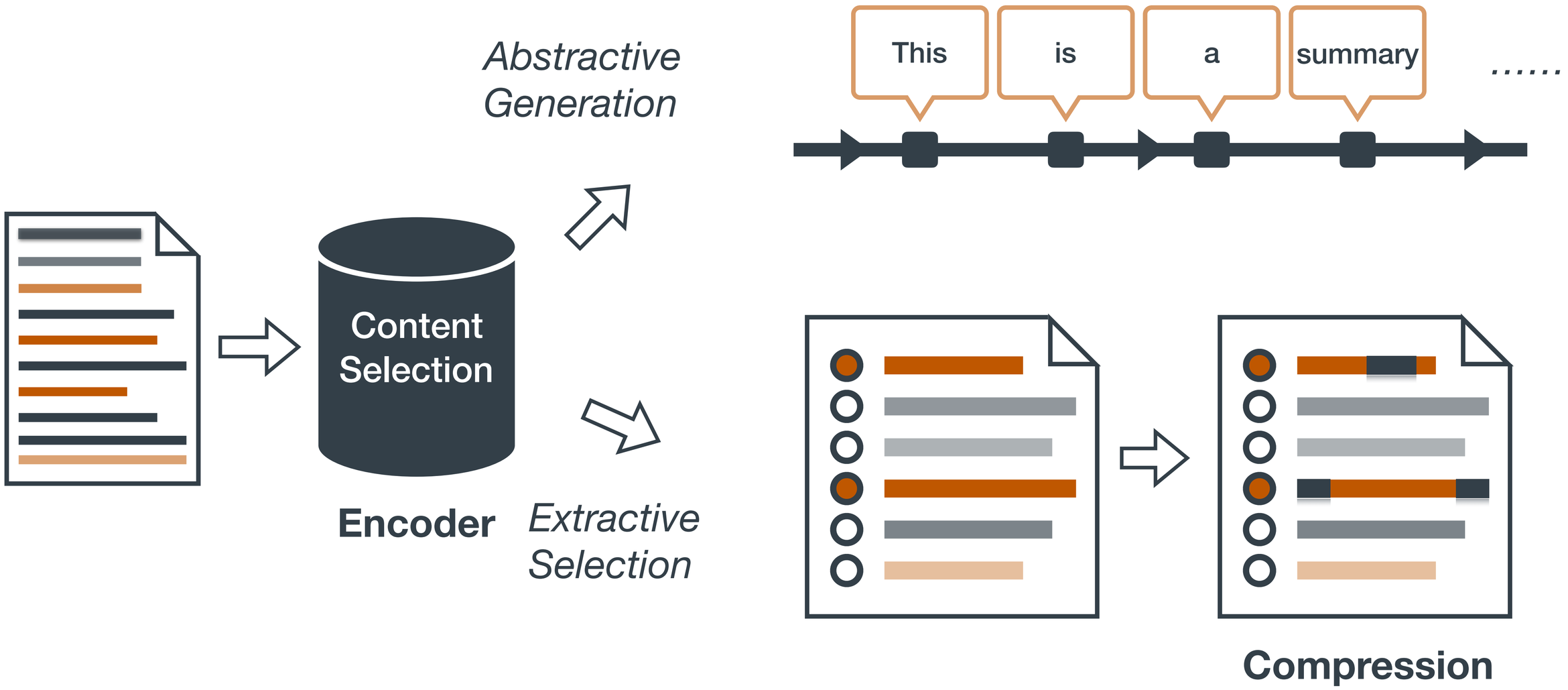
Learning Discrete Compressions for Neural Extractive Text Summarization

Jiacheng Xu

Research Preparation Exam

September 30, 2019

Neural Text Summarization



From Ferguson to Ebola-stricken West Africa: The gripping photos that won the 2015 Pulitzer Prize...as tiny South Carolina newspaper takes prestigious top honor

It is one of the most prestigious honors bestowed upon journalists and people in the arts. And today, the Pulitzer prize for journalism went to The Post and Courier newspaper of Charleston, South Carolina - which has a tiny staff of just 80 and a daily circulation of 85,000.

The paper's powerful photo series entitled 'Till Death Do Us Part,' on domestic violence scooped the top award for its exploration into why South Carolina is among the deadliest states for women.

The newspaper was awarded the gold medal for public service, the first time in five years that the prize has gone to such a small newspaper.

.....

106 /913 words shown here

Reference Summary

The Post and Courier newspaper of Charleston, South Carolina was awarded the gold medal for public service.

[59 words omitted ...]

From Ferguson to Ebola-stricken West Africa: The gripping photos that won the 2015 Pulitzer Prize...as tiny South Carolina newspaper takes prestigious top honor

It is one of the most prestigious honors bestowed upon journalists and people in the arts. ~~And today,~~ **the Pulitzer prize for journalism went to The Post and Courier newspaper of Charleston, South Carolina** ~~, which has a tiny staff of just 80 and a daily circulation of 85,000.~~

The paper's powerful photo series entitled 'Till Death Do Us Part,' on domestic violence scooped the top award for its exploration into why South Carolina is among the deadliest states for women.

The newspaper was awarded the gold medal for public service, the first time in five years that the prize has gone to such a small newspaper.

.....

106 /913 words shown here

Reference Summary

The Post and Courier newspaper of Charleston, South Carolina was awarded the gold medal for public service.

[59 words omitted ...]

Criteria

Better

Abstractive Model

Extractive Model

- **Redundancy**
- Grammaticality
- Factuality
- Scalability

Too Much Detail

"~~And today~~, the Pulitzer prize for journalism went to The Post and Courier newspaper of Charleston, South Carolina ~~- which has a tiny staff of just 80 and a daily circulation of 85,000.~~"

Repeated Content

*"The **Pulitzer** committee at **Columbia University** in New York …"*

*"The **Pulitzer** prizes, awarded annually by **Columbia University** …"*

Criteria

Better

Abstractive Model

Extractive Model

Redundancy

Grammaticality

Factuality

Scalability

Grammar Mistakes (Only Problem)

“The Pulitzer prizes, awarded annually by Columbia University in New York.”

Criteria

Better

 *Abstractive Model*

 *Extractive Model*

Redundancy

Grammaticality

Factuality

Scalability

Wrong Fact or Statement ( Only Problem)

"The Pulitzer prizes, awarded annually by New York University, recognize extraordinary work …"

Criteria

Better

 *Abstractive Model*

 *Extractive Model*

 Redundancy

 Grammaticality

 Factuality

 Scalability

Long Document Understanding

(Sentence # 5) … Pulitzer prizes for international reporting and feature photography …

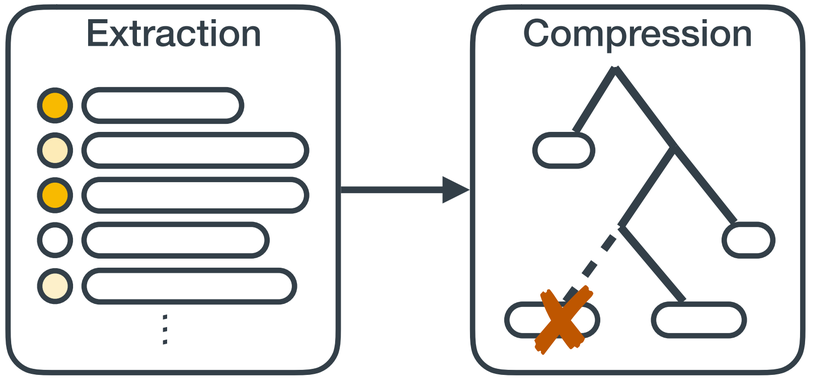
……

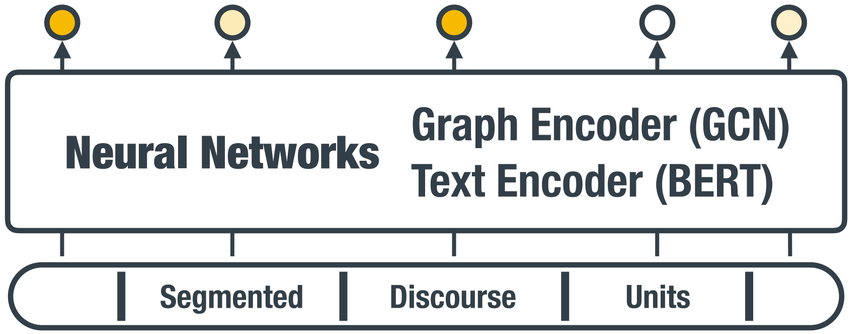
(Sentence # 34) … received the Pulitzer prize in photography …

……

Our Approaches

- Neural-based summarization models with joint content selection and text compression
- Pruning constituency parse tree or RST discourse tree
- Strong performance on benchmark datasets

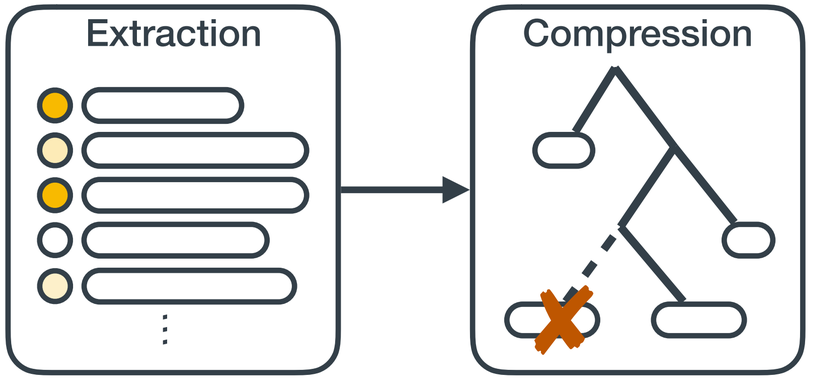




Our Approaches

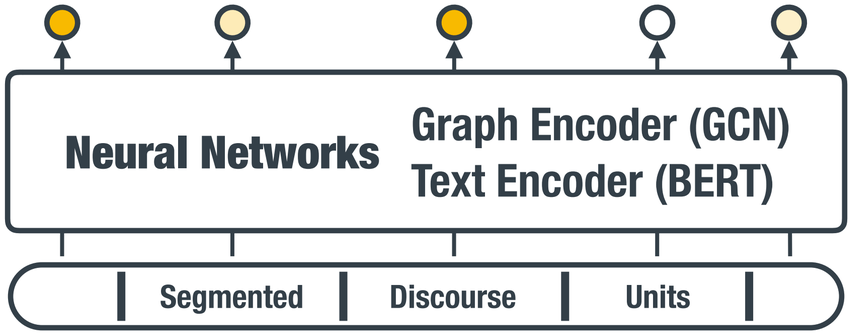
Neural Extractive Text Summarization with Syntactic Compression

