



## Dataset Creation and Evaluation of Aspect Based Sentiment Analysis in Telugu

### ABSTRACTS

In recent years, sentiment analysis has gained popularity as it is essential to moderate and analyse the information across the internet. It has various applications like opinion mining, social media monitoring, and market research. Aspect Based Sentiment Analysis (ABSA) is an area of sentiment analysis which deals with the sentiment at a finer level. ABSA classifies sentiment with respect to each aspect to gain greater insights into the sentiment expressed. Significant contributions have been made in ABSA, but this progress is limited only to a few languages with adequate resources. Telugu lags behind in this area of research despite being one of the most spoken languages in India and an enormous amount of data being created each day. In this paper, we create a reliable resource for aspect based sentiment analysis in Telugu. The data is annotated for three tasks namely Aspect Term Extraction, Aspect Polarity Classification and Aspect Categorisation. Further, we develop baselines for the tasks using deep learning methods demonstrating the reliability and usefulness of the resource.

### Data Creation and Annotation

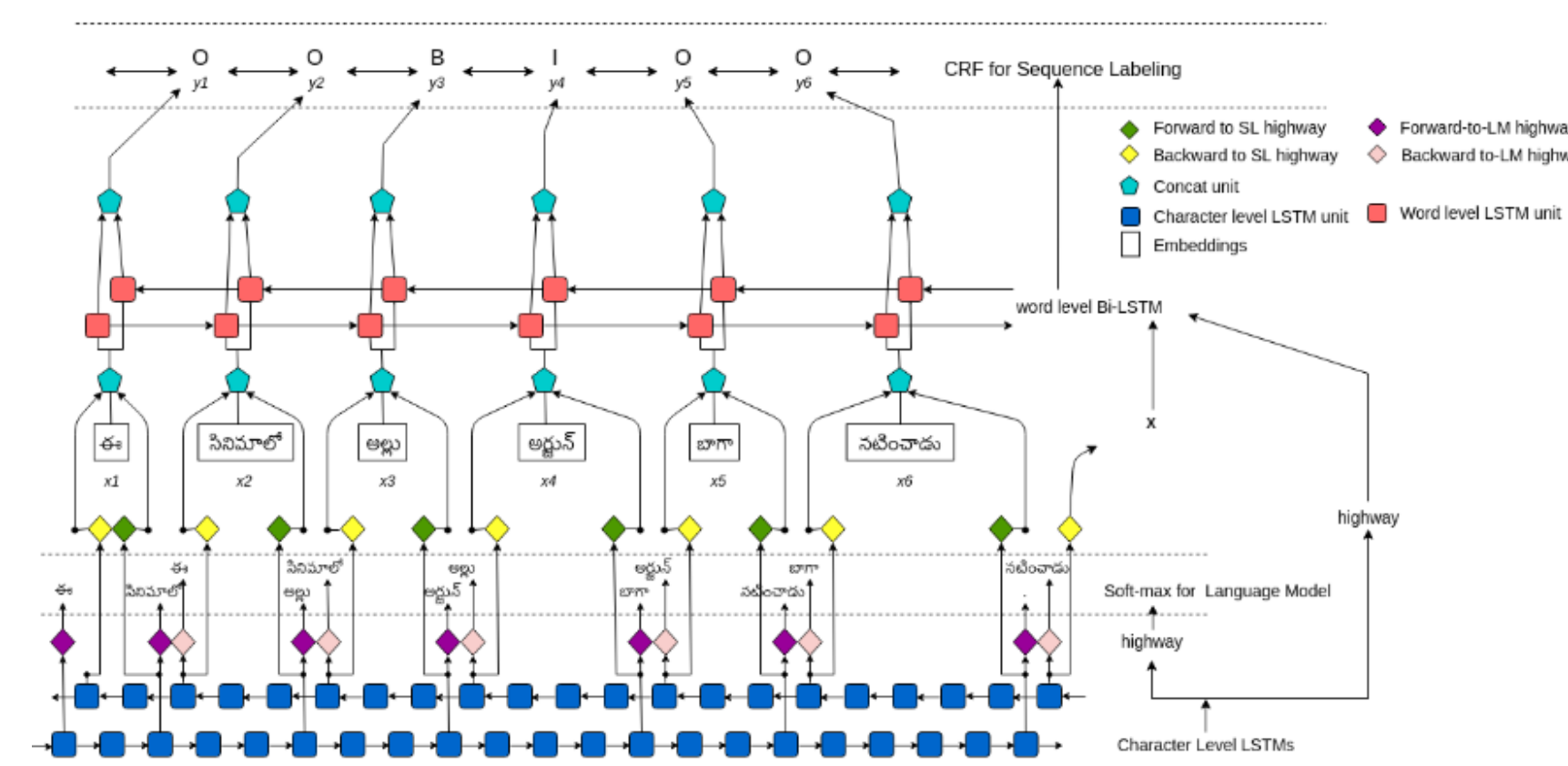
- We crawled several movie review websites such as 123telugu.com, eenadu.net, telugu.samayam.com. Initially, there were 10000 sentences from the scraped data. There were 5027 review sentences after the pre-processing steps.
- Annotation:
  - Identifying the aspect terms in each sentence
  - Assigning polarity to each aspect term, either positive, negative or neutral
  - Categorising the aspect term into one of the six categories, viz. story, acting, direction, music, technical and general.

	Review Sentences	Aspect Terms	Polarity	Category
Telugu Script	ఈ సినిమా లో లోకేష్లు అందంగా ఉన్నారు.	లోకేష్లు	positive	general
Transliterated	ee cinema lo locationlu andanga unnayi.	(locationlu)		
English	In this movie, locations are good.			

