Summarization

1. Paper01： MatchSum （SOTA）
2. Paper02： GSum （SOTA）
3. Paper03： StructSum

zylei-AntNLP

Extractive Summarization as Text Matching

(ACL-20)

Ming Zhong[*], Pengfei Liu[*], Yiran Chen, Danqing Wang, Xipeng Qiu[†], Xuanjing Huang

Shanghai Key Laboratory of Intelligent Information Processing, Fudan University

School of Computer Science, Fudan University
825 Zhangheng Road, Shanghai, China

{mzhong18,pfliu14,yrchen19,dqwang18,xpqiu,xjhuang}@fudan.edu.cn

Motivation

1, Score and extract sentences(or smaller semantic unit) one by one.

2, As sequence labelling problem, make decisions for each sentence. result redundancy.

3, Auto-regressive decoder、Trigram Blocking（BertExt）

4, Sentence-level to summary-level (MatchSum), as a text matching problem(better to estimate semantic similarity).

A Dataset-dependent Analysis

We investigate the gap between sentence-level and summary-level methods on six benchmark datasets

Definition:

D = {s1,…,sn}, single document

C = {s1,…,sk | si ∈ D }, candidate summary

C* : gold summary

Sentence-level score: $$\mathrm{g}^{sen}(C) = \frac{1}{|C|}\sum_{s \in C} \mathrm{R}(\mathrm{s}, \mathrm{C}^*), \tag{1}$$

Summary-level score: $$\mathrm{g}^{sum}(C) = \mathrm{R}(C, C^*), \tag{2}$$

Pearl-Summary: has a lower sentence-level score but a higher summary-level score.

(eg. $\mathrm{g}^{\mathrm{sen}}(\mathrm{C}') > \mathrm{g}^{\mathrm{sen}}(\mathrm{C})$ while $\mathrm{g}^{\mathrm{sum}}(\mathrm{C}') < \mathrm{g}^{\mathrm{sum}}(\mathrm{C})$.)

Best-Summary: $$\hat{C} = \underset{C \in \mathcal{C}}{\operatorname{argmax}}\, \mathrm{g}^{sum}(C),$$

A Dataset-dependent Analysis

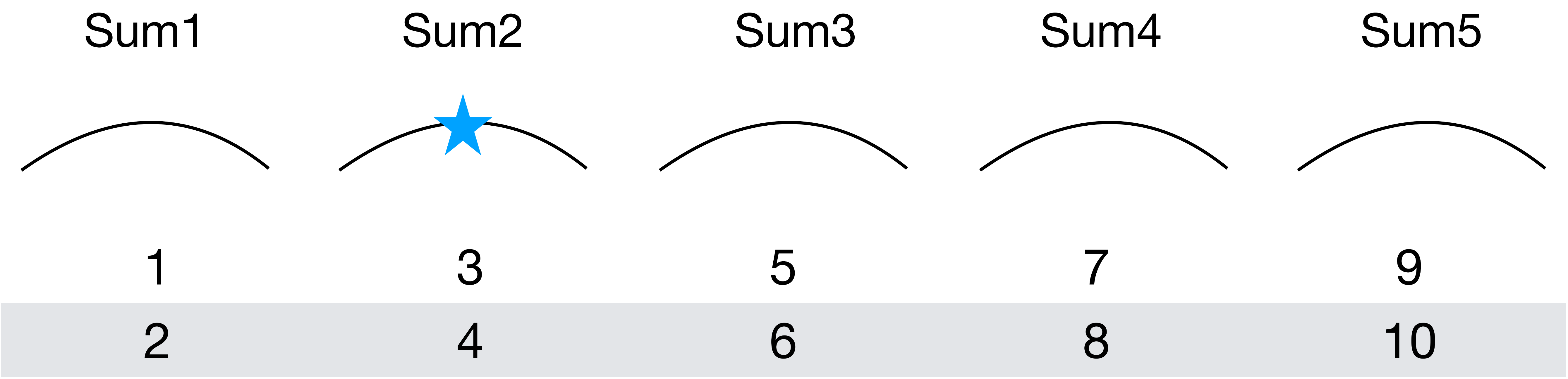
Datasets	Source	Type	# Pairs			# Tokens		# Ext
			Train	Valid	Test	Doc.	Sum.	
Reddit	Social Media	SDS	41,675	645	645	482.2	28.0	2
XSum	News	SDS	203,028	11,273	11,332	430.2	23.3	2
CNN/DM	News	SDS	287,084	13,367	11,489	766.1	58.2	3
WikiHow	Knowledge Base	SDS	168,126	6,000	6,000	580.8	62.6	4
PubMed	Scientific Paper	SDS	83,233	4,946	5,025	444.0	209.5	6
Multi-News	News	MDS	44,972	5,622	5,622	487.3	262.0	9

Table 1: Datasets overview. SDS represents single-document summarization and MDS represents multi-document summarization. The data in Doc. and Sum. indicates the average length of document and summary in the test set respectively. # Ext denotes the number of sentences should extract in different datasets.