Jiacheng Xu and Greg Durrett, *EMNLP 2019*



Discourse-Aware Neural Extractive Model for Text Summarization

Jiacheng Xu, Zhe Gan, Yu Cheng and Jingjing Liu, Submitted to *AAAI 2020*



Neural Extractive Text Summarization with Syntactic Compression

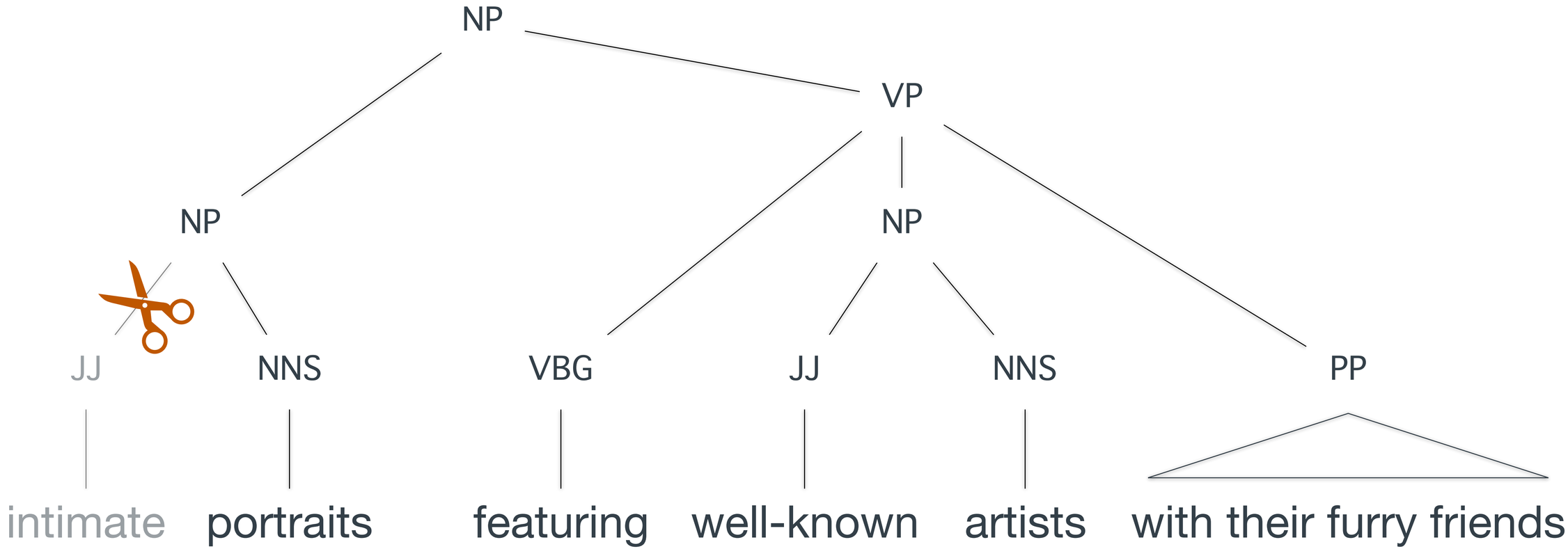
Learning Discrete Compressions for Neural Extractive Text Summarization

Related Work

- Joint model of extraction and compression
  - sparse features,
  - Integer Linear Programming (ILP) as the optimization framework,
  - based on constituency parses *(Berg-Kirkpatrick et al., 2011; Wang et al., 2013; Li et al., 2014)* or RST trees *(Hirao et al., 2013; Durrett et al., 2016)*.
- Neural-based deletion-based sentence compression model without using tree structure *(Filippova et al., 2015).*

- We propose:
  - Neural networks based model
  - End-to-end MLE training & inference
  - Pruning constituency parse tree

