

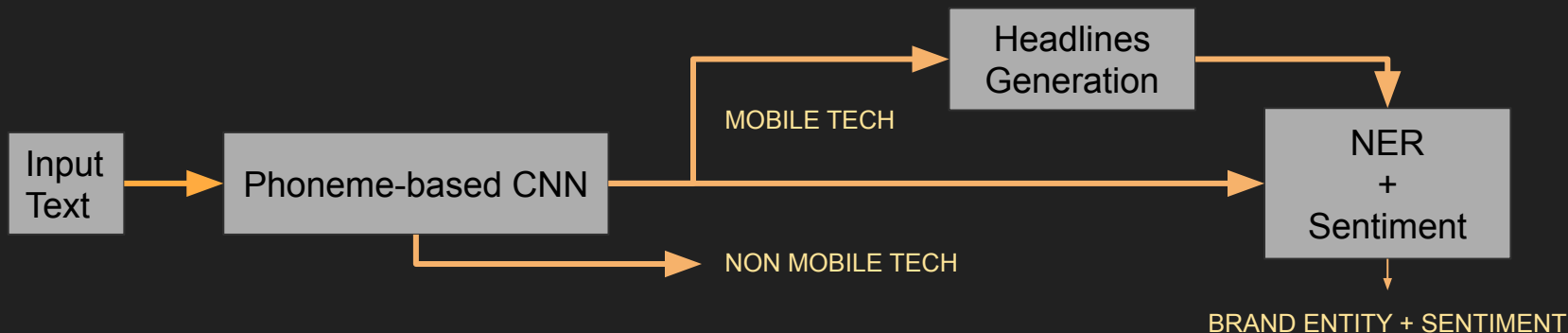
Bridgei2i's

Automatic Headline and Sentiment Generator

Our Solution

We divide our solution for the problem into three sections:

- Text Classification - Phoneme based CNN
- Headline Generation for Articles - Fine-tuned mBART with data augmentation
- Entity Recognition and Sentiment Assigning - Aspect Based Sentiment Analysis with a Rule Based Approach

