



# **DETECTING POLITICAL BIAS IN NEWS ARTICAL USING HEADLINE ATTENTION**



**統碩一 112354032 王冀鋼**



# 實驗簡介



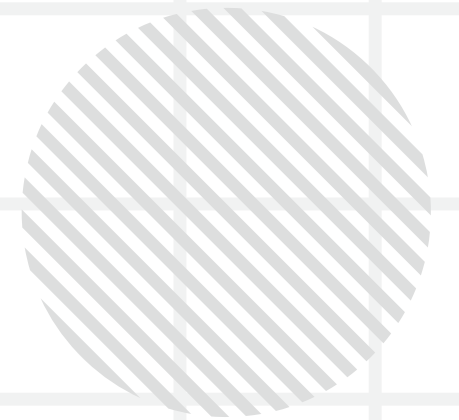
# 簡介

- 報章雜誌可能因為政治立場而有所偏頗
- 建立模型分類文章所偏頗的政黨
- 泰盧固文做為分析語言
- 印度的 Telangana、Andhra Pradesh 兩個邦
- 人工分類文章政黨偏向並作為資料集





# 資料介紹



# 資料介紹



- 蒐集 1329 篇泰盧固日報 (Telugu Newspaper)
- 分成BJP、TDP、Congress、TRS、YCP 和 None 六大類
- 資料種類分成 only headline、only article、concatenation of headline and article 三種型態

# 資料介紹

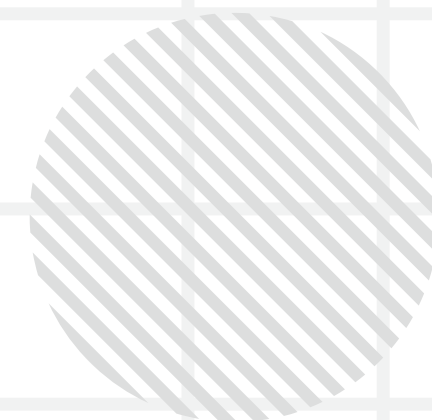


<b>Parties</b>	<b>Documents</b>	<b>Sentences</b>	<b>Words</b>	<b>avg #w in headline</b>	<b>avg #w in article</b>
BJP	182	2244	24863	4.13	132.48
Congress	82	1031	11410	4.06	135.08
TRS	151	1860	21685	4.09	139.52
TDP	361	3484	40495	3.86	108.3
YCP	335	1958	22370	3.79	62.98
Unbiased	218	1546	19245	4.09	65.14
Total	1329	12123	140068	3.95	98.3