Compression on Constituency Parse Tree



NP = Noun Phrase

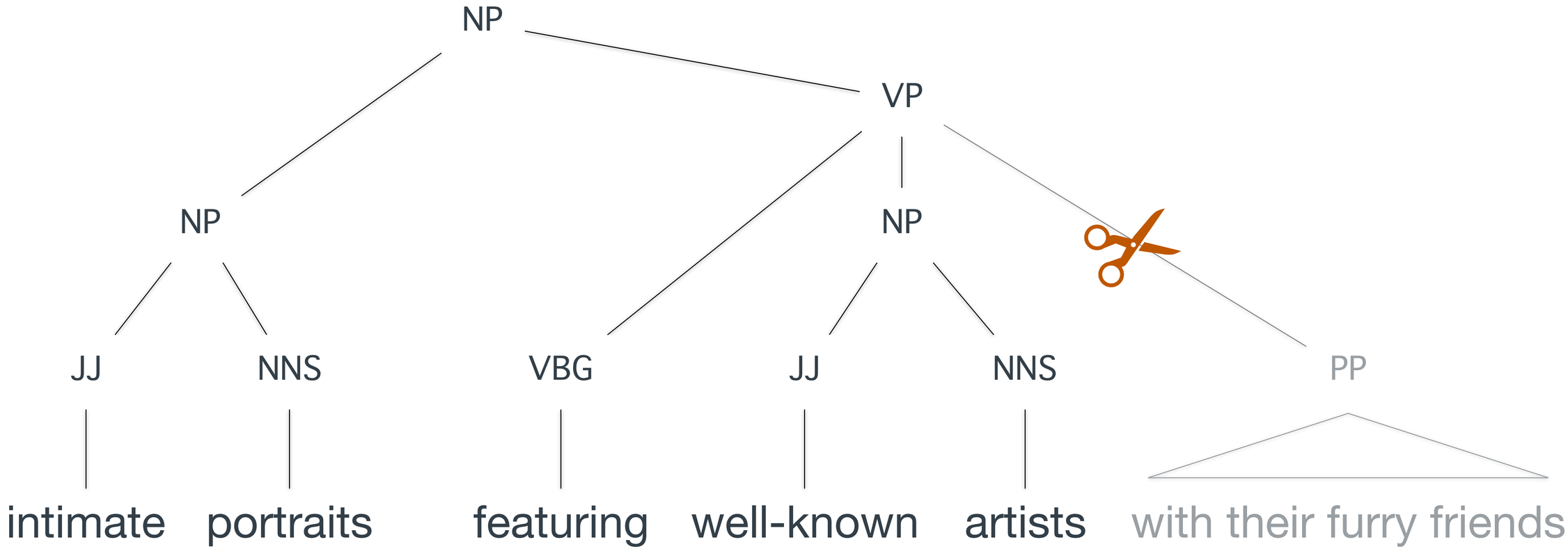
JJ = Adjective

NNS = plural noun

PP = Prepositional Phrase

VBG=present participle, G-form

Compression on Constituency Parse Tree



NP = Noun Phrase

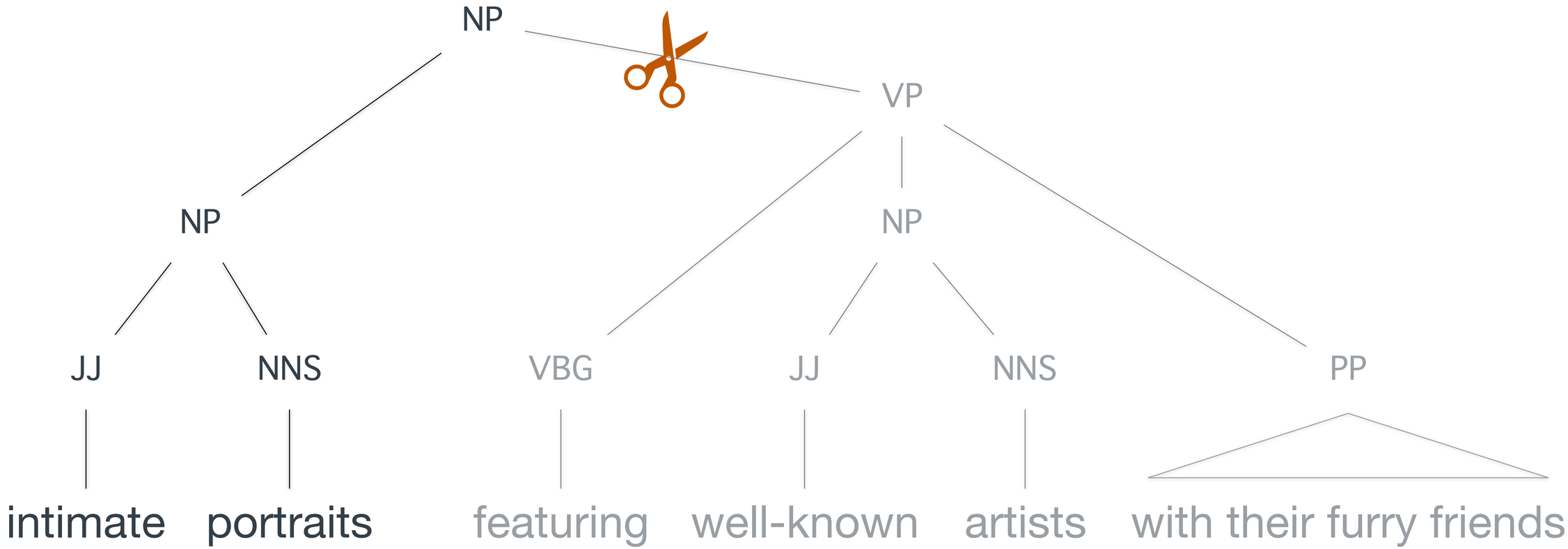
JJ = Adjective

NNS = plural noun

PP = Prepositional Phrase

VBG=present participle, G-form

Compression on Constituency Parse Tree



NP = Noun Phrase

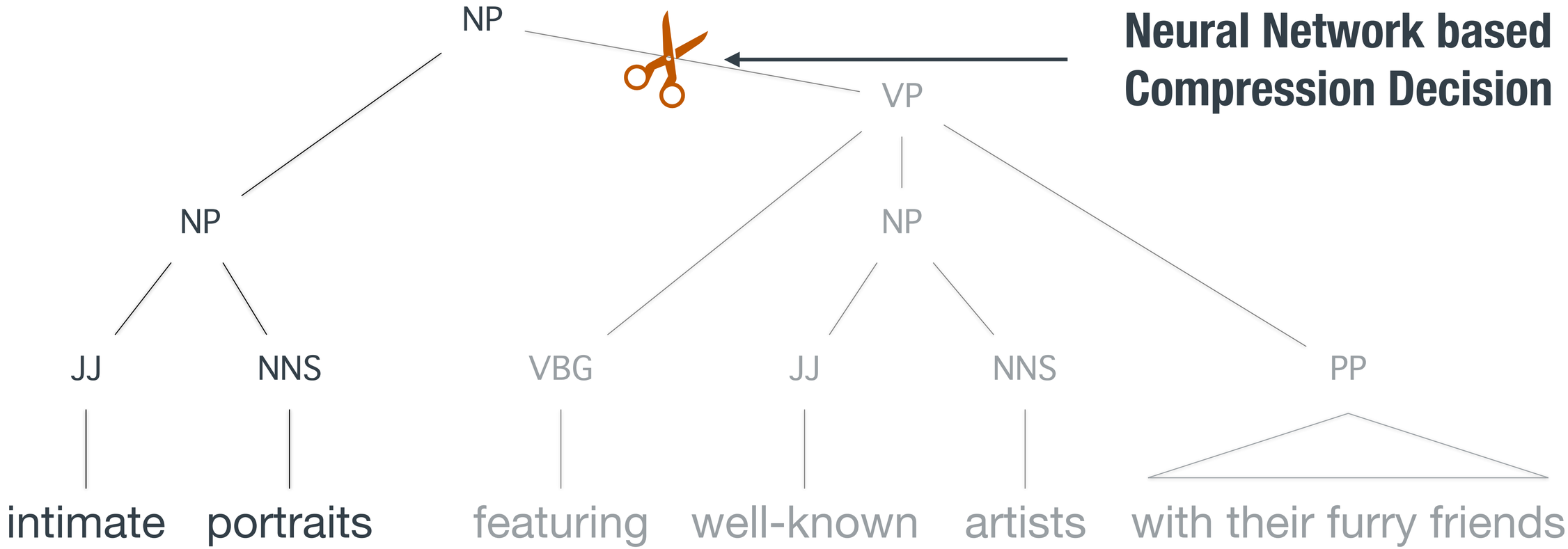
JJ = Adjective

NNS = plural noun

PP = Prepositional Phrase

VBG=present participle, G-form

Compression on Constituency Parse Tree



Neural Network based Compression Decision

NP = Noun Phrase

JJ = Adjective

NNS = plural noun

PP = Prepositional Phrase

VBG=present participle, G-form

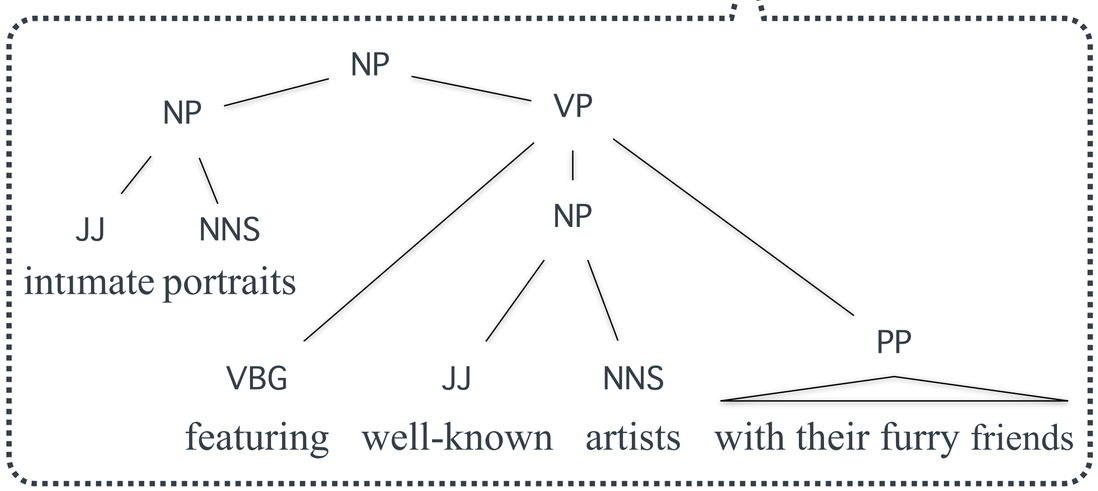
Preprocessing - Compression Extraction

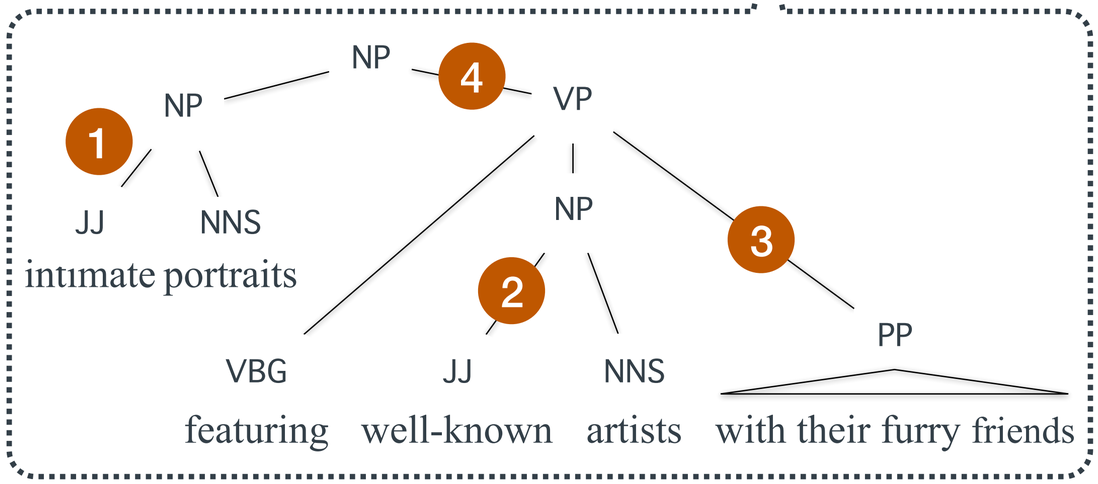
intimate portraits featuring well-known artists with their furry friends

Node	License	Condition
JJ	Conditional	parent = ‘NP’
VP	Conditional	(child[0] = ‘VBG’ or ‘VBN’) & (parent = ‘NP’)
...	...	...

1. [JJ] intimate
2. [JJ] well-known
3. [PP] with their furry friends
4. [VP] featuring well-known artists with their furry friends

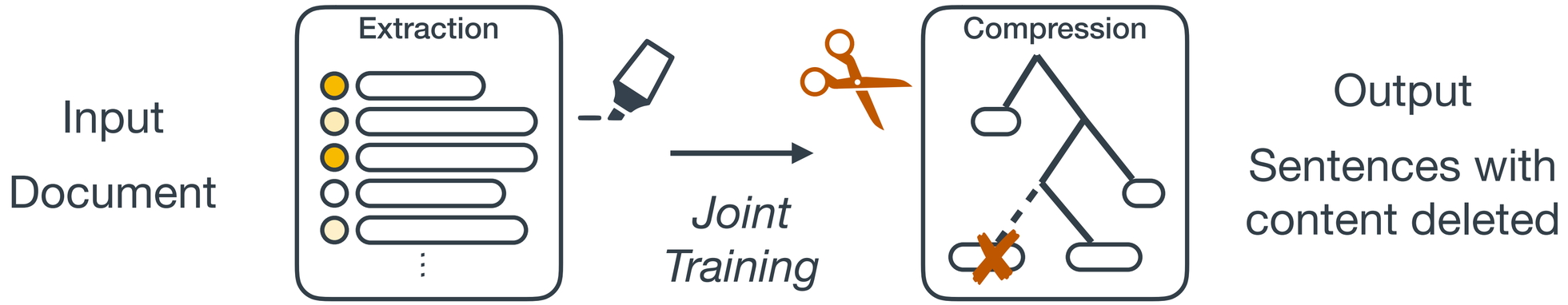
Text → **Constituency Parses** → **Compression Rules** → **Extracted Compression**





Model

Joint Extraction & Compression Summarization (JECS)



Sentence Selection Module

- Encoding: encodes the document with BLSTM & CNN
- Decoding: predicts a sequence of chosen sentences indexes
- Training with extractive oracle

Text Compression Module

- Encoding: encodes the target compression phrases
- Prediction: $f(v_{\text{compression}}, v_{\text{sentence}}, \cdots)$
- Training with compression oracle

Oracle - Searching for *gold* Combination

$$\text{R-1-Precision} = \frac{\text{Number of overlapping words}}{\text{Number of words in the system summary}}$$

$$\text{R-1-Recall} = \frac{\text{Number of overlapping words}}{\text{Number of words in the reference summary}}$$

$$\text{R-1-}F_1 = \text{Harmonic Mean of recall and precision}$$

Given the reference summary and the document,

F_1(oracle combination) ≥ any other F_1(sentences combination) of the document.

With beam search, we greedily pick up the sentences to raise the F_1 value until it does not go up anymore.

Oracle - Searching for *gold* Combination

R-1 Oracle	Document
3.7 ○	It is one of the most prestigious honors.
45.2 ●	And today, the Pulitzer prize for journalism went to **The Post and Courier newspaper of Charleston, South Carolina** - which has a tiny staff of just 80 and a daily circulation of 85,000.
5.2 ○	The paper's powerful photo series entitled 'Till Death Do Us Part'.

Reference Summary

The Post and Courier newspaper of Charleston, South Carolina was awarded the gold medal for public service.

.....

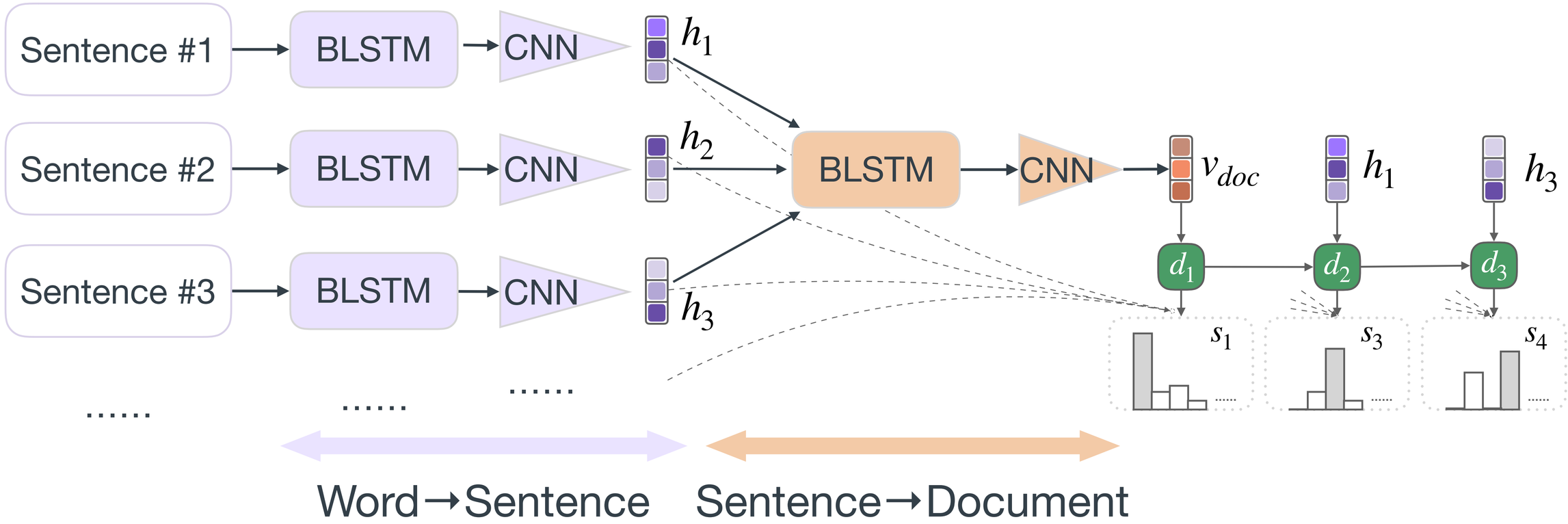
Sentence to Compress

And today, the Pulitzer prize for journalism went to The Post and Courier newspaper of Charleston, South Carolina - which has a tiny staff of just 80 and a daily circulation of 85,000.