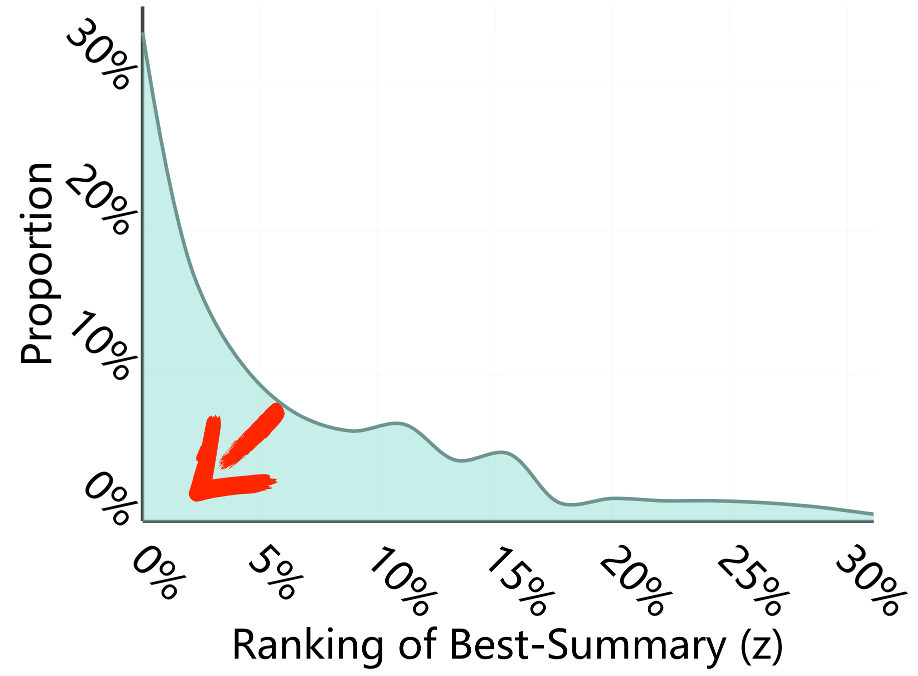
Five Candidates summaries

Best summary's index = 2

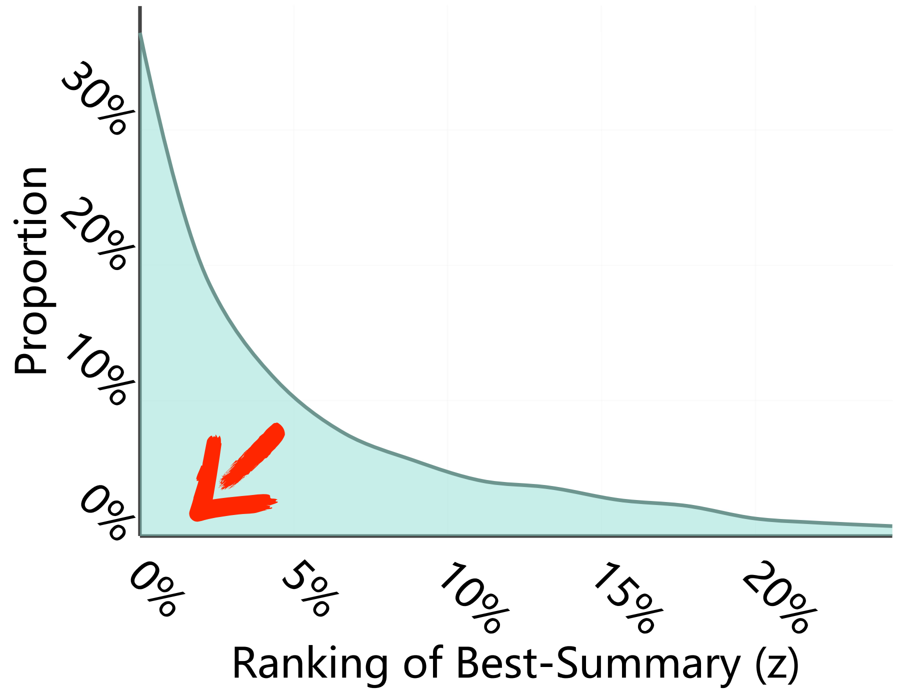
Belong to Pearl Summary



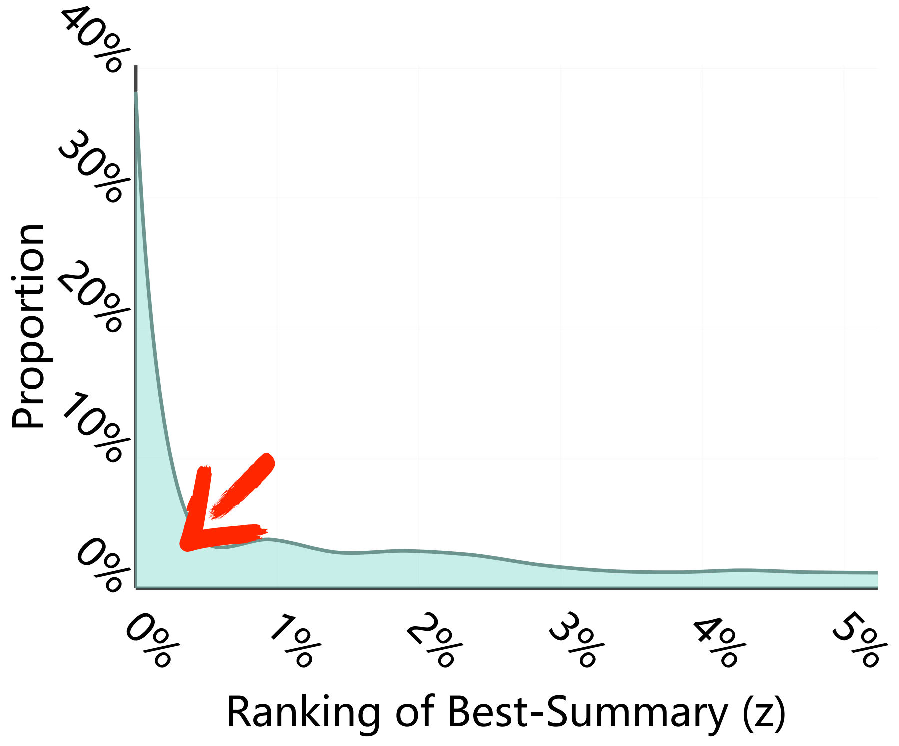
A Dataset-dependent Analysis



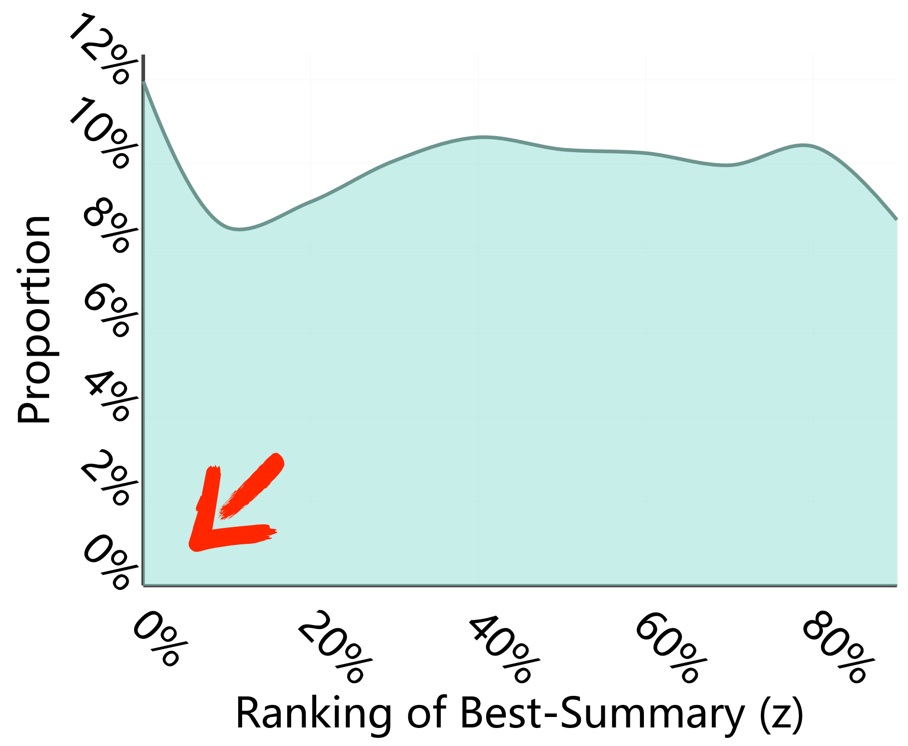
(a) Reddit



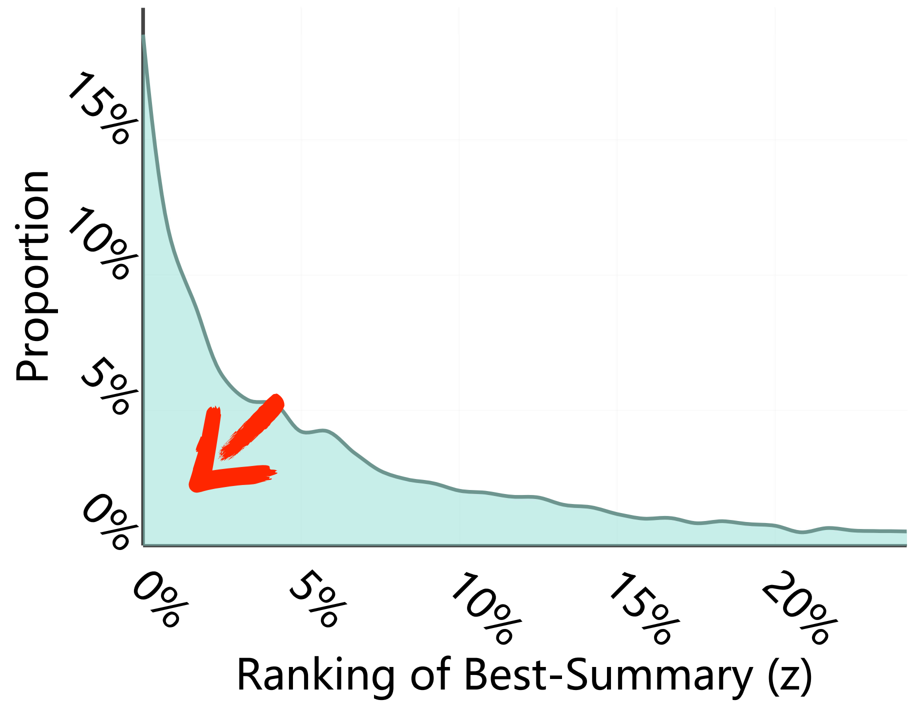
(b) XSum



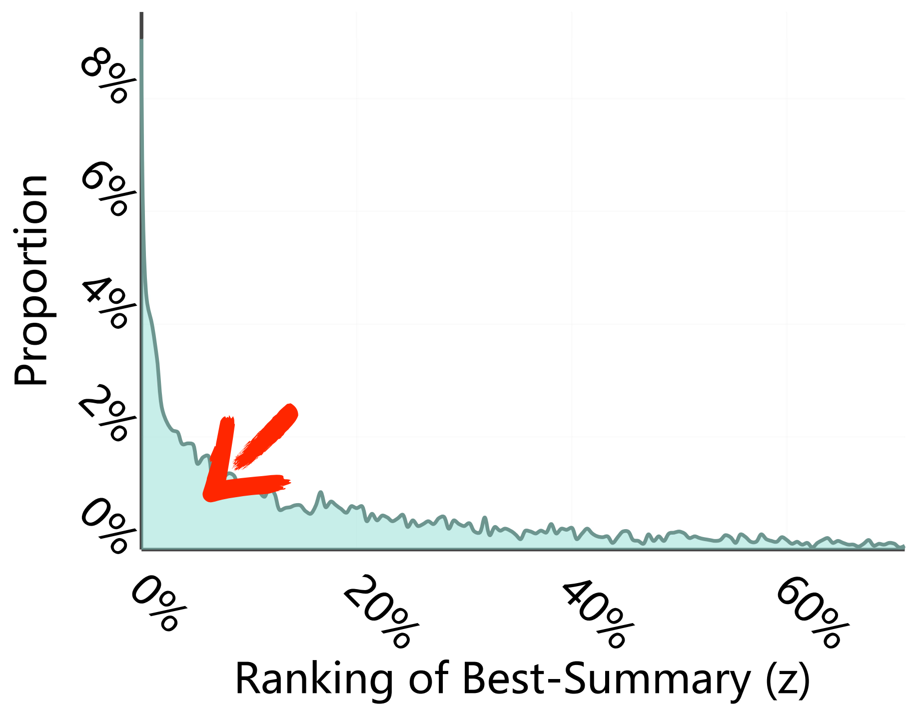
(e) PubMed



(f) Multi-News



(c) CNN/DM



(d) WikiHow

Figure 2: Distribution of $z(\%)$ on six datasets. Because the number of candidate summaries for each document is different (short text may have relatively few candidates), we use z / *number of candidate