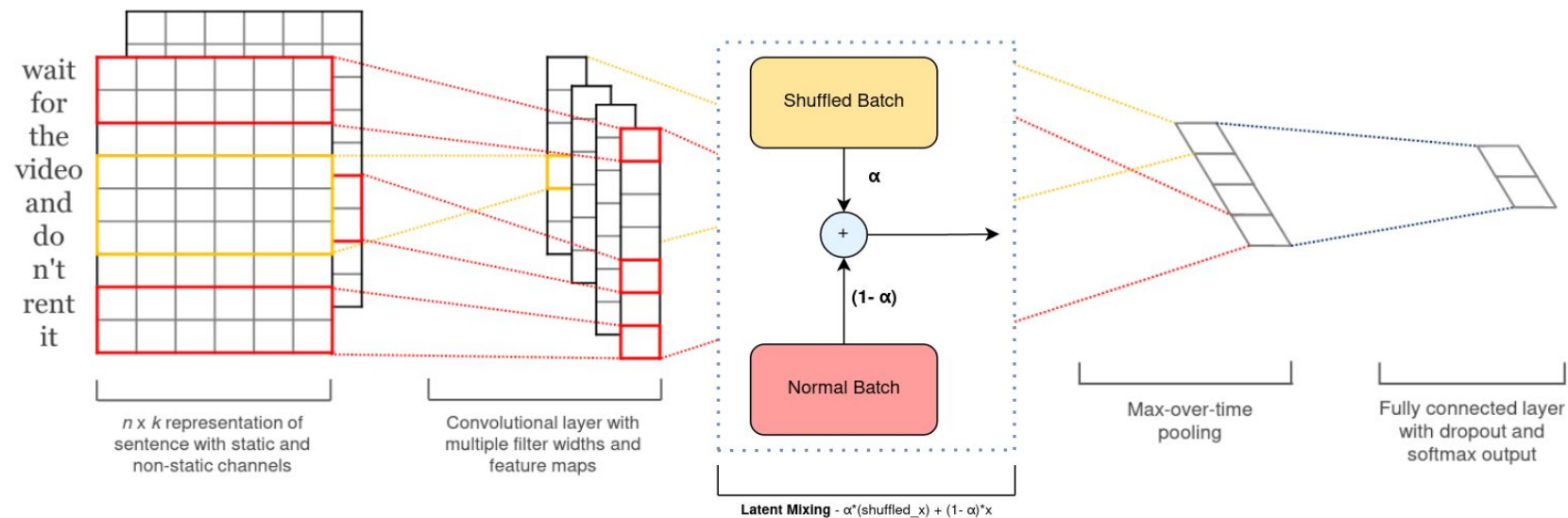


Preprocessing

- **Word Expansion** - “*aren’t*” to “*are not*”, “*you’ve*” to “*you have*” etc.
- **Basic Cleaning:**
 - Remove tags/mentions
 - Remove hyperlinks + html tags
 - Remove hashtags but retain text
 - Remove retweets (RT), quote tweet (QT)
- **Remove Emojis**
- **Remove Stopwords** (with some probability)

Theme identification using Text Classification

- Utilises **Phonemes** - smallest unit of distinct sound; *a unified representation across multiple languages*.
- Encoder based on CNN model (**Kim-CNN variant**)
- **Small size, Faster Inference speeds** compared to other SOTA
- Introduced learnable augmentation (Latent Mix) scheme to **increase variance** in samples, induce **better generalization capacity**



RESULTS:

Data Type	Train Accuracy	Validation Accuracy
Articles	99.97%	92.2%
Tweets	96.12%	93.72%

Results

of Params: 0.94 M

Inference Speed* (in secs):

0.0019 (phoneme conversion) + 0.0015 (model inference) for a seq len of 550 phonemes.

Sample outputs

RT @aajtak: 20 हजार के अंदर 5G स्मार्टफोन्स आ चुके हैं, क्या इससे भी सस्ता 5G स्मार्टफोन लाएगा Nokia? पढ़ें ये रिपोर्ट.. #Smartphones
#Nokia <https://t.co/of7KIH3IKS>

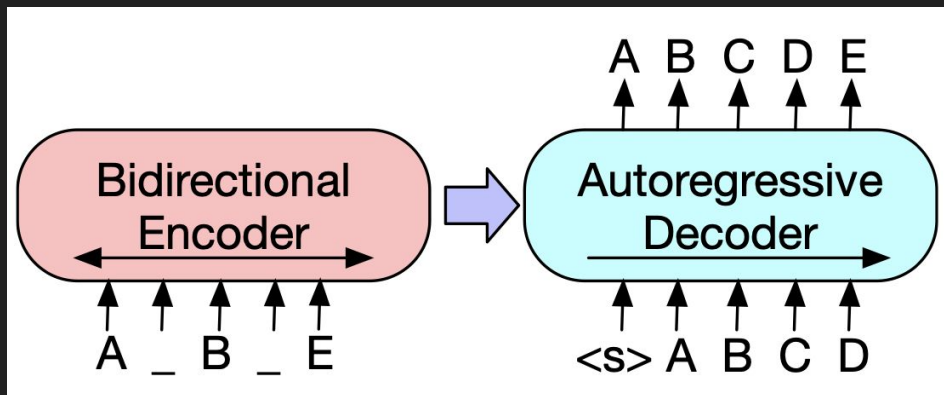
Predicted probabilities: [1.6838817e-25, 1.0000000e+00] (**predicted theme: Mobile Tech**)

February is American Heart Month! ❤️ Celebrate with Raleigh Parks by taking time to show your heart some love! We'll be sharing tips all month to improve your overall health and lessen your risk for heart disease! <https://t.co/HwzyaGTlgy> #HeartMonth
#RaleighParks <https://t.co/xkiBzP6DND>