Candidate Compression Phrase

- Compare the phrase and the summary
- If no significant overlapping, then delete
- In this case, R-1 raises from 45.2 to 68.4

Module - Sentence Selection



Encoder
Encoding Words & Sentences with separate [BLSTM + CNN] modules

Decoder
Sentence index prediction with pointer network

Similar architecture from previous work (Zhou et al., 2018)

Module - Text Compression

Task:

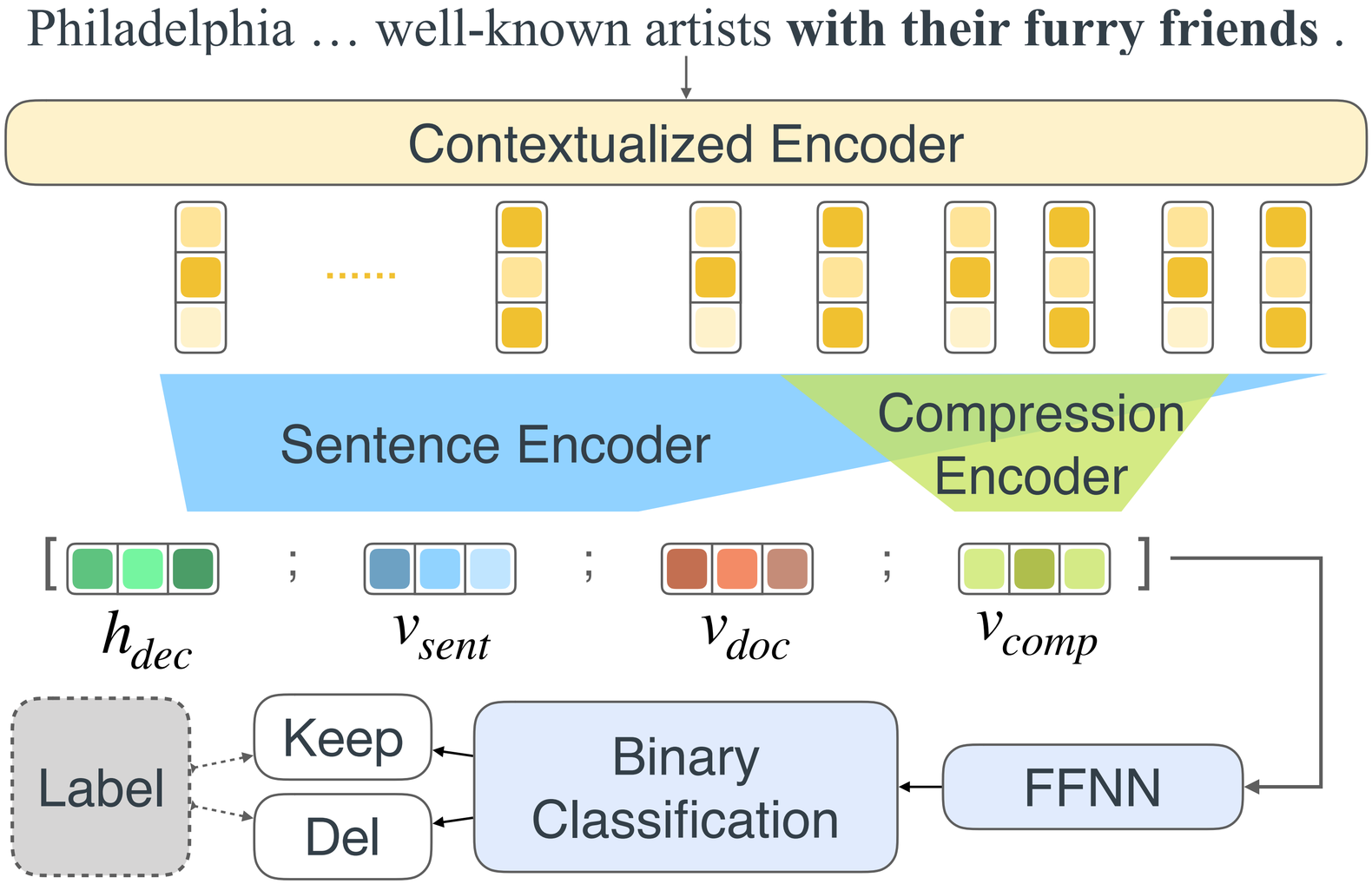
Decide if we are going to keep '**with their furry friends**' given the context

Input:

The original sentence and the target compression phrase

Output:

Keep or Del



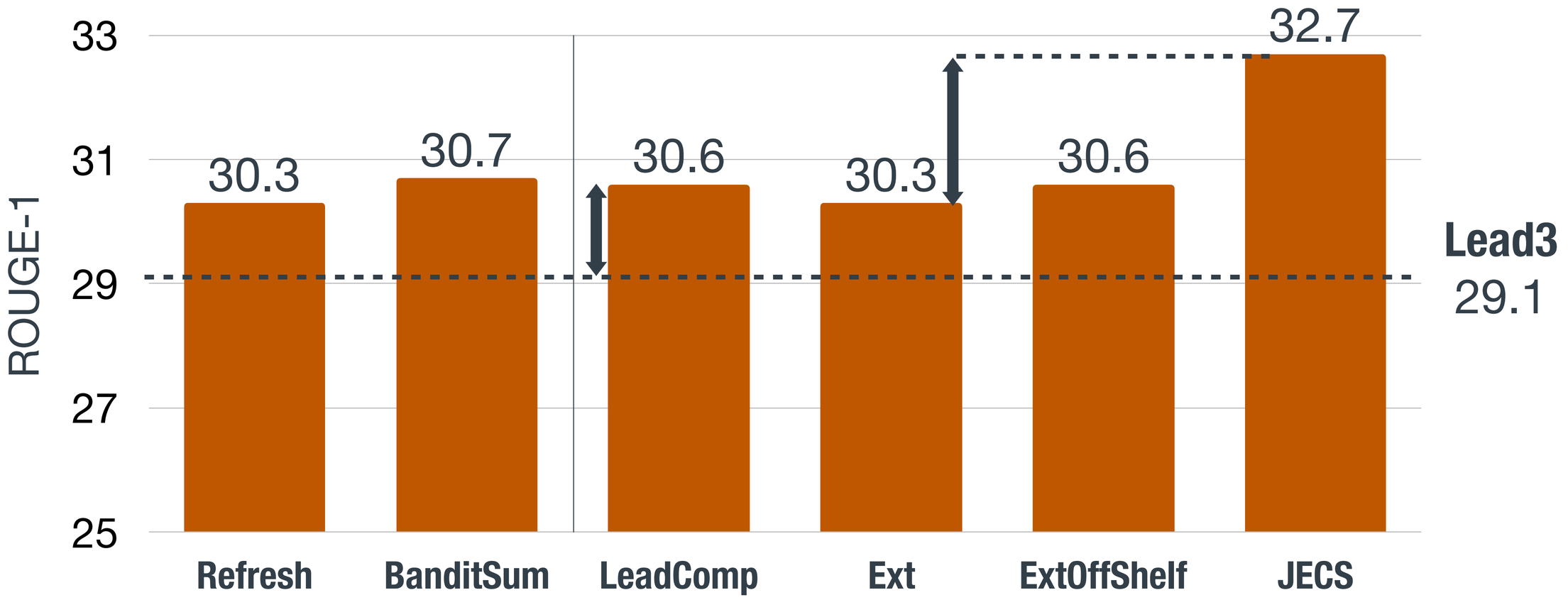
Experiments

- Task: Single Document Text Summarization
- Datasets: CNNDM, CNN (more abstractive), NYT
- Metric: ROUGE *(Lin 2004)*
  - Measuring Word & phrases overlapping w.r.t. reference summary
  - 0 (No overlapping) - 100 (Exact Same)
- Lead3 is a baseline where we treat the first three sentences of the document as the summary.

Comparison

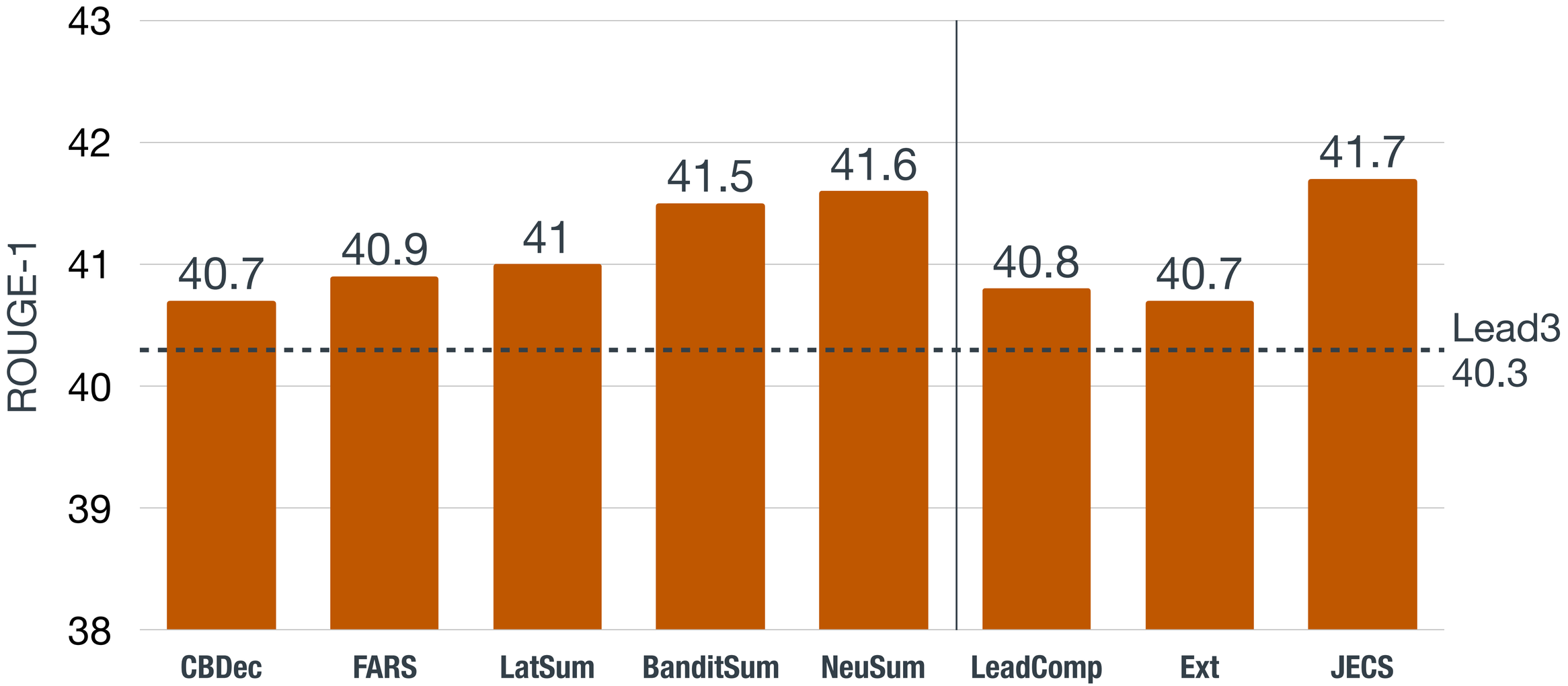
- Extractive Models:
  - **Refresh** (Narayan et al., 2018) & **BanditSum** (Dong et al., 2018)
- Ours:
  - [Compression] **LeadComp**: learning-based text compression on Lead3
  - [Pipeline System] **ExtOffShelf**: our extraction module + off-the-shelf compression model from *(Filippova et al., 2015)*
  - [Joint ext & comp] **JECS**: our model

