captured by extrinsic methods