Table 1: Data statistics: #w denotes the number of words per document



# 模型介紹



# HEADLINE ATTENTION NETWORK

- headline encoder
- article encoder
- headline attention layer
- bias detection

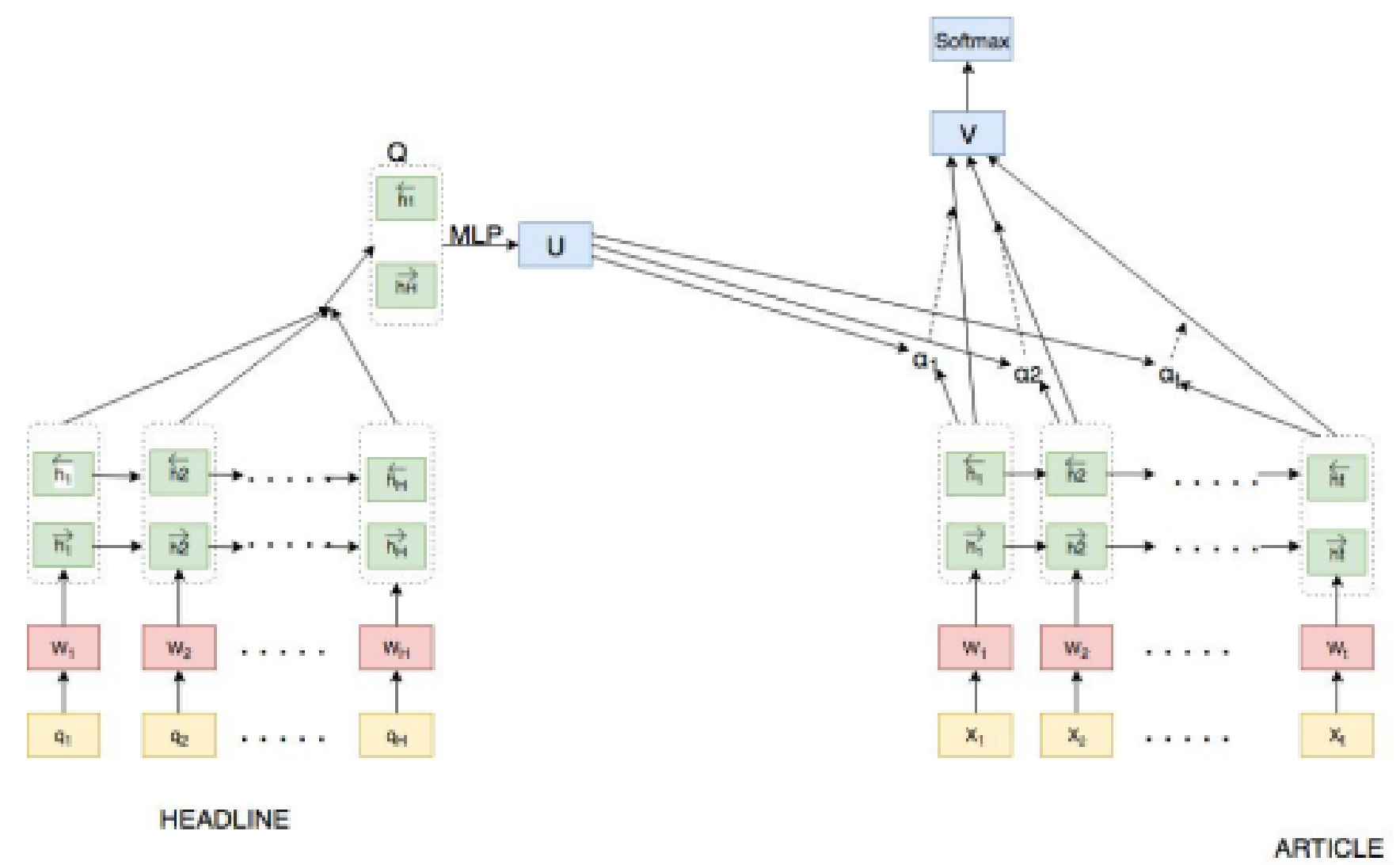


Figure 2: Headline Attention Network



# HEADLINE ATTENTION NETWORK

- headline encoder
  - 將標題的單字轉換成向量形態
  - 將轉換後的向量放進  
bidirectional LSTM (雙向長短記憶網路)

$$x_i = W_e q_i, i \in [1, H] \quad (1)$$

$$\vec{h}_i = \overrightarrow{LSTM}(x_i) i \in [1, H] \quad (2)$$

$$\overleftarrow{h}_i = \overleftarrow{LSTM}(x_i) i \in [H, 1] \quad (3)$$

# HEADLINE ATTENTION NETWORK

- article encoder
  - 將文章內的單字轉換成向量形態
  - 將轉換後的向量放進 bidirectional LSTM

$$x_i = W_e w_i, i \in [1, T] \quad (4)$$

$$\vec{h}_i = \overrightarrow{LSTM}(x_i), i \in [1, T] \quad (5)$$

$$\overleftarrow{h}_i = \overleftarrow{LSTM}(x_i), i \in [T, 1] \quad (6)$$

# HEADLINE ATTENTION NETWORK

- headline attention layer
  - 將 article encoder 的輸出結果依重要性作加權

$$u_i = \tanh(W_w h_i + b_w) \quad (7)$$

$$\alpha_i = \frac{\exp(u_i^T \cdot U)}{\sum_i \exp(u_i^T \cdot U)} \quad (8)$$

$$v = \sum_i \alpha_i h_i \quad (9)$$

# HEADLINE ATTENTION NETWORK

- bias detection
  - 將 headline encoder 的輸出結果和加權後的 article encoder 的輸出結果放進 Softmax function 處理
  - 計算 loss function 並判斷偏向的政黨