summaries* as the X-axis. The Y-axis represents the proportion of the best-summaries with this rank in the test set.

A Dataset-dependent Analysis

Calculating the potential gain between sentence-level and summary-level:

$$\alpha^{sen}(D) = \max_{C \in \mathcal{C}_D} \mathrm{g}^{sen}(C), \tag{3}$$

$$\alpha^{sum}(D) = \max_{C \in \mathcal{C}_D} \mathrm{g}^{sum}(C), \tag{4}$$

Gap:

$$\Delta(D) = \alpha^{sum}(D) - \alpha^{sen}(D). \tag{5}$$

$$\Delta(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{D \in \mathcal{D}} \Delta(D), \tag{6}$$

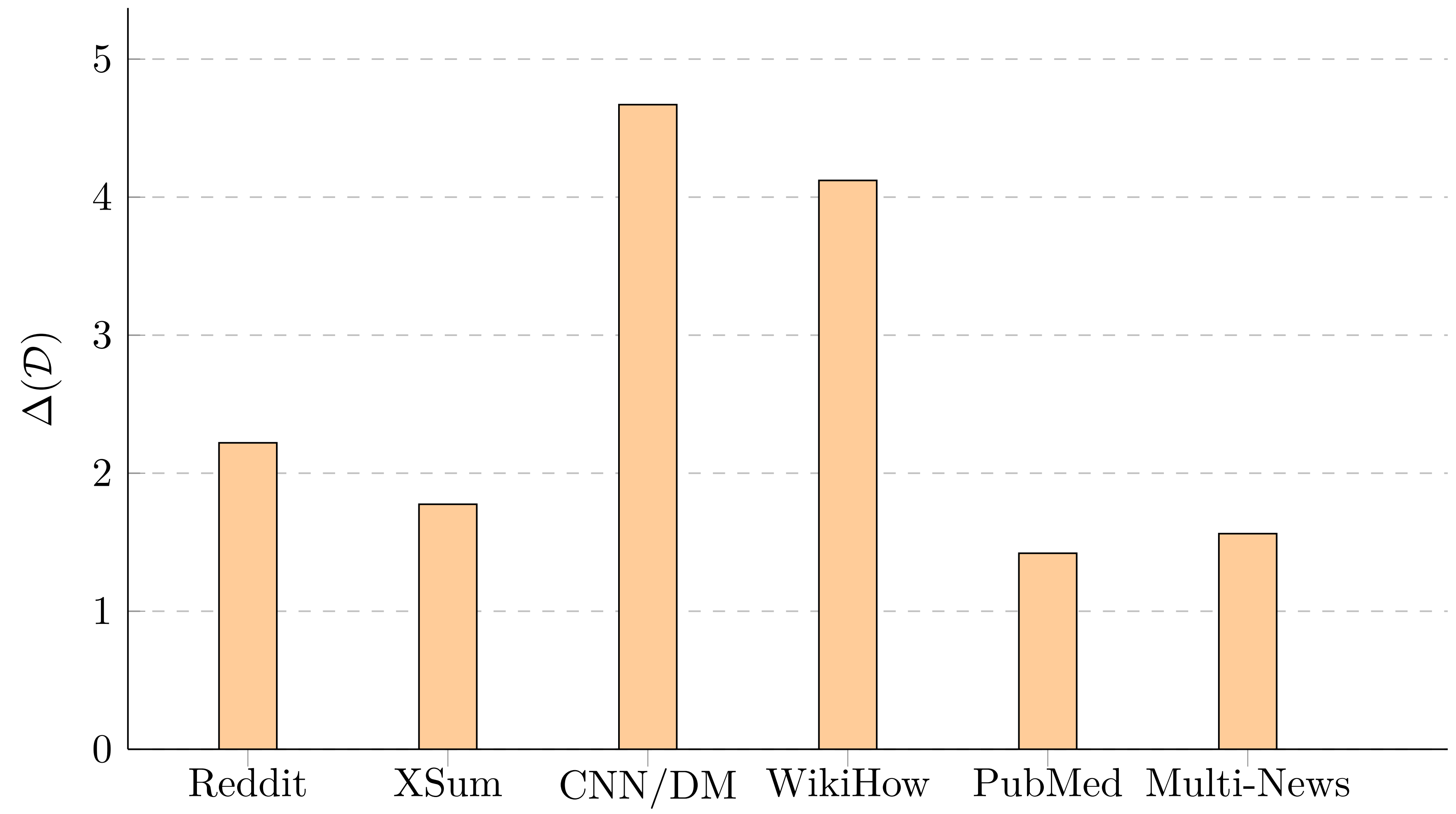


Figure 3: $\Delta(\mathcal{D})$ for different datasets.