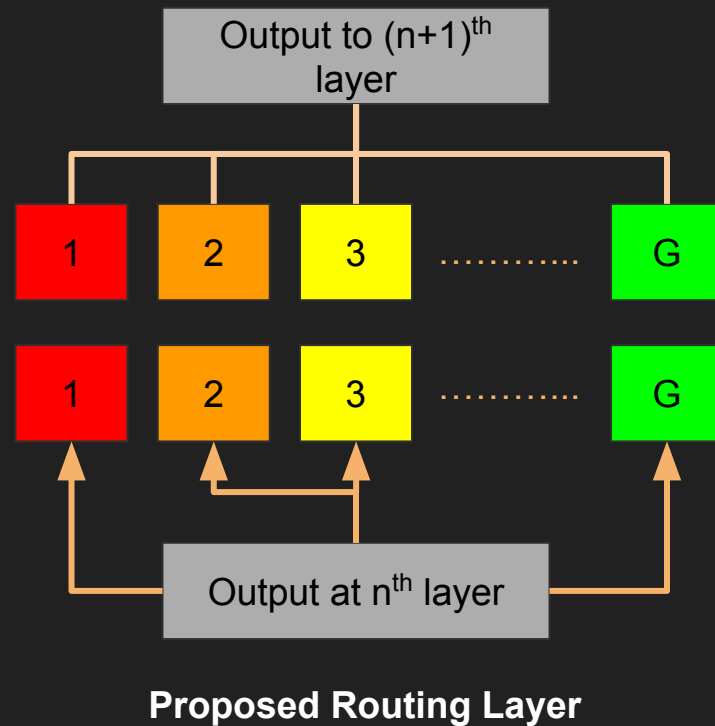
Predicted probabilities: [9.999875e-01, 1.3529364e-06] (**predicted theme: Non-Mobile Tech**)

Headline Generation

- Encoder-Decoder model based on **mBart** fine tuned on the given dataset.
- Data size is increased by **augmenting the text** (randomly replacing some words with their synonyms).
- We propose to introduce a **Routing-based layer** for modelling suitable interdependencies between languages. It can also be beneficial for low-resource languages.



mBART is an encoder-decoder model which was pretrained with the objective of denoising multiple languages simultaneously.



RESULTS:

Data Type	Rouge-L	BLUE Score	Headline similarity
Validation samples with Mobile_Tech_Flag = 1	Precision = 0.46 Recall = 0.49 F1 Score = 0.52	23.46	0.75
All validation samples	Precision = 0.39 Recall = 0.43 F1 Score = 0.42	20.53	0.73
Training samples	Precision = 0.45 Recall = 0.49 F1 Score = 0.47	28.23	0.75

of Params: 610.85 M

Inference Time*(in secs): 0.82

Sample outputs

ID: article_0176

Predicted headline: Global Vertical Farming Market to Increase Demand for Organic Food in the Forecast Period 2021-2026

