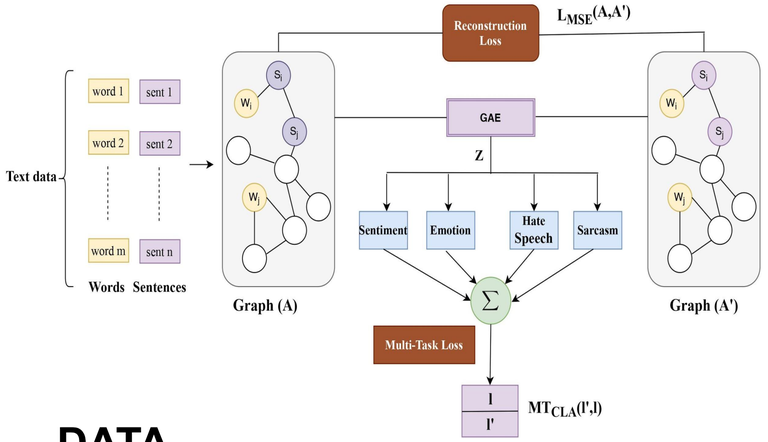
INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD
R&D SHOWCASE 2021
Technology, Social Impact

Multi-Task Text Classification using Graph Convolutional Neural Networks for Resource-Poor Language

ABSTRACT

- Many resource-poor languages fail to reap the benefits of recent developments of deep feature representations in Natural Language Processing (NLP) due to the lack of abundant annotated corpus.
- We create an annotated dataset for multiple NLP tasks such as Sentiment Analysis, Emotion Identification, Hate-Speech Detection, and Sarcasm Detection for Telugu language.
- We propose to use a Multi Task Graph Convolutional Neural Networks (MT-Text GCN) on Telugu to: (i) learn word embedding representation generated from reconstructed graph encoder, (ii) learn sentence representation for multi-task text classification for the above mentioned tasks.

METHOD



DATA

Task	# Total instances	# Classes
Sentiment	16,234	2
Emotion	9,582	4
Hate-speech	35,142	2
Sarcasm	35,142	2

RESULTS

Tasks→	SA			EI			HS			SAR		
Feature set↓	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BoW	0.73	0.69	0.70	0.48	0.38	0.39	0.57	0.51	0.51	0.49	0.50	0.50
TF-IDF	0.74	0.70	0.70	0.48	0.38	0.40	0.61	0.52	0.53	0.50	0.50	0.50
Word2Vec	0.80	0.79	0.79	0.55	0.43	0.44	0.62	0.51	0.52	0.49	0.50	0.50
GloVe	0.80	0.79	0.79	0.49	0.42	0.43	0.54	0.51	0.51	0.59	0.52	0.52
DeepWalk	0.57	0.56	0.56	0.31	0.31	0.30	0.49	0.50	0.50	0.49	0.50	0.50
Node2Vec WS2	0.77	0.77	0.77	0.50	0.47	0.48	0.54	0.51	0.52	0.62	0.53	0.54
Node2Vec WS3	0.78	0.77	0.78	0.48	0.46	0.47	0.59	0.53	0.55	0.60	0.56	0.55
Node2Vec WS4	0.78	0.78	0.78	0.47	0.46	0.46	0.61	0.53	0.54	**0.63**	**0.56**	**0.58**
MT-GCN	**0.83**	**0.83**	**0.83**	**0.58**	**0.48**	**0.50**	**0.61**	**0.54**	**0.55**	0.61	0.56	0.56

P = Precision, R = Recall, F1 = F1-score

SA: Sentiment Analysis, **EI**: Emotion Identification, **HS**: Hate-speech, **SAR**: Sarcasm

CONCLUSION

1. We formulate the problem as a graph to learn word and sentence embeddings, and perform multi-task text classification jointly with GCN.
2. We are the first to create a GCN model on Telugu language and our MT-Text GCN outperforms the earlier methods.

Authors: Mounika Marreddy, Lakshmi Sireesha Vakada, Venkata Charan Chinni, Subbareddy Oota, Radhika Mamidi.

NIPS-WiML

Research Center Name: LTRC