Experiments - CNN

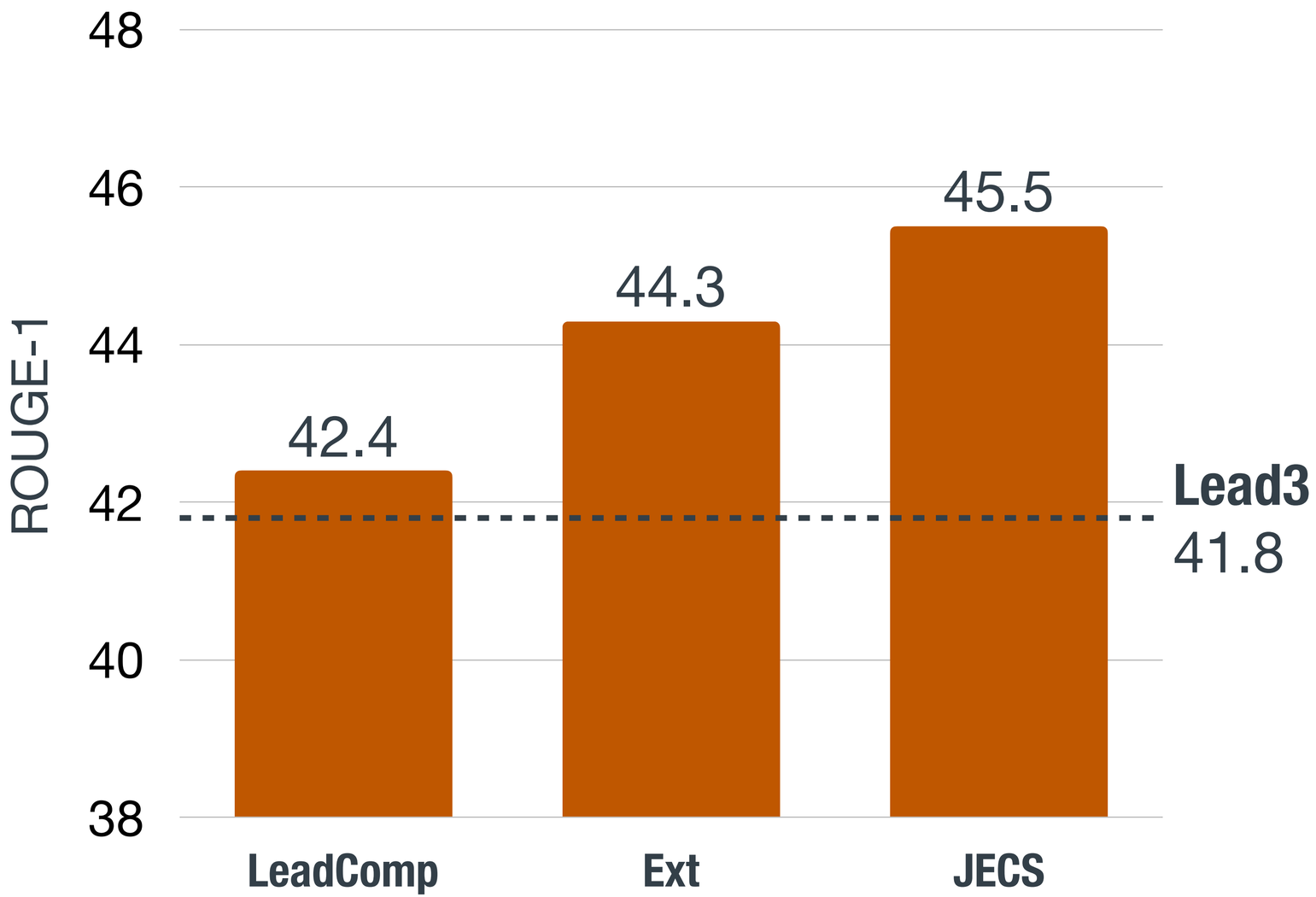


- ExtOffShelf = Extraction + off-the-shelf compression model (Fillipova)
- Learning-based compression helps on **Lead3** & **Ext**(raction)
- Joint learning matters (**JECS** > **ExtOffShelf**)

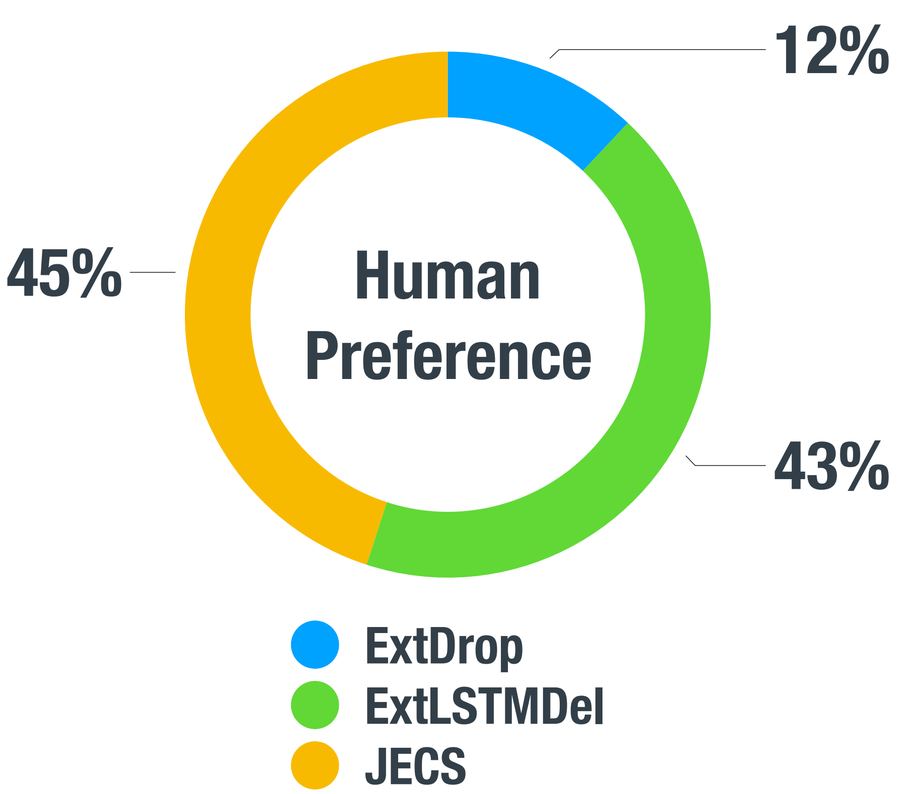
Experiments - CNNDM

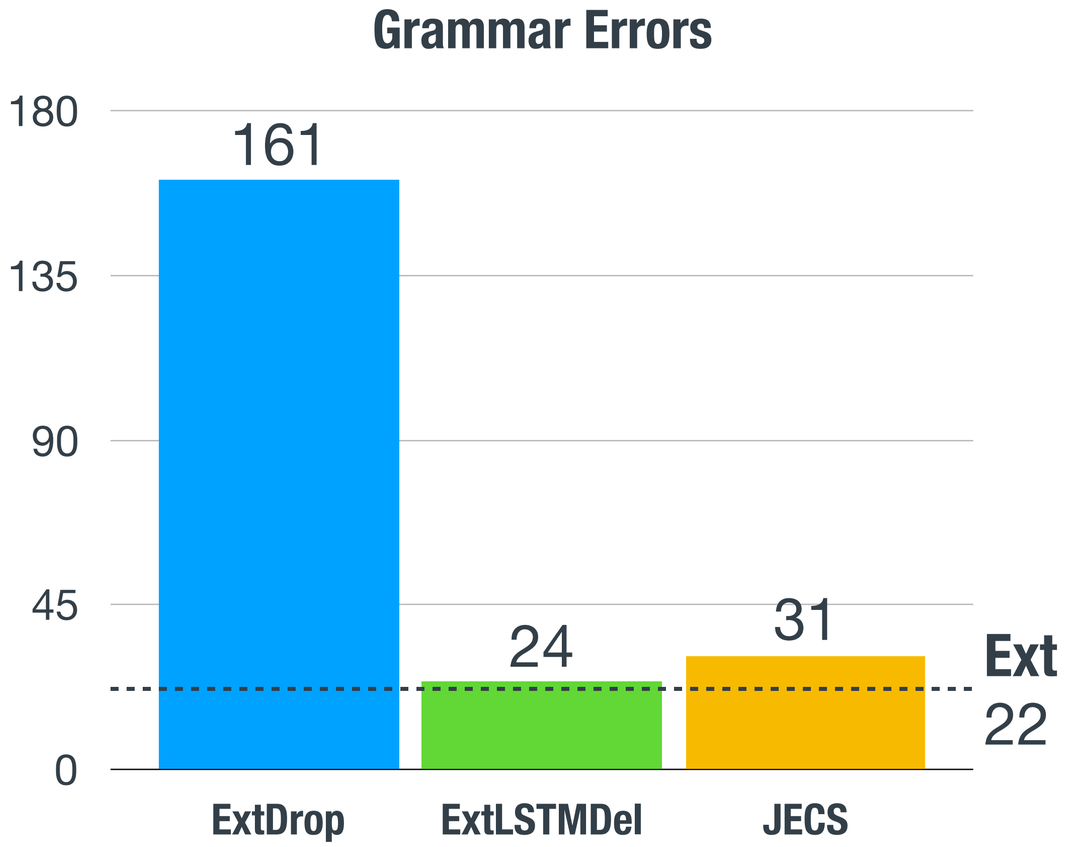


Experiments - NYT



Human Evaluation





Human preference judged by Amazon Turkers based on **grammaticality**.