ID: article_0179

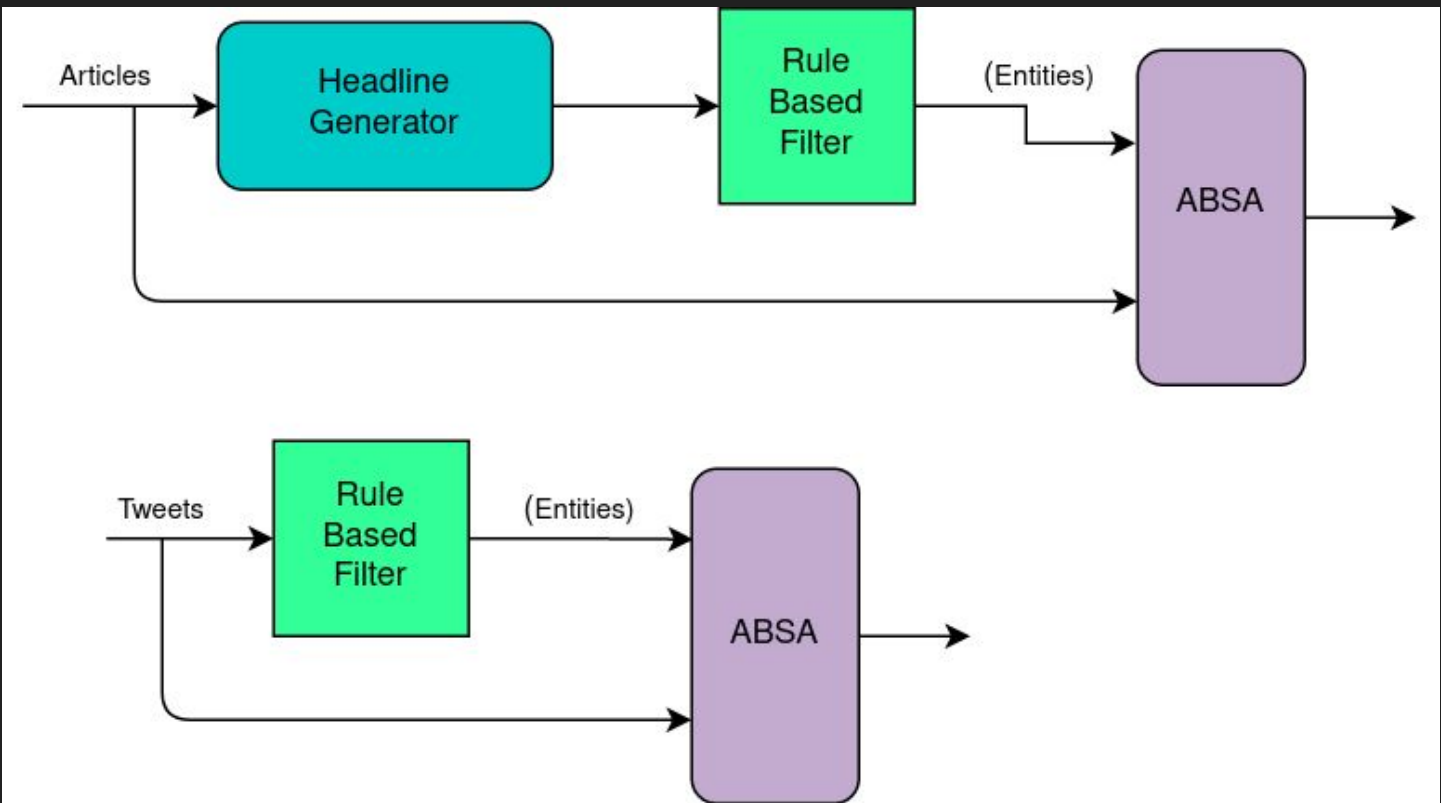
Predicted headline: ISL 7: Jamshedpur FC puts winless run to bed with spectacular Mobashir's goal

ID: article_1605

Predicted headline: Samsung Galaxy M02, Galaxy A32 and Galaxy A32 Pro 5G can be purchased online at Rs. 10,000.

Entity Recognition and Sentiment Analysis

- Identify entities using NER, cross verified with exhaustive list of mobile companies.
- Aspect Based Sentiment Analysis (ABSA) is used to predict the corresponding sentiment for each entity.
- In the case of articles, the summarized headlines are used to identify the entities and their sentiments.



RESULTS:

- 1) **Tweet:** '#Samsung is now making another addition into it's A series, and the word is out that it's upcoming variant Galaxy A72 4G has awesome battery and charger with brilliant specs . . . For more tech updates Follow 🖱️ @MobileNerve . . . #smartphone #mobile #mobilephone #mobilephoto <https://t.co/b8UO2Z6mC4>'
Predicted Entity: ['samsung'] **Predicted Sentiment:** ['Positive']
- 2) **Tweet :** 'Vivo X60 5G: The snapdragon 888 is coming to India in March or April. Check its full specs, price, pros & cons. <https://t.co/fRMzGq4Glt> #Vivo #VivoX60ProPlus #vivoY51s #VivoY31 #VivoX60 #VivoX60Pro #X60 #X60Pro #X60ProPlus #vivov20 #vivov20se #VivoIndia #Smartphone #thewidgetguy'
Predicted Entity: ['Vivo'] **Predicted Sentiment:** ['Neutral']
- 3) **Tweet:** 'RT @RethinkWireless: LG may exit smartphone market after 23 quarters of losses...<https://t.co/PwCJnJIsY7> #Wireless #5G #wifi #wifi7 #cellular #operators #spectrum #RAN #rethinktechnology #cloud @LGUS <https://t.co/ztKZhmRzzH>'
Predicted Entity: ['LG'] **Predicted Sentiment:** ['Negative']