Method

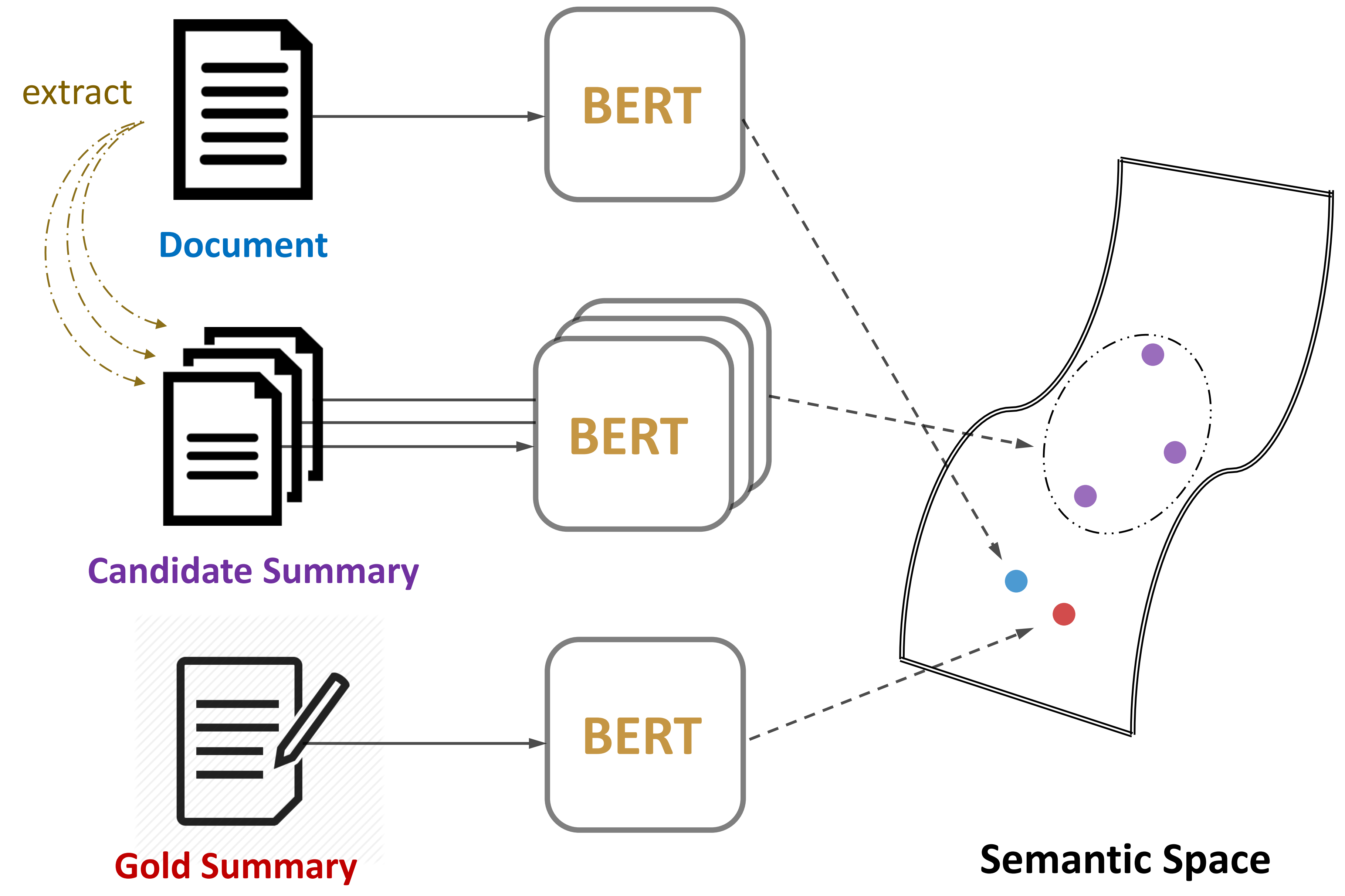
Extract candidate summary： Use BertSum （include pruning）

$$f(D,C) = \text{cosine}(\mathbf{r}D, \mathbf{r}C)$$

Inspired by siamese network structure, **Siamese-BERT**:

Document representation： **r**D

Candidates representation： **r**C



$$\hat{C} = \arg\max_{C \in \mathcal{C}} f(D, C).$$

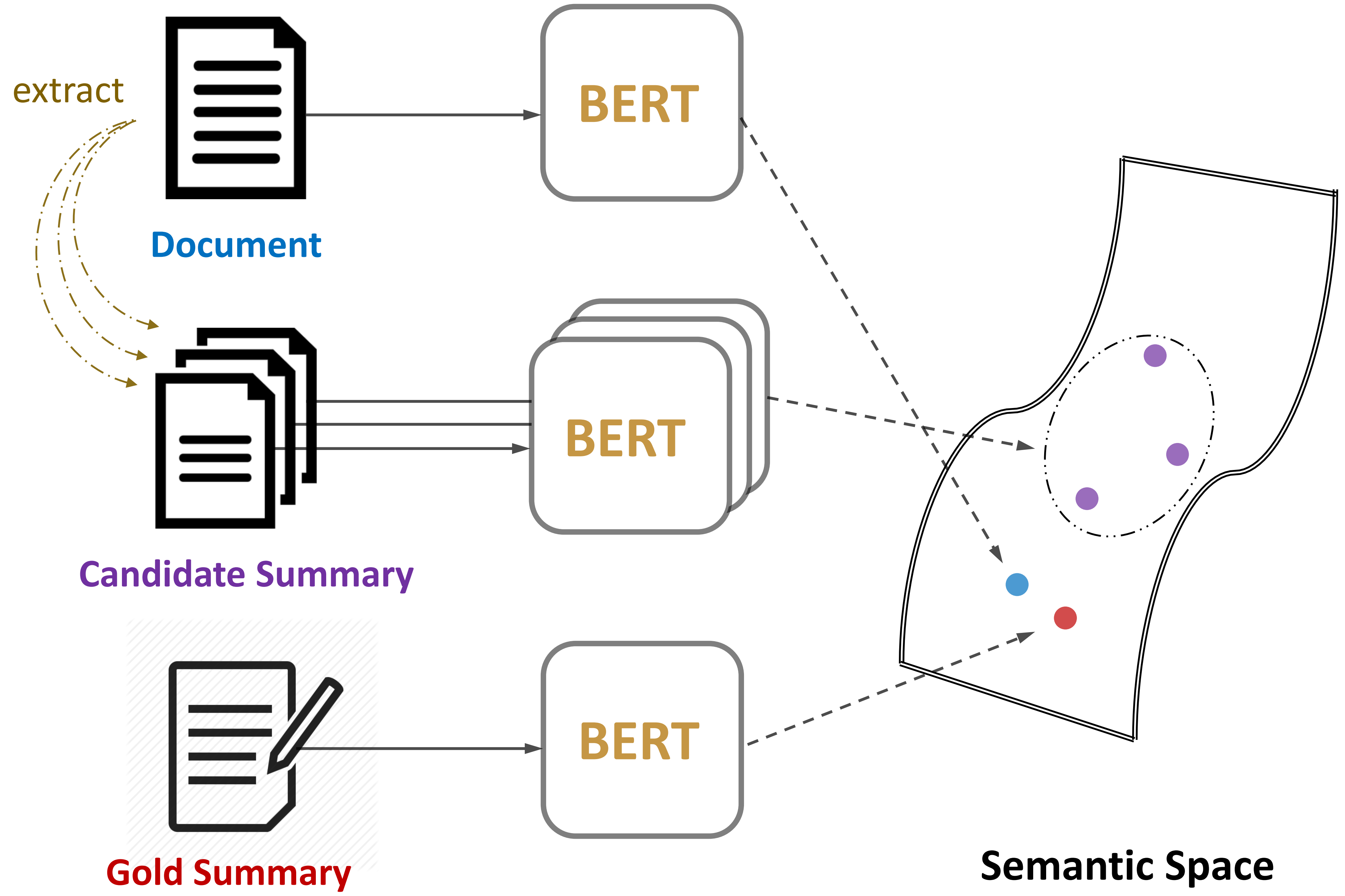
Method

Margin-based Triple Loss(update the weights):