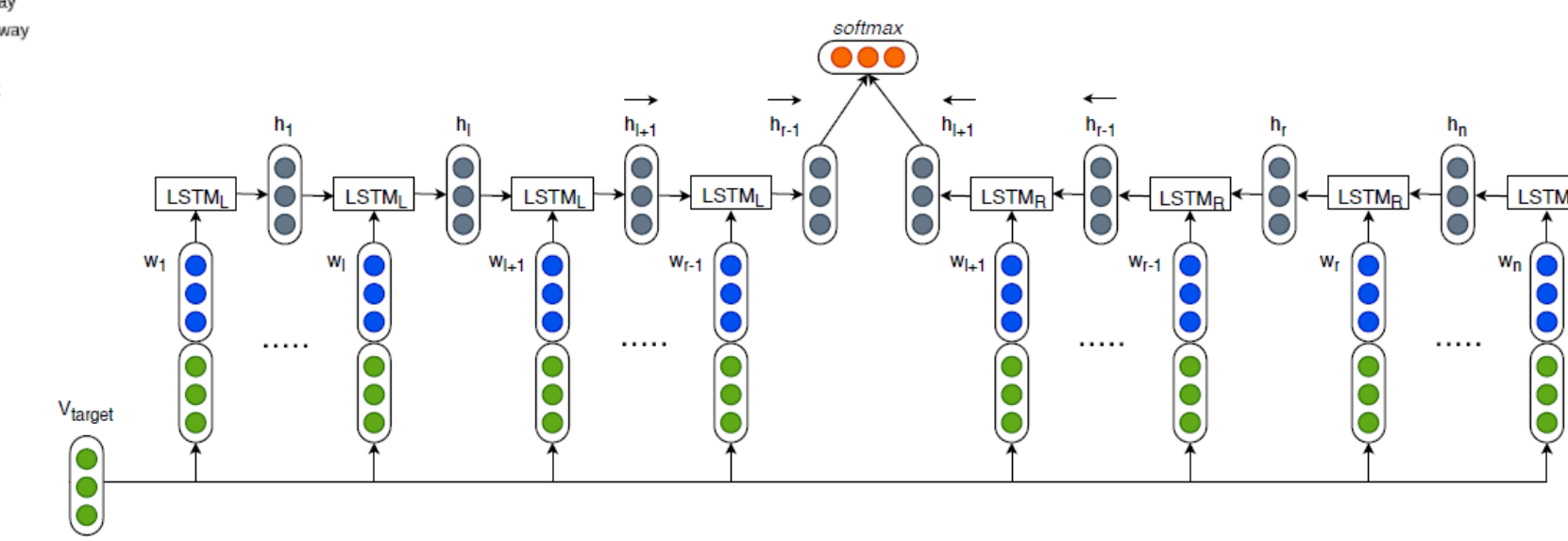
O	O	O	B	I	O	O	O
ఈ	సినిమా	లో	అల్లు	అర్జున్	బాగా	నటిం	చాడు .
ee	cinema	lo	allu	arjun	bAgA	naTinchADu	.
In this movie, Allu Arjun acted well.							

#Sentences		5027
#tokens		92848
Aspect Terms	#positive	3521
	#negative	2480
	#neutral	1129
	total	7130
Aspect Categories	#Story	548
	#Action	603
	#Direction	301
	#Music	382
	#Technical	554
	#General	4742
total		7130

### Method



(a) LM-LSTM-CRF architecture for Aspect Term Identification



(b) TC-LSTM model architecture for Aspect Polarity Classification and Categorisation

### Results

Methods	Precision	Recall	F1 score
LSTM + CRF	74.6%	69.2%	70.7%
bi-LSTM + CRF	79.3%	74.9%	75.8%
LM-LSTM-CRF + random Embeddings	81.3%	77.4%	77.7%
LM-LSTM-CRF + Pre-trained Telugu word2vec	82.3%	83.0%	81.5%
LM-LSTM-CRF + BPEmb	84.1%	82.6%	82.4%
LM-LSTM-CRF + Fasttext Embeddings	84.4%	84.2%	83.1%

(a) Aspect Term Identification

Methods	Precision	Recall	F1 score	Accuracy
SVM + TFIDF + Unigrams	70.8%	60.81%	65.42%	60.81%
SVM + TFIDF + Bigrams	69.81%	63.62%	66.57%	64.7%
Naive Bayes	62.16%	41.30%	46.40%	41.3%
LSTM	73.34%	67.91%	68.92%	74.79%
TD-LSTM	72.99%	73.77%	72.82%	76.33%
TC-LSTM	74.54%	72.58%	73.36%	79.71%
ATAE-LSTM	73.79%	68.36%	69.83%	75.91%
IAN	70.60%	68.96%	69.65%	74.93%
Deep Memory Networks	72.21%	66.19%	67.49%	74.37%

(b) Aspect Polarity Classification

Methods	Precision	Recall	F1 score	Accuracy
SVM + TFIDF + Unigrams	56.72%	51.31%	51.04%	51.36%
SVM + TFIDF + Bigrams	58.74%	52.33%	52.46%	52.33%
Naive Bayes	56.31%	51.9%	52.22%	51.9%
LSTM	69.9%	67.16%	68.18%	73.88%
TD-LSTM	70.34%	70.06%	66.92%	77.32%
TC-LSTM	74.65%	72.04%	72.32%	79.68%
ATAE-LSTM	68.05%	69.69%	68.61%	73.03%
IAN	67.66%	67.92%	67.78%	73.53%
Deep Memory Networks	71.50%	68.49%	69.58%	74.72%

(c) Aspect Categorisation