$$p = \text{Softmax}(W_c v + b_c) \quad (10)$$

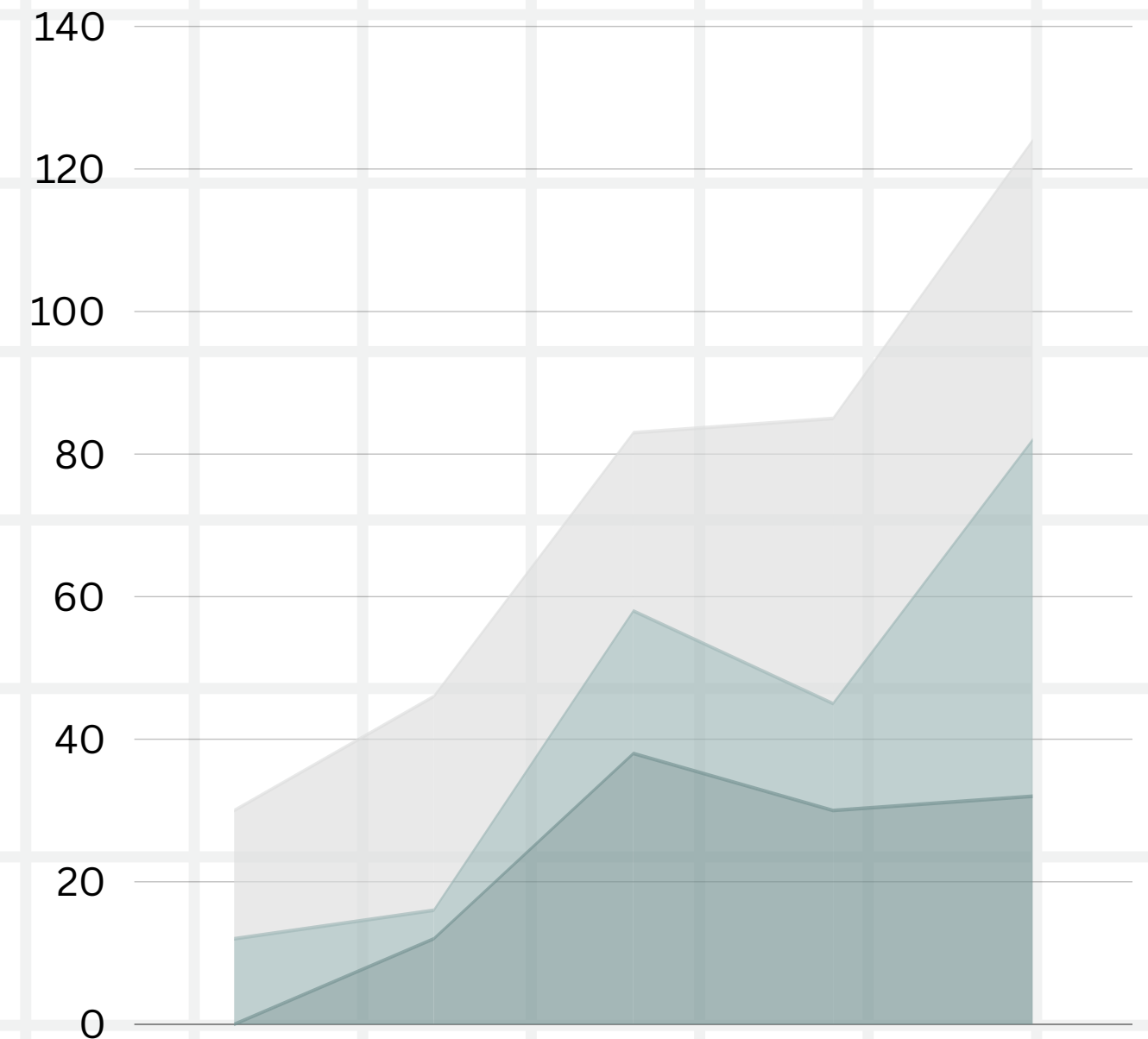
$$L = - \sum_d \text{Log}(p_{di}) \quad (11)$$

where  $i$  is the label of document  $d$ .

# 比較模型種類



- Naive Bayes
- SVM (Support Vector Machine)
- Neural Network
  - CNN
  - Branched CNN
  - LSTM
  - GRU





# 分析結果



# 分析結果



Methods	Only Headline	Only article	Concatenation of headline and article	Maximum
Naive Bayes+TFIDF+Unigrams	39	58	59	59
Naive Bayes+TFIDF+Bigrams	29	32	33	33
Naive Bayes+Bag-of-means	49	63	63	63
SVM+TFIDF+Unigrams	41	69	69	69
SVM+TFIDF+Bigrams	55	76	71	76
SVM+AverageSG	57	69	66	69
CNNs	80	80.5	81.7	81.7
Branched CNNs	83.33	84.52	84.6	84.6
LSTM	84	85.25	85.32	85.32
GRU	81	82.7	82.7	82.7
Headline Attention Network without attention layer	-	-	-	85.25
Headline Attention Network	-	-	-	89.54



Table 2: Bias Detection Accuracy in percentage. Maximum is the best value among the three divisions of our dataset for baselines.

# 分析結果




- **神經網路模型較傳統分類模型準確**
- **Headline Attention Network 有最高的正確率**
- **單看標題與看完文章內容的預測結果差異不大**





# 結論

- 標題及內文影響效果差異小
  - 可利用標題作為判斷依據
  - **Headline Attention Network 模型可以推廣至其他文字分類**  
(例如：部落格和線上趨勢分析文章)
- 

# 文獻參考



## 🔍 REFERENCES

**Detecting Political Bias in News Articles Using Headline Attention**

**Editors:**

Tal Linzen, Grzegorz Chrupała, Yonatan Belinkov, Dieuwke Hupkes

**Month/year:**

August/2019

**Pages:**

77–84

**THANK YOU**