$$L_1 = \max(0, f(D,C) - f(D,C^*) + \gamma_1)$$

$$L_2 = \max(0, f(D,C_j) - f(D,C_i) + (j - i) * \gamma_2) \quad (i < j)$$

$$L = L_1 + L_2.$$



$$\hat{C} = \arg\max_{C \in \mathcal{C}} f(D, C).$$

Experiment

Dataset（after pruning）：

	Reddit	XSum	CNN/DM	Wiki	PubMed	M-News
Ext	5	5	5	5	7	10
Sel	1, 2	1, 2	2, 3	3, 4, 5	6	9
Size	15	15	20	16	7	9

Table 2: Details about the candidate summary for different datasets. *Ext* denotes the number of sentences after we prune the original document, *Sel* denotes the number of sentences to form a candidate summary and *Size* is the number of final candidate summaries.

Model	R-1	R-2	R-L
LEAD	40.43	17.62	36.67
ORACLE	52.59	31.23	48.87
MATCH-ORACLE	51.08	26.94	47.22
BANDITSUM (Dong et al., 2018)	41.50	18.70	37.60
NEUSUM (Zhou et al., 2018)	41.59	19.01	37.98
JECS (Xu and Durrett, 2019)	41.70	18.50	37.90
HIBERT (Zhang et al., 2019b)	42.37	19.95	38.83
PNBERT (Zhong et al., 2019a)	42.39	19.51	38.69
PNBERT + RL	42.69	19.60	38.85
BERTEXT† (Bae et al., 2019)	42.29	19.38	38.63
BERTEXT† + RL	42.76	19.87	39.11
BERTEXT (Liu, 2019)	42.57	19.96	39.04
BERTEXT + Tri-Blocking	43.23	20.22	39.60
BERTSUM* (Liu and Lapata, 2019)	43.85	20.34	39.90
BERTEXT (Ours)	42.73	20.13	39.20
BERTEXT + Tri-Blocking (Ours)	43.18	20.16	39.56
MATCHSUM (BERT-base)	44.22	20.62	40.38
MATCHSUM (RoBERTa-base)	**44.41**	**20.86**	**40.55**

Result on CNN/DM test set

Short Summary :

Model	R-1	R-2	R-L
Reddit			
BERTEXT (Num = 1)	21.99	5.21	16.99
BERTEXT (Num = 2)	23.86	5.85	19.11
MATCHSUM (Sel = 1)	22.87	5.15	17.40
MATCHSUM (Sel = 2)	24.90	5.91	20.03
MATCHSUM (Sel = 1, 2)	**25.09**	**6.17**	**20.13**
XSum			
BERTEXT (Num = 1)	22.53	4.36	16.23
BERTEXT (Num = 2)	22.86	4.48	17.16
MATCHSUM (Sel = 1)	23.35	4.46	16.71
MATCHSUM (Sel = 2)	24.48	4.58	18.31
MATCHSUM (Sel = 1, 2)	**24.86**	**4.66**	**18.41**

Result on Reddit and XSum test set

Experiment

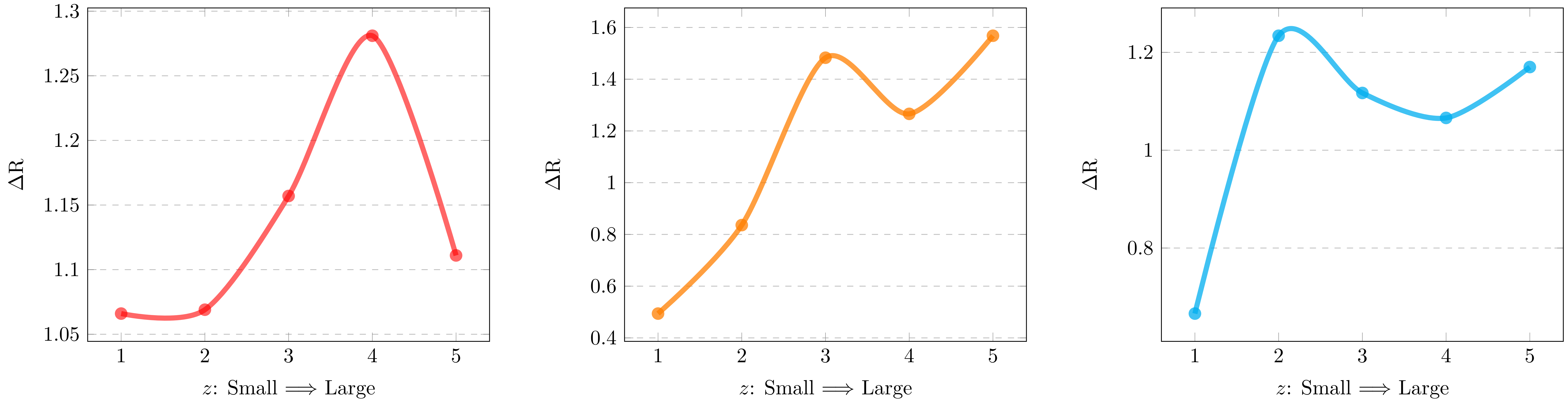
Long Summary：

Model	WikiHow			PubMed			Multi-News		
	R-1	R-2	R-L	R-1	R-2	R-L	R-1	R-2	R-L
LEAD	24.97	5.83	23.24	37.58	12.22	33.44	43.08	14.27	38.97
ORACLE	35.59	12.98	32.68	45.12	20.33	40.19	49.06	21.54	44.27
MATCH-ORACLE	35.22	10.55	32.87	42.21	15.42	37.67	47.45	17.41	43.14
BERTEXT	30.31	8.71	28.24	41.05	14.88	36.57	45.80	16.42	41.53
+ 3gram-Blocking	30.37	8.45	28.28	38.81	13.62	34.52	44.94	15.47	40.63
+ 4gram-Blocking	30.40	8.67	28.32	40.29	14.37	35.88	45.86	16.23	41.57
MATCHSUM (BERT-base)	**31.85**	**8.98**	**29.58**	**41.21**	**14.91**	**36.75**	**46.20**	**16.51**	**41.89**

Table 5: Results on test sets of WikiHow, PubMed and Multi-News. MATCHSUM beats the state-of-the-art BERT model with Ngram Blocking on all different domain datasets.

Analysis

Dataset Splitting Testing :



(a) XSum (b) CNN/DM (c) WikiHow

Figure 4: Datasets splitting experiment. We split test sets into five parts according to z described in Section 3.2. The X-axis from left to right indicates the subsets of the test set with the value of z from small to large, and the Y-axis represents the ROUGE improvement of MATCHSUM over BERTEXT on this subset.