Grammar errors reported by Grammarly®

Examples

Reference Summary	Prediction with Compressions
Rebecca Francis' photo with a giraffe was shared by Ricky Gervais. Francis was threatened on Twitter for the picture. Francis, a hunter, said the giraffe was "close to death" and became food for locals.	(CNN) Five years ago, Rebecca Francis posed for a photo while lying next to a dead giraffe. The trouble started Monday, when comedian Ricky Gervais tweeted the photo with a question. Francis, who has appeared on the NBC Sports Network outdoor lifestyle show "Eye of the Hunter" and was the subject of an interview with Hunting Life in late March, responded in a statement to HuntingLife.com on Tuesday, which was posted on its Facebook page.
Mullah Omar, the reclusive founder of the Afghan Taliban, is still in charge, a new biography claims. An ex-Taliban insider says there have been rumors that the one-eyed militant is dead.	(CNN) Mullah Mohammed Omar is "still the leader" of the Taliban's selfdeclared Islamic Emirate of Afghanistan. The Taliban's "Cultural Commission" released the 11-page document in several different translations on the movement's website, ostensibly to commemorate the 19th anniversary of an April 4, 1996, meeting in Afghanistan's Kandahar province when an assembly of Afghans swore allegiance to Omar.

Discourse-Aware Neural Extractive Model for Text Summarization

Learning Discrete Compressions for Neural Extractive Text Summarization

Challenges

Hand-coding Compression Rules

- 11 categories of rules for text compression.
- Move away from handwritten rules?

A Two-Stage System

- No global optimization & decision bridging the two stages;
- Error propagation.

Compression Strategy with Length Bias

- Too aggressive on small deletions
  - eg: a good, solid agreement
- Not aggressive on big deletions

DiscoBERT

Elementary Discourse Units (EDUs) from RST as the minimal selection units

- Rhetorical Structure Theory: a theory of text structure as a theoretical basis for computational text planning
- Grounding sentence compression into phrase selection
- Compression at sub-sentence level
- Encoding the document with BERT and Self-Attentive Span Extractor

Discourse relations

- Using RST discourse graph & coreference mentions graph
- Better understanding of the document structure
- Encoding the graphs with Graph Convolutional Networks

Text Segmentation

- The whole document can be segmented into contiguous, adjacent and non-overlapping text spans called Elementary Discourse Units.
- Dependencies represent their rhetorical relations.

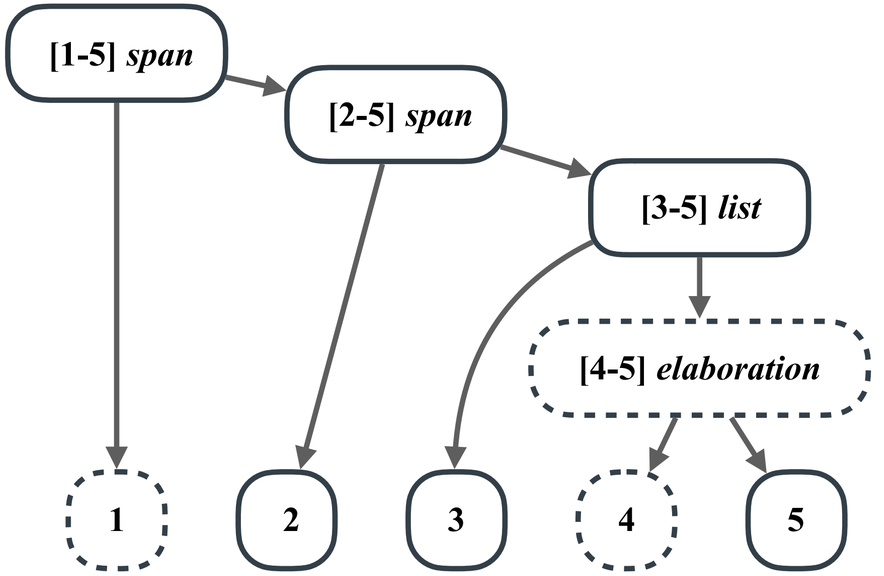
Original Text

Winner: This iconic photo by New York Times photographer Daniel Berehulak, was part of a winning series, and shows James Dorbor, 8, suspected of being infected with Ebola, being carried by medical staff to an Ebola treatment center in Monrovia, Liberia.

Segmented EDUs

[1] Winner: **[2]** This iconic photo by New York Times photographer Daniel Berehulak, was part of a winning series, **[3]** and shows James Dorbor, 8, **[4]** suspected of being infected with Ebola, being carried by medical staff to an Ebola treatment center in Monrovia, Liberia.

RST Discourse Tree



EDU as Selection Unit

Document

1. [It is one of the most prestigious honors]$_1$ [bestowed upon journalists and people in the arts.]$_2$
2. [And today, the Pulitzer prize for journalism went to The Post and Courier newspaper of Charleston, South Carolina,]$_1$ [which has a tiny staff of just 80 and a daily circulation of 85,000.]$_2$

……

5. [Winner:]$_1$ [This iconic photo by New York Times photographer Daniel Berehulak, was part of a winning series,]$_2$ [and shows James Dorbor, 8,]$_3$ [suspected of being infected with Ebola,]$_4$ [being carried by medical staff to an Ebola treatment center in Monrovia, Liberia.]$_5$

……

20. [The Pulitzer prizes,]$_1$ [awarded annually by Columbia University,]$_2$ [recognize extraordinary work in U.S. journalism, literature, drama, and other categories.]$_3$

……

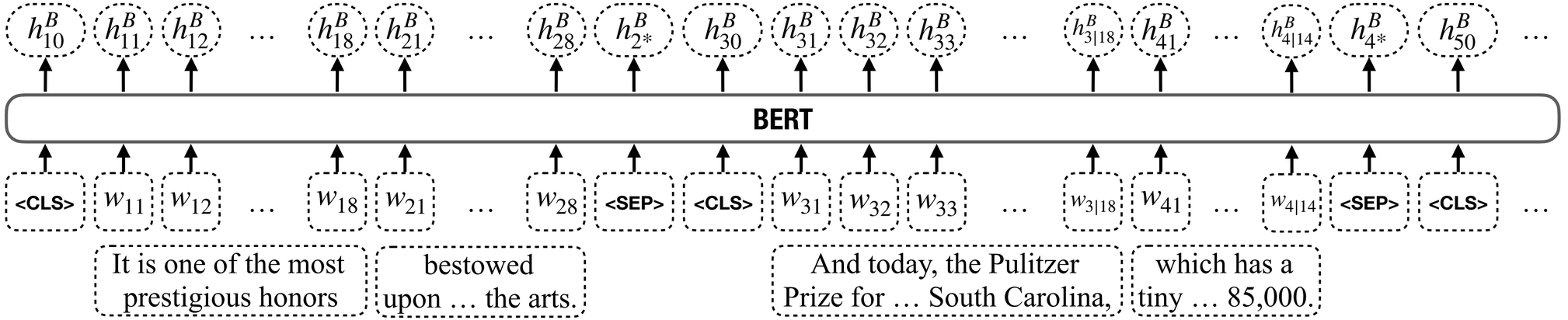
22. [Other winners of the coveted award included the St. Louis Post-Dispatch.]$_1$

……

	R-1
Sentence Selection	37.4
[EDU Selection]	44.3

Text Encoder - BERT

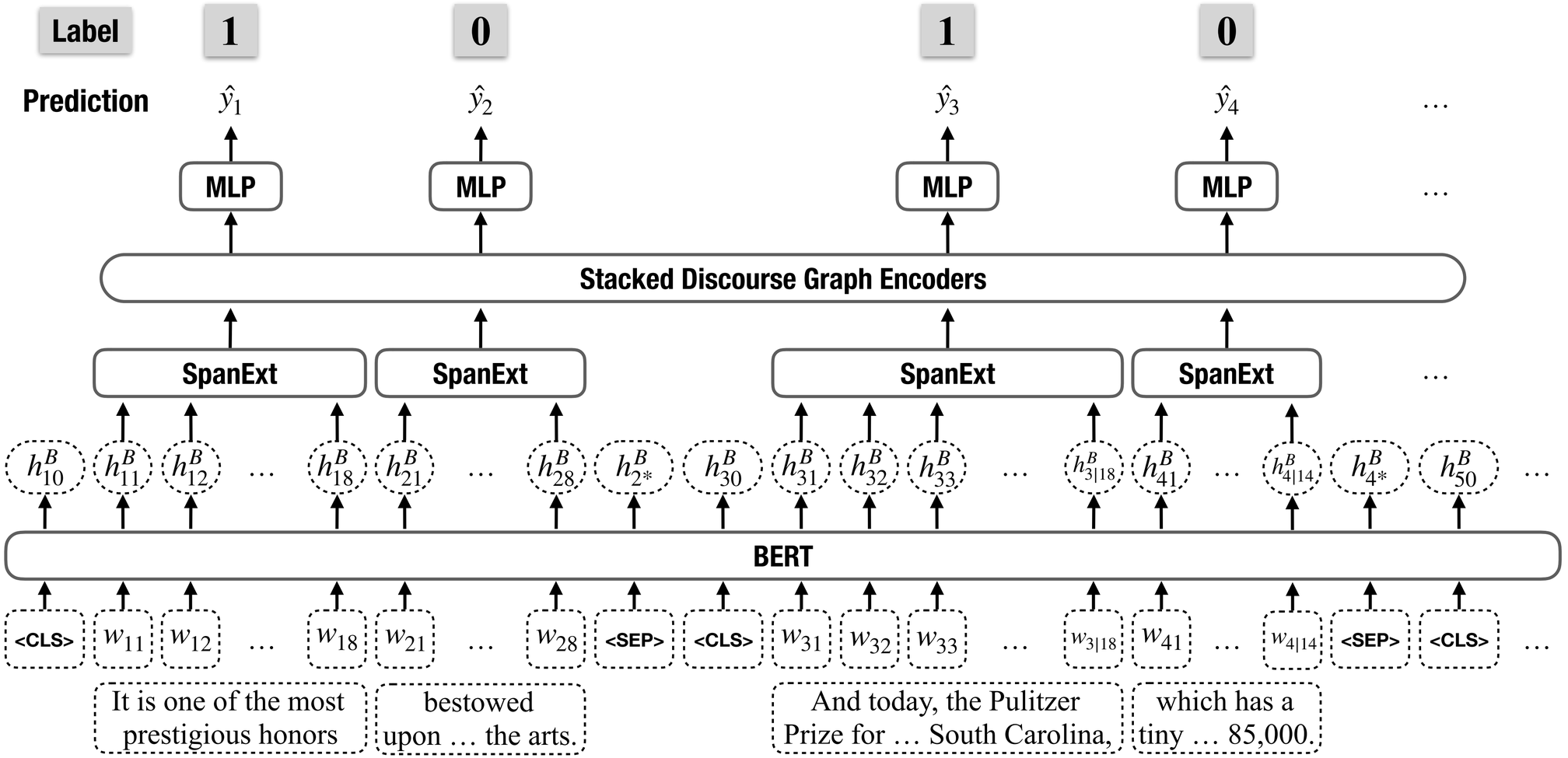
- BERT *(Devlin et al., 2019)* = Bidirectional **Encoder** Representations from Transformers
  - Pre-train deep bidirectional representations from unlabeled text
  - State-of-the-art on major NLP tasks
- We use it as the text encoder (which was LSTM before)
- We fine tune the BERT model on the summarization task



Google AI

DiscoBERT

Slightly different decoding strategy: binary classification for each time step.

