国际人工智能会议 AAAI 2021论文北京预讲会

Exploring Explainable Selection to Control Abstractive Summarization Wang Haonan¹, Gao Yang¹, Bai Yu¹, Mirella Lapata², Huang Heyan¹

¹School of Computer Science and Technology, Beijing Institute of Technology ²Institute for Language, Cognition and Computation, School of Informatics, University of Edinburgh

3

6

Introduction

Rely on **selecting** informative content (**extractor**) as well as aggregating into a summary in line with linguistic expression (**abstractor**). The extractors are largely **black-box decisions** without a rationale of what is **informative content**. We need methods for identifying the sentence **informativeness**, identifying whether a sentence is **relevant** to a document and, if so, to what extent. Another importance influence is the **novelty** of the **contribution** a sentence makes to a summary. Therefore, to reveal more of the inner workings of these black-box models so as to inject a level of control into the **substance** and **integrity** of the final summary, we developed a novel select-and-generate framework, called **ESCA**, that focuses on **explain-ability**.

The key to the framework is an interaction matrix that highlights the decisions made about

Proposed Model



each sentence, which can be decoupled into three explicit components, **informativeness**, **relevance**, **novelty**. Further, which content is selected for extraction can be **controlled** by setting **thresholds** for novelty or relevance, and applying **Mask Matrix** to adjust the probability of extraction.

Contributions

2

- Proposed an explainable content selection module for document summarization.
- Ability to extract the **appropriate content** for generating a desired summary based on **explicit** and **quantified** measures of **informativeness**, **relevance** and **novelty** to the final summary.
- Automatically creating synthetic datasets w.r.t novelty and relevance for exercising controllable inference without the need to retrain the entire system.

4 Crucial formula

Pair-wise Learning Loss for Extractor

$$L_{ext} = -\sum_{i=1}^{m} \sum_{j=1}^{m} \left(\hat{P}_{ij} \log r_{ij} + (1 - \hat{P}_{ij}) \log(1 - r_{ij}) \right)$$

Construct interaction matrix of directional sentence pair relation.

$$q_{i,j}(h_i, a_i, h_j, d) = \sigma(\underbrace{W_c h_i}_{\text{infor.}} + \underbrace{h_i^T W_r d}_{\text{rel.}} + \underbrace{h_i^T W_s h_j - h_i^T W_n \tanh(a_i)}_{\text{nov.}} + b_m)$$

Hidden state & accumulated summary representation

$$h_t^* = \alpha_t Z_e \qquad a_i = \frac{1}{s} \sum_t^{i-1} \sum_{k=1}^s h_t \cdot q_{tk}$$

Output distribution of decoder

$$P_{vocab}(w) = P(y_t | y < t, x; \theta) = softmax(W_2(W_1[s_t, h_t^*] + b_1) + b_2)$$

Conclusion & Case Study

- We present a novel hybrid framework for document summarization
 ESCA, equipped with a pair-wise ranking extractor that connects with an abstractor armed with a sentence-level attention pointer.
- ESCA is designed to explicitly explain why sentences are marked for extraction and to control which sentences are extracted according to novelty and relevance score.
 - The relationship between sentence pairs contributes more to the degree of sentence importance than the relationship between sentence and document.

Controllable Inference

$$\hat{Q}^{s} = Q^{s} \odot M$$
, where $M_{ij} = \begin{cases} 1, val \ge \epsilon \\ 0, val < \epsilon \end{cases}$

Experiment

 Abstractive summary is a trade-off between controllability and quality of summary, high-quality summary should focus on more attributes information.

Gold Summary

																		 <u>Police officers have shut down an enormous 1000 rave in Sydney's east.</u> They were called to abandoned industrial area in botany on Saturday night.
Model	CNN/DailyMail				NYT50		Control	Threshold	R-1	R-2	R-L	Models	04	Criteria ranking				 Police were forced to use capsicum spray on the group after back up came. One officer had glass removed from his headafter the crowd threw bottles.
	R-1	R-2	R-L	R-1	R-2	R-L		0	44.78	35.39	30.33	woders	QA	Infor.	Nov.	Rel.	Flu.	• A woman was arrested and is being questioned after assaulting an officer . FGIM
PG+Coverage	39.53	17.28	36.38	43.71	26.40	37.79		0.2	15.000	26.004	42.054	PG+Cov	26. <i>o</i>	-0.28	-0.43	-0.05	-0.39	 have sustained injuries after attempting to close down an enormous 1000 rave One officer had to have a piece of glass removed from his head after having a bottle thrown at him
Select-Reinforce	40.88	17.80	38.54	_	_	-	$Nov(\epsilon_n)$	0.5 45.00	45.66†	30.28	43.05†	Bottom-Up	31.3	-0.07	0.02	-0.08	-0.02	Control by Relevance ($\epsilon_r = 0.5$) • sustained injuries after attempting to close <u>down an enormous 1000 person rave in Sydney's ea</u>
Inconsistency-Loss	40.68	17.97	37.13	_	_	_		0.4	45.26↑	36.08↑	42.67↑	Inconsistency	29.8	-0.10	-0.12	-0.15	-0.14	 Police were forced to use capsicum spray on and one officer had to have a piece of glass A 26-year-old woman was arrested after she allegedly assaulted an officer .
Bottom-Up	41.22	18.68	38.34	47.38	31.23	41.81		0.5	45.28↑	35.90↑	42.71↑	FGIM-BERT	39.2	0.15	0.14	0.15	0.12	Gold Summary • Jeralean Talley was born on may 23, 1899
Explicit-Select	41.54	18.18	36.47	-	-	_		0	41.35	18.50	38.57	Gold	\sim	0.30	0.40	0.13	0.48	 She credits her longevity to her faith. Inherited the title of world's oldest person following the death of Arkansas woman
SENECA	41.52	18.36	38.09	47.94	31.77	44.34		0.2	41.414	10.57*	29 (2*	Bottom-up	~	-0.23	-0.07	-0.15	~	FGIM • Jeralean Talley was born in rural montrose on may 23, 1899, and credits her long life to her faith
BERTSumAbs	41.72	19.39	38.76	48.92	30.84	45.41	$\operatorname{Rel}(\epsilon_r)$	0.5	41.41	18.3/1	38.627	FGIM-BERT	\sim	0.10	0.03	0.05	\sim	 <u>Asked for her key to longevity</u>, the Detroit free press reports . Gertrude Weaver, a 116-year-old arkansas woman who was the oldest
FGIM-Transformer	41.65	18.89	37.94	47.63	30.10	43.94		0.5	41.52↑	18.67↑	38.55↓	$FGIM(\epsilon_n = 0.3)$	\sim	0.05	0.10	0.02	\sim	Control by Novelty $(\epsilon_n = 0.3)$ • tops a list maintained by Geron planck , which tracks the world's longest-living prople .
FGIM-BERT	42.12	19.52	39.07	49.41	32.22	45.83		0.7	41.27↓	18.44↓	38.43↓	$FGIM(\epsilon_r = 0.5)$	\sim	0.07	-0.02	0.07	\sim	 Talley's five generations of her family have lived in the Detroit area . Talley was born on may 23, 1899 , and credits her long life to her faith.