Analysis

Comparison Across Dataset:

$$\Delta(D)^* = \mathrm{g}^{sum}(C_{MS}) - \mathrm{g}^{sum}(C_{BE}),$$

$$\Delta(\mathcal{D})^* = \frac{1}{|\mathcal{D}|}\sum_{D\in\mathcal{D}} \Delta(D)^*,$$

$$\psi(\mathcal{D}) = \Delta(\mathcal{D})^* / \Delta(\mathcal{D}), \qquad (14)$$

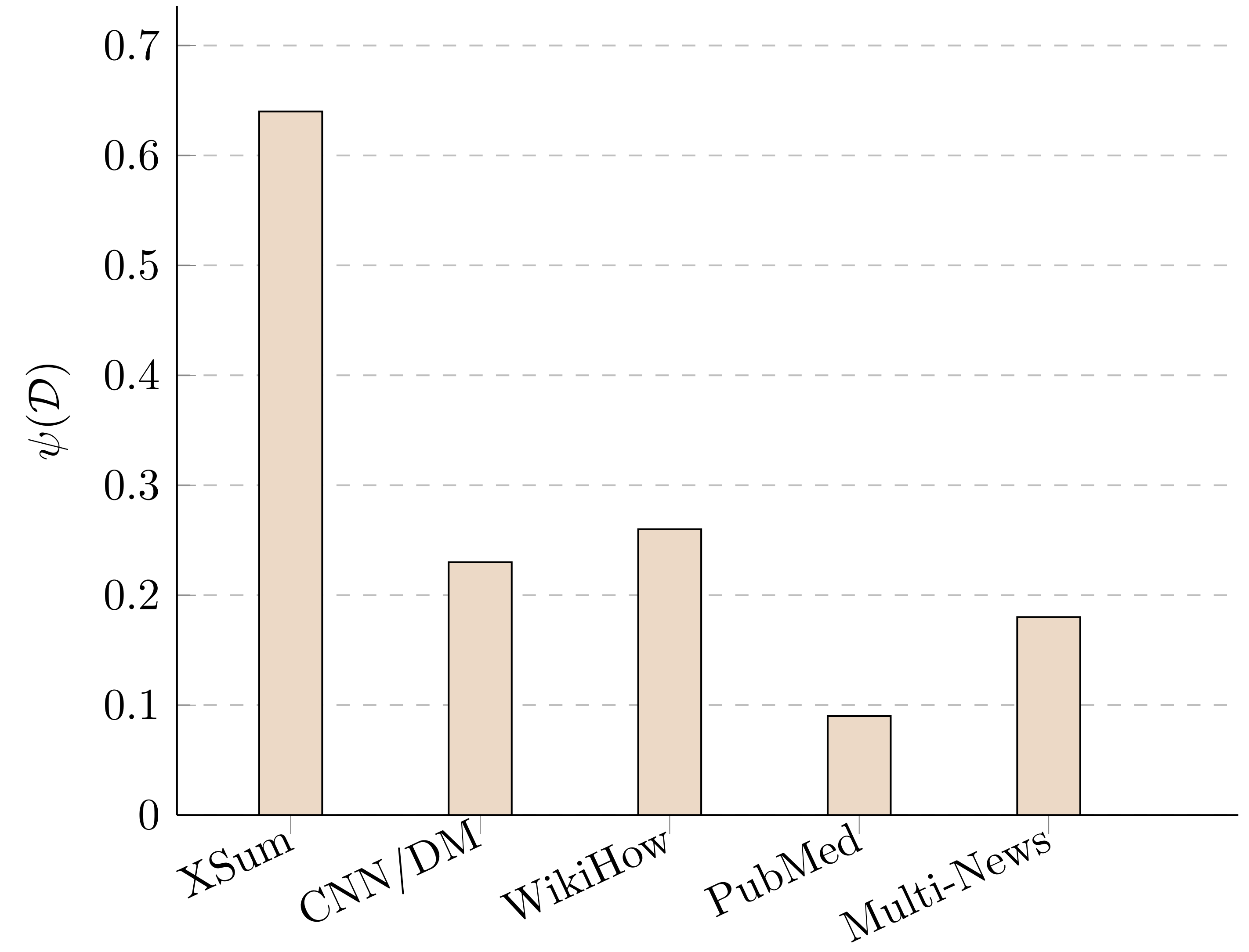


Figure 5: ψ of different datasets. Reddit is excluded because it has too few samples in the test set.

Conclusion

1, Propose a novel summary-level framework

2, Investigate whether extractive models must do summary-level extraction, quantify the inherent gap between sentence-level and summary-level methods

3, Achieved superior performance compared with strong baselines on six benchmark datasets even with base BERT.

GSum: A General Framework for Guided Neural Abstractive Summarization

(NAACL-21)

Zi-Yi Dou, Pengfei Liu, Hiroaki Hayashi, Zhengbao Jiang, Graham Neubig
Language Technologies Institute, Carnegie Mellon University
{zdou, pliu3, hiroakih, zhengbaj, gneubig}@cs.cmu.edu

Motivation

1， Compared with extractive algorithms， Abstractive Summarization are more flexible

2， Problems in Abstractive algorithms： unfaithful summaries ； difficult to control the content

3， previous guided methods： length； keywords； retrieve and reference relevant summaries

Work	Guidance Form			
	Tokens	**Triples**	**Sentences**	**Summaries**
Kikuchi et al. (2016)	✓ (length tokens)	✗	✗	✗
Cao et al. (2018)	✗	✗	✗	✓ (retrieved sums.)
Li et al. (2018)	✓ (keywords)	✗	✗	✗
Liu et al. (2018a)	✗	✗	✓ (highlighted sents.)	✗
Liu et al. (2018b)	✓ (length tokens)	✗	✗	✗
Fan et al. (2018)	✓ (length, entity, style tokens)	✗	✗	✗
Zhu et al. (2020)	✗	✓ (relations)	✗	✗
Jin et al. (2020)	✗	✓ (relations)	✗	✗
Saito et al. (2020)	✓ (keywords)	✗	✓ (highlighted sents.)	✗
Ours	✓ (keywords)	✓ (relations)	✓ (highlighted sents.)	✓ (retrieved sums.)

Methods

Select guidance:

Training: Oracle

Testing: Auto Ext/User-specified

(BertExt/MatchSum/BertAbs)