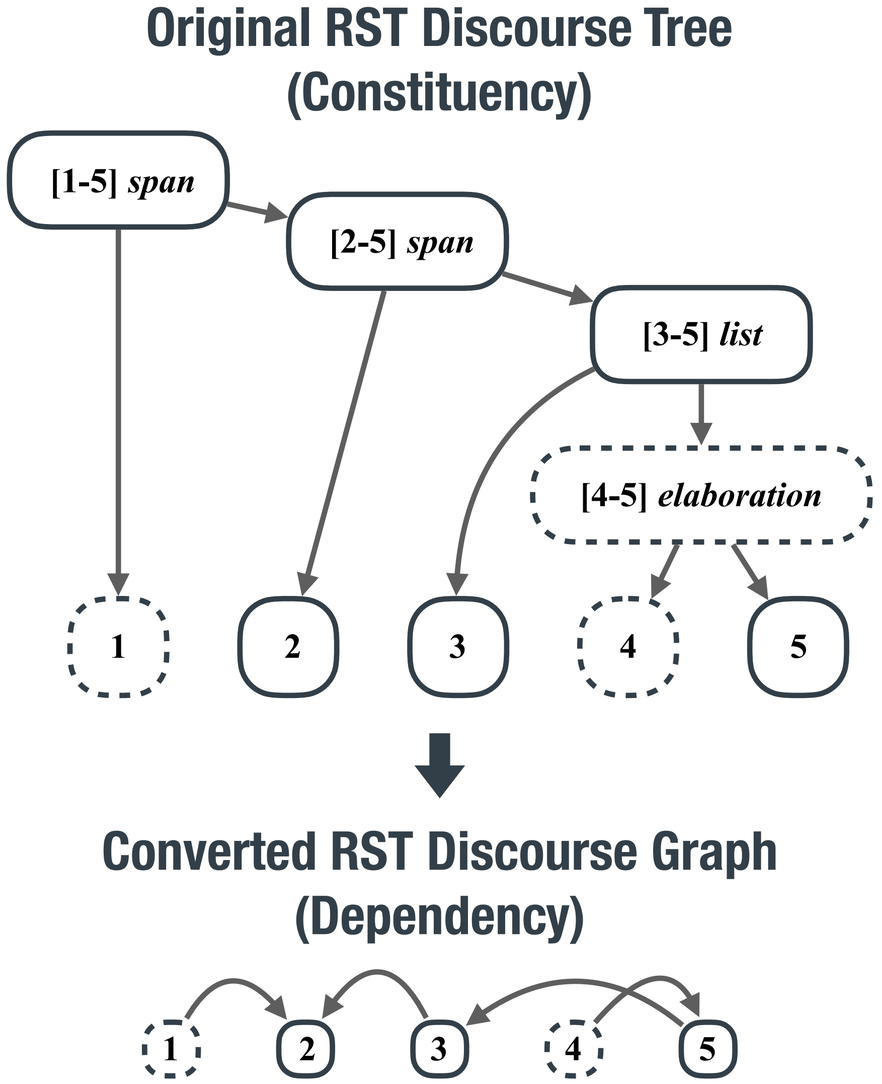


(Liu 2019)

Discourse-Aware Graphs

- Motivation
  - Modeling entity occurrences and relations
  - Capturing long-range context in the document
  - Building a neural layer for EDUs with prior knowledge injected
- Graphs proposed
  - $\mathcal{G}_R$: RST Discourse Graphs
    - Induced from RST discourse trees
    - Uni-directional dependency graph
  - $\mathcal{G}_C$: Coreference Mention Graphs
    - Induced from entity coreference mentions

RST Discourse Graph



Document

1. [It is one of the most prestigious honors]$_1$ [bestowed upon journalists and people in the arts.]$_2$
2. [And today, the Pulitzer prize for journalism went to The Post and Courier newspaper of Charleston, South Carolina,]$_1$ [which has a tiny staff of just 80 and a daily circulation of 85,000.]$_2$

.....

5. [Winner:]$_1$ [This iconic photo by New York Times photographer Daniel Berehulak, was part of a winning series,]$_2$ [and shows James Dorbor, 8,]$_3$ [suspected of being infected with Ebola,]$_4$ [being carried by medical staff to an Ebola treatment center in Monrovia, Liberia.]$_5$

.....

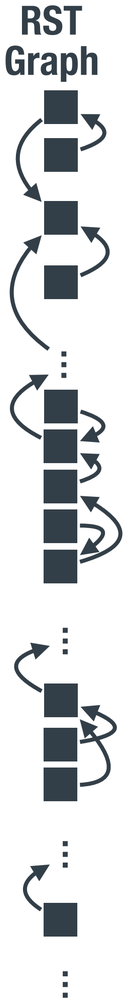
20. [The Pulitzer prizes,]$_1$ [awarded annually by Columbia University,]$_2$ [recognize extraordinary work in U.S. journalism, literature, drama, and other categories.]$_3$

.....

22. [Other winners of the coveted award included the St. Louis Post-Dispatch.]$_1$

.....

Sentence Selection [EDU Selection]



Coreference Mention Graph

Document

1. [It is one of the most prestigious honors]$_1$ [bestowed upon journalists and people in the arts.]$_2$
2. [And today, the Pulitzer prize for journalism went to The Post and Courier newspaper of Charleston, South Carolina,]$_1$ [which has a tiny staff of just 80 and a daily circulation of 85,000.]$_2$

……

5. [Winner:]$_1$ [This iconic photo by New York Times photographer Daniel Berehulak, was part of a winning series,]$_2$ [and shows James Dorbor, 8,]$_3$ [suspected of being infected with Ebola,]$_4$ [being carried by medical staff to an Ebola treatment center in Monrovia, Liberia.]$_5$

……

20. [The Pulitzer prizes,]$_1$ [awarded annually by Columbia University,]$_2$ [recognize extraordinary work in U.S. journalism, literature, drama, and other categories.]$_3$

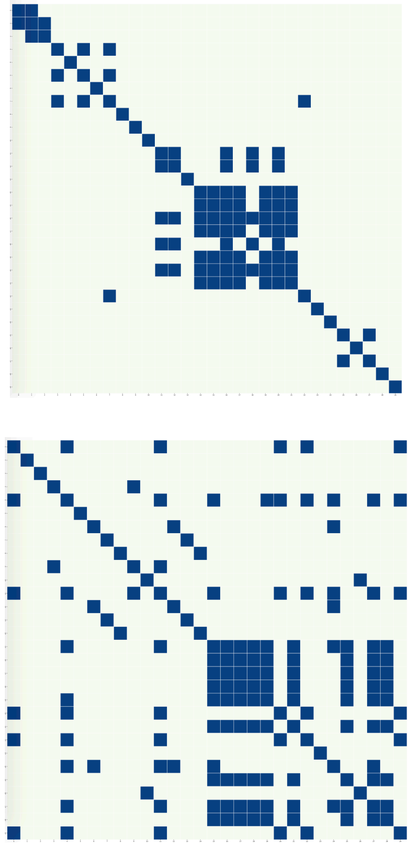
……

22. [Other winners of the coveted award included the St. Louis Post-Dispatch.]$_1$