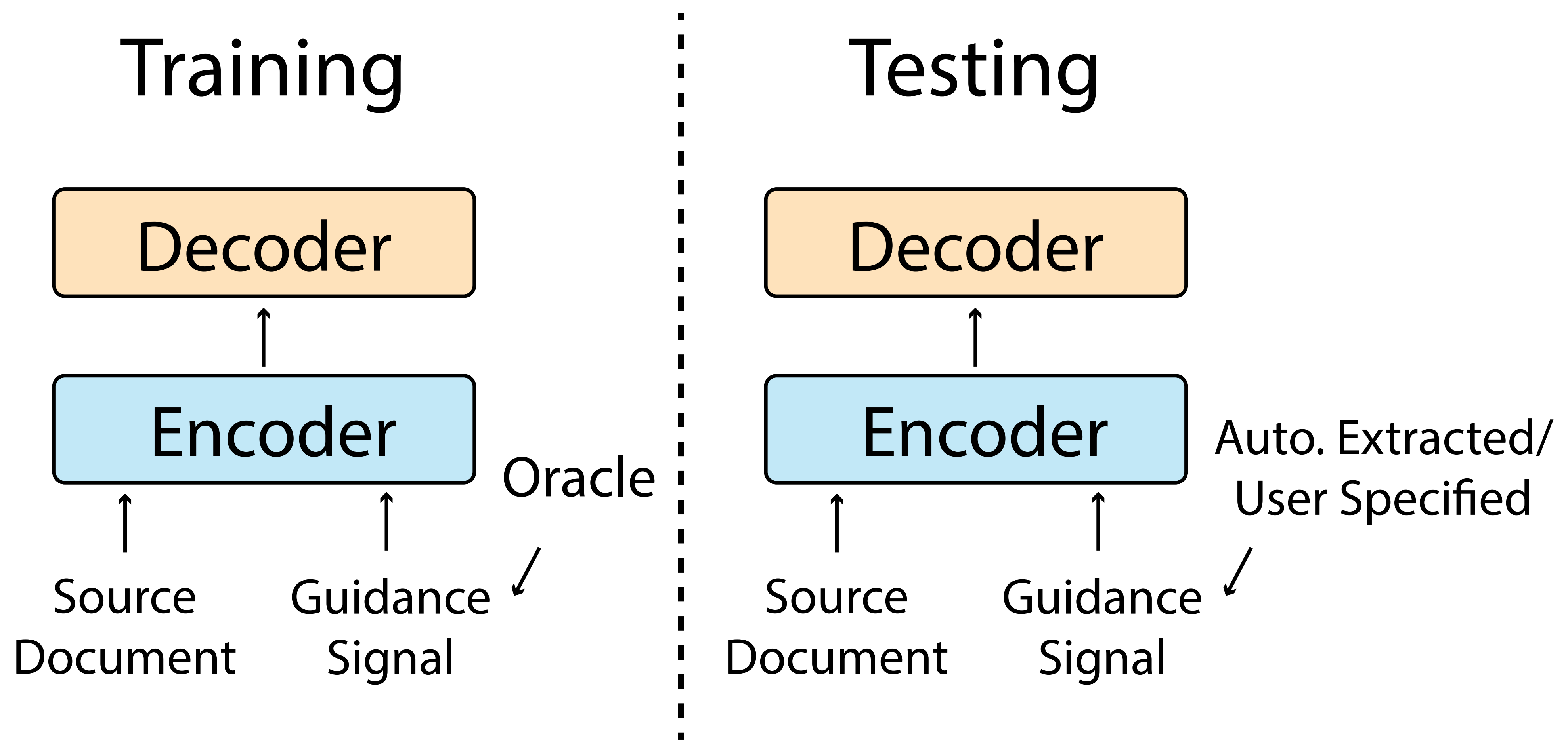


Figure 1: Our framework generates summaries using both the source document and separate guidance signals. We use an oracle to select guidance during training and use automatically extracted or user-specified guidance at test time.

Methods

Neural abstractive summarization:

$$\arg\max_{\theta} \sum_{\langle \mathbf{x}^i, \mathbf{y}^i \rangle \in \langle \mathcal{X}, \mathcal{Y} \rangle} \log p(\mathbf{y}^i \mid \mathbf{x}^i; \theta).$$

with guidance:

$$\arg\max_{\theta} \sum_{\langle \mathbf{x}^i, \mathbf{y}^i, \mathbf{g}^i \rangle \in \langle \mathcal{X}, \mathcal{Y}, \mathcal{G} \rangle} \log p(\mathbf{y}^i \mid \mathbf{x}^i, \mathbf{g}^i; \theta).$$

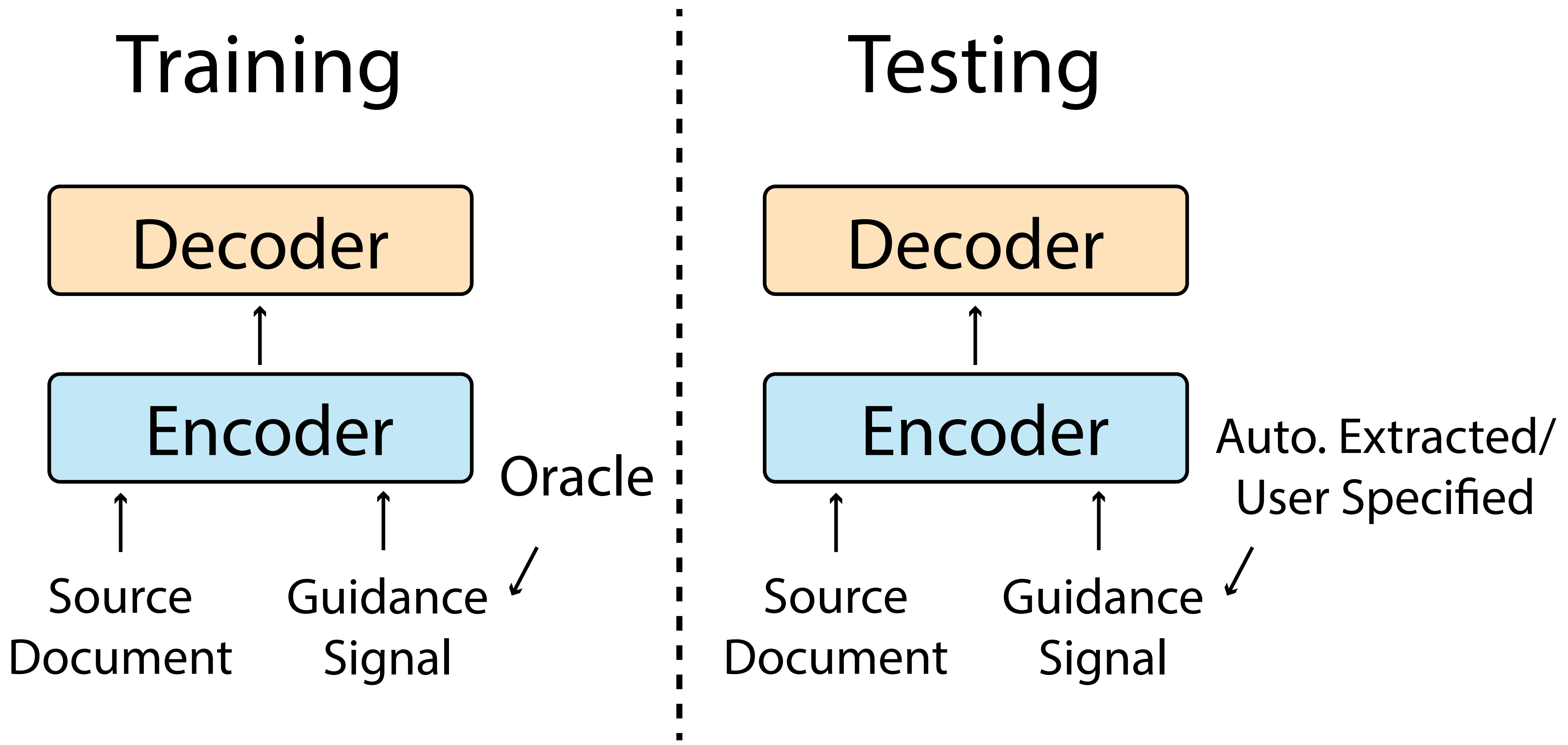
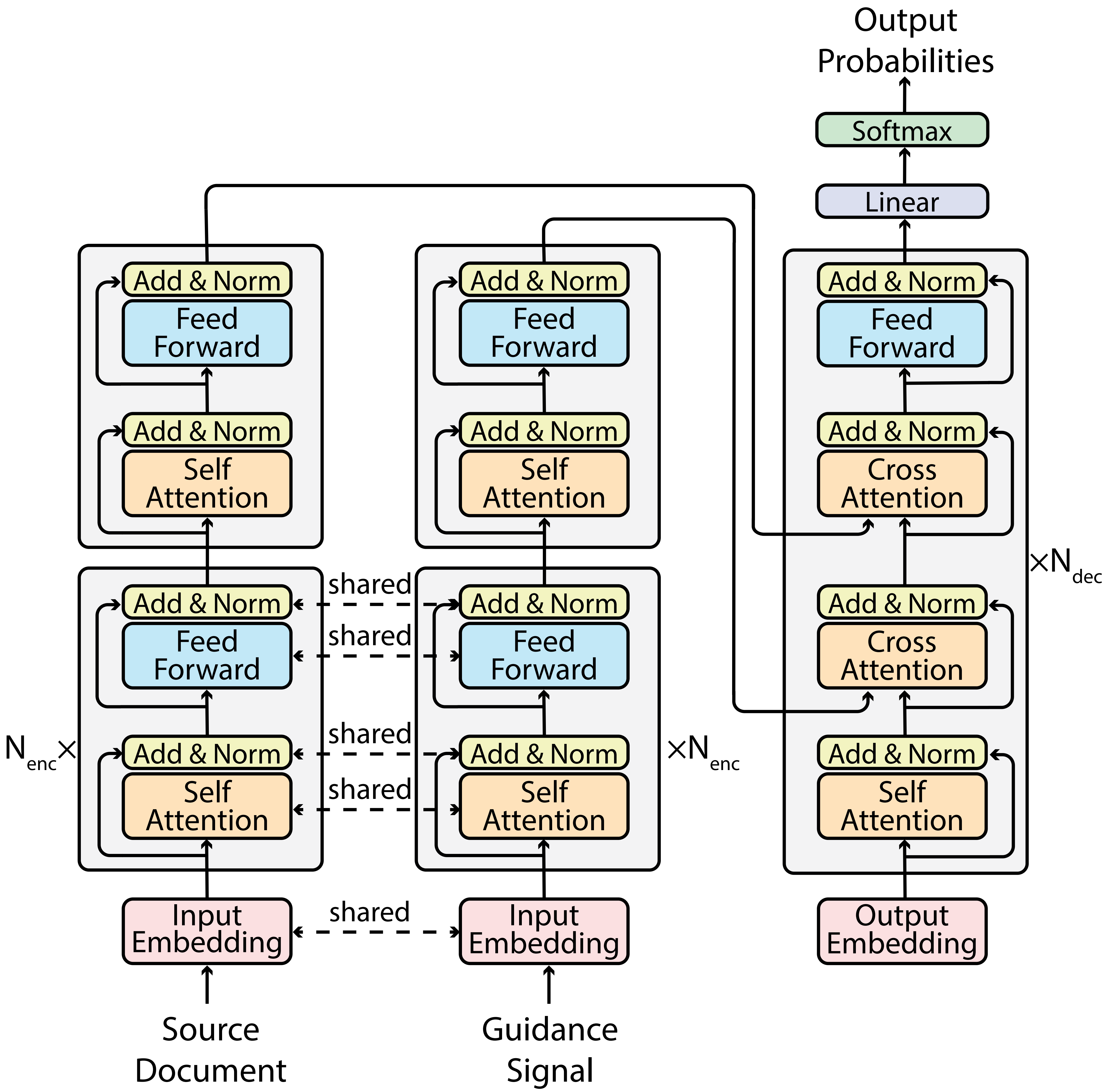