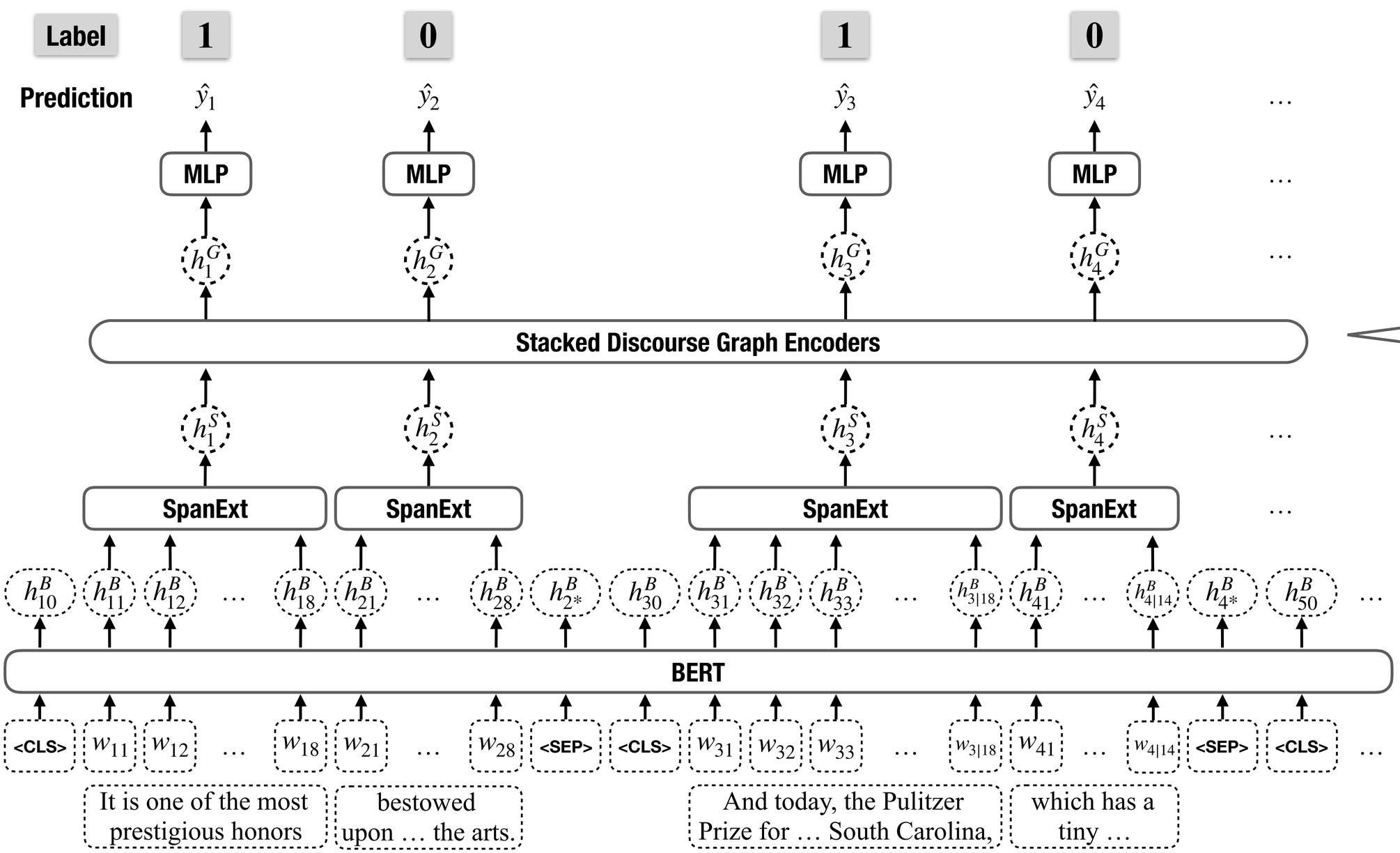
……

Coref Graph

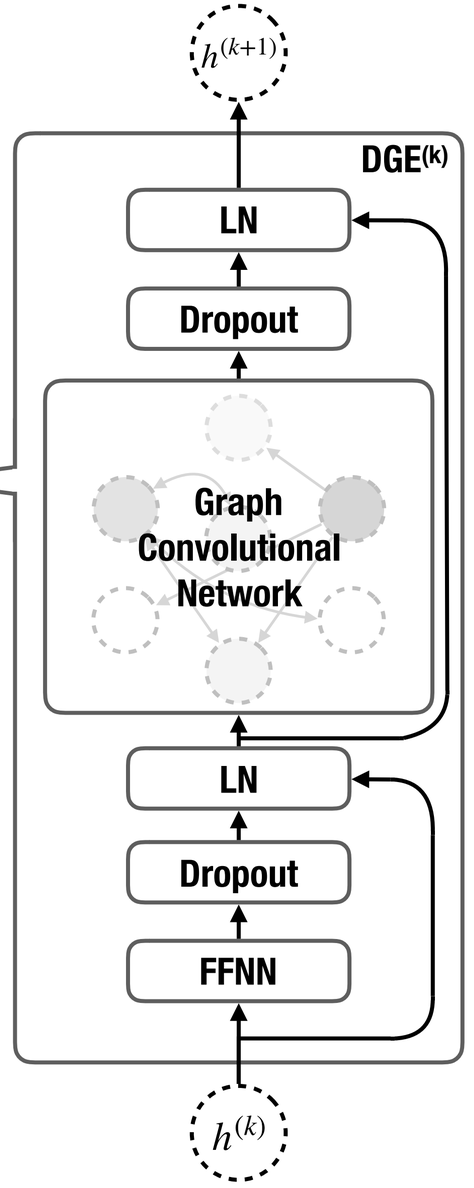
Build a graph based on entities and their mentions



Sentence Selection [EDU Selection] Mentions of ‘Pulitzer prizes’

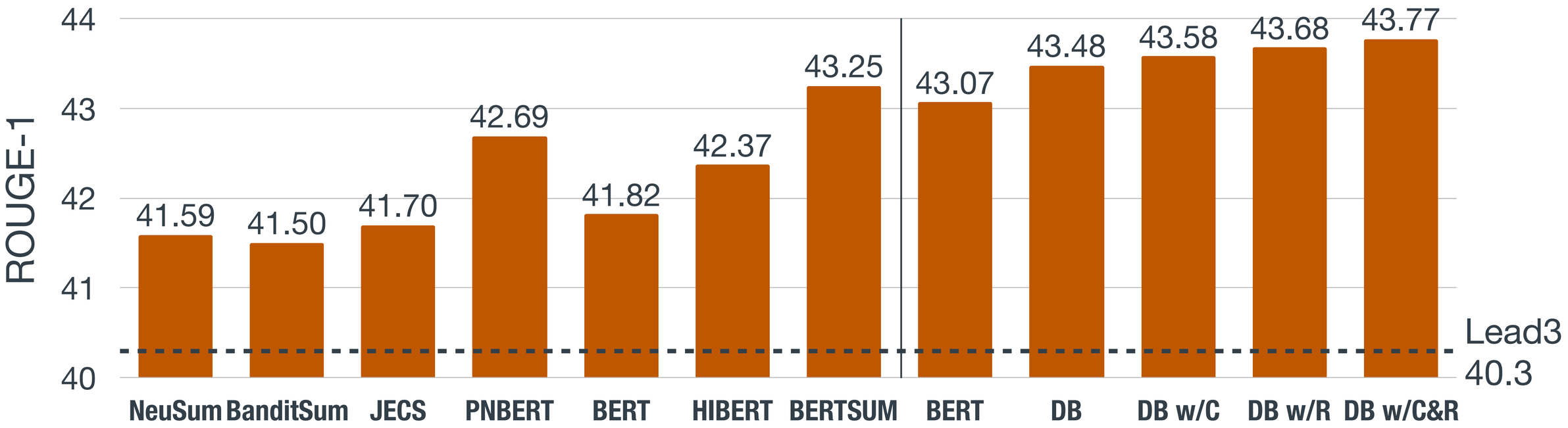


Model architecture of DiscoBERT



Discourse Graph Encoder Module

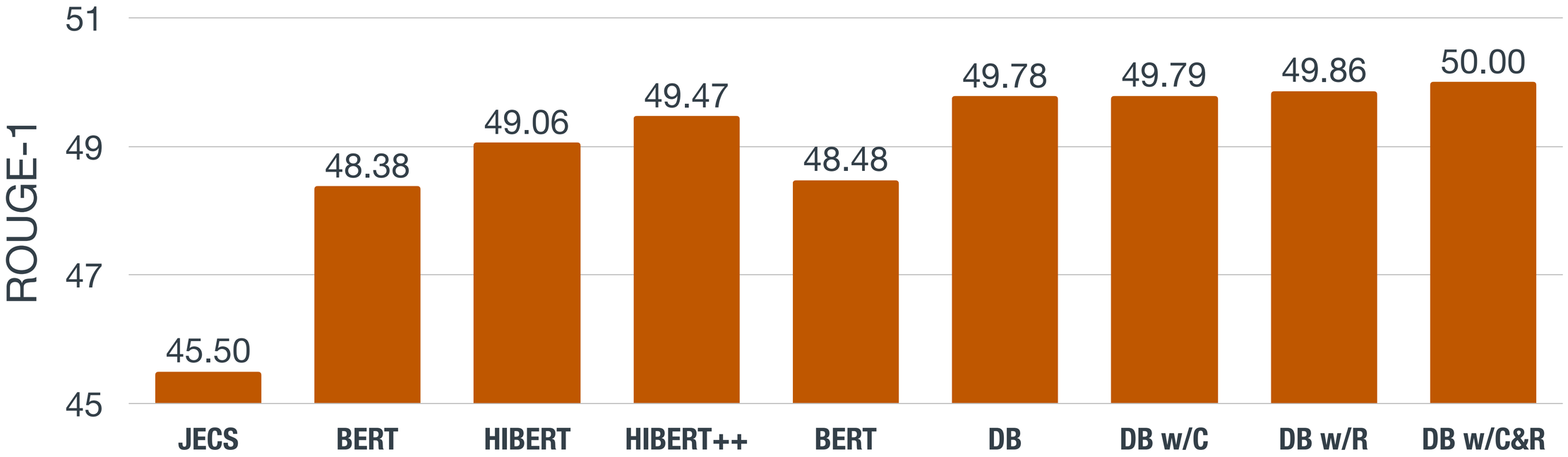
Experiments - CNNDM



- BERT-based models beat all non-BERT models by a significant margin
- DiscoBERT (**DB**) outperforms BERT
- DiscoBERT with Graphs (**C**: Coref & **R**: RST Discourse) works better than the vanilla **DB**

(Zhou et al., 2018; Dong et al., 2018; Xu and Durrett 2019; Zhong et al., 2019; Zhang et al., 2019; Liu 2019)

Experiments - NYT



BERT << {DB, DB w/C, DB w/R} < DB w/C&R

Discrete compression brings a huge jump;

Graphs might not always be that useful.

(Xu and Durrett 2019; Zhang et al., 2019)

Conclusion

Learning Discrete Compressions for Neural Extractive Text Summarization

Take-home Message

- Discrete compressions reduce the redundancy and improves the conciseness
- Interpretable learning-based compressions makes the system more customizable
- The injection of prior knowledge could help text summarization

Future Work

- Neural Template-based Text Summarization & Generation
  - Mining text patterns from the corpus
  - Assemble text pieces to coherent summaries
- Weakly Supervised Document Structure Induction
  - Using the section boundaries in the document
  - Inducing the document structure with pre-training techniques

Reference

- Taylor Berg-Kirkpatrick, Dan Gillick, and Dan Klein. ACL 2011. Jointly Learning to Extract and Compress.
- Yue Dong, Yikang Shen, Eric Crawford, Herke van Hoof, and Jackie Chi Kit Cheung. EMNLP 2018. BanditSum: Extractive Summarization as a Contextual Bandit.
- Greg Durrett, Taylor Berg-Kirkpatrick, and Dan Klein. ACL 2016. Learning-Based Single-Document Summarization with Compression and Anaphoricity Constraints.
- Katja Filippova, Enrique Alfonseca, Carlos A. Colmenares, Lukasz Kaiser, and Oriol Vinyals. EMNLP 2015. Sentence Compression by Deletion with LSTMs.
- Tsutomu Hirao, Yasuhisa Yoshida, Masaaki Nishino, Norihito Yasuda, and Masaaki Nagata. EMNLP 2013. Single-Document Summarization as a Tree Knapsack Problem.
- Junyi Jessy Li, Kapil Thadani, and Amanda Stent. SIGDIAL 2016. The role of discourse units in near-extractive summarization.

Reference

- Shashi Narayan, Shay B. Cohen, and Mirella Lapata. NAACL 2018. Ranking sentences for extractive summarization with reinforcement learning.
- Lu Wang, Hema Raghavan, Vittorio Castelli, Radu Florian, and Claire Cardie. ACL 2013. A Sentence Compression Based Framework to Query-Focused Multi-Document Summarization.
- Xingxing Zhang, Mirella Lapata, Furu Wei, and Ming Zhou. EMNLP 2018. Neural Latent Extractive Document Summarization.
- Ming Zhong, Pengfei Liu, Danqing Wang, Xipeng Qiu, and Xuanjing Huang. ACL 2019. Searching for effective neural extractive summarization: What works and what's next.
- Qingyu Zhou, Nan Yang, Furu Wei, Shaohan Huang, Ming Zhou, and Tiejun Zhao. ACL 2018. Neural Document Summarization by Jointly Learning to Score and Select Sentences.

Q&A

Thank you!