Figure 1: Our framework generates summaries using both the source document and separate guidance signals. We use an oracle to select guidance during training and use automatically extracted or user-specified guidance at test time.

Methods

Backbone： Transformer

instantiated： BERT or BART



Methods

Share the parameters of the bottom N layers

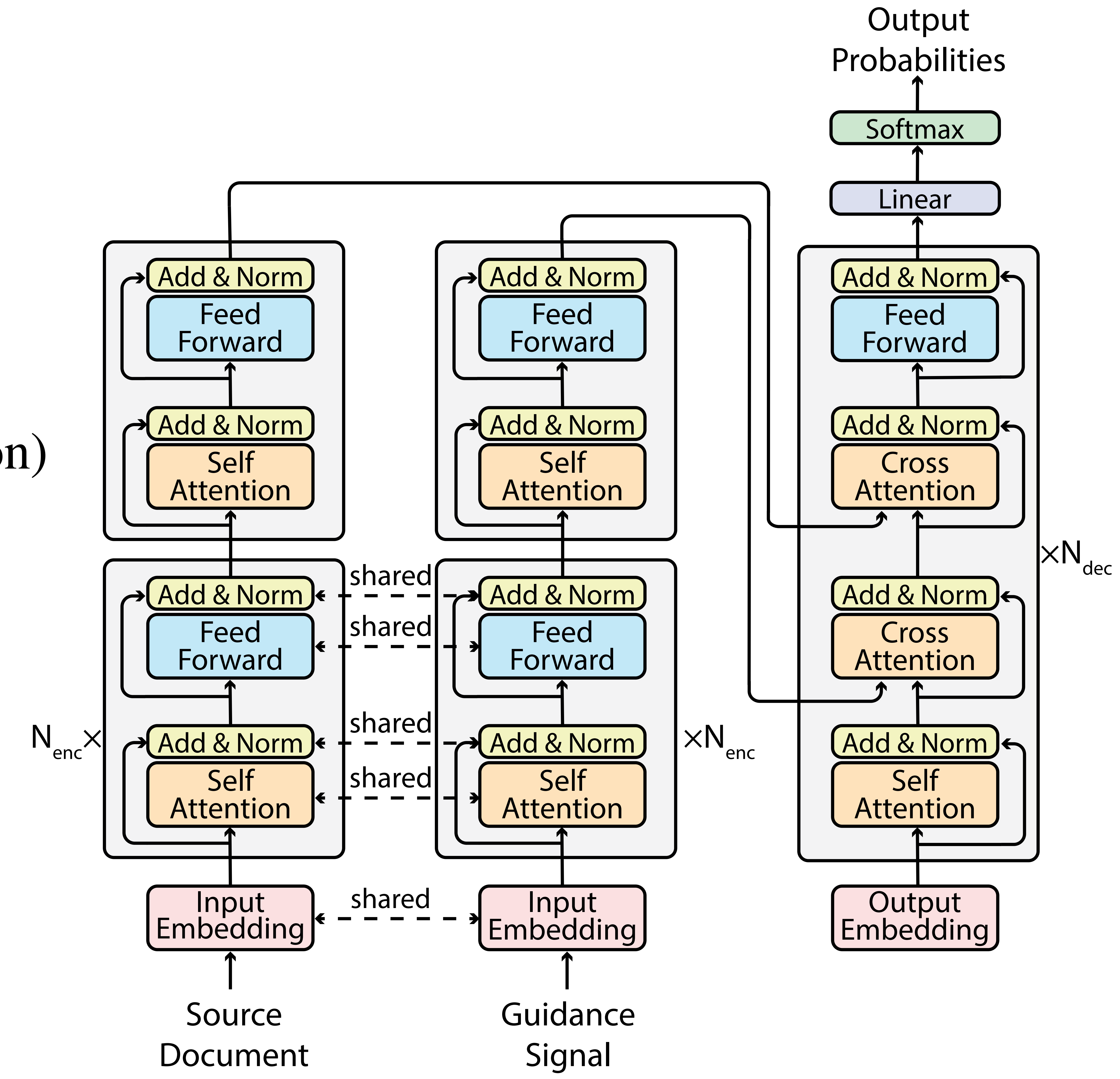
Because:

1， reduce computation and memory(computation)

2， conjecture the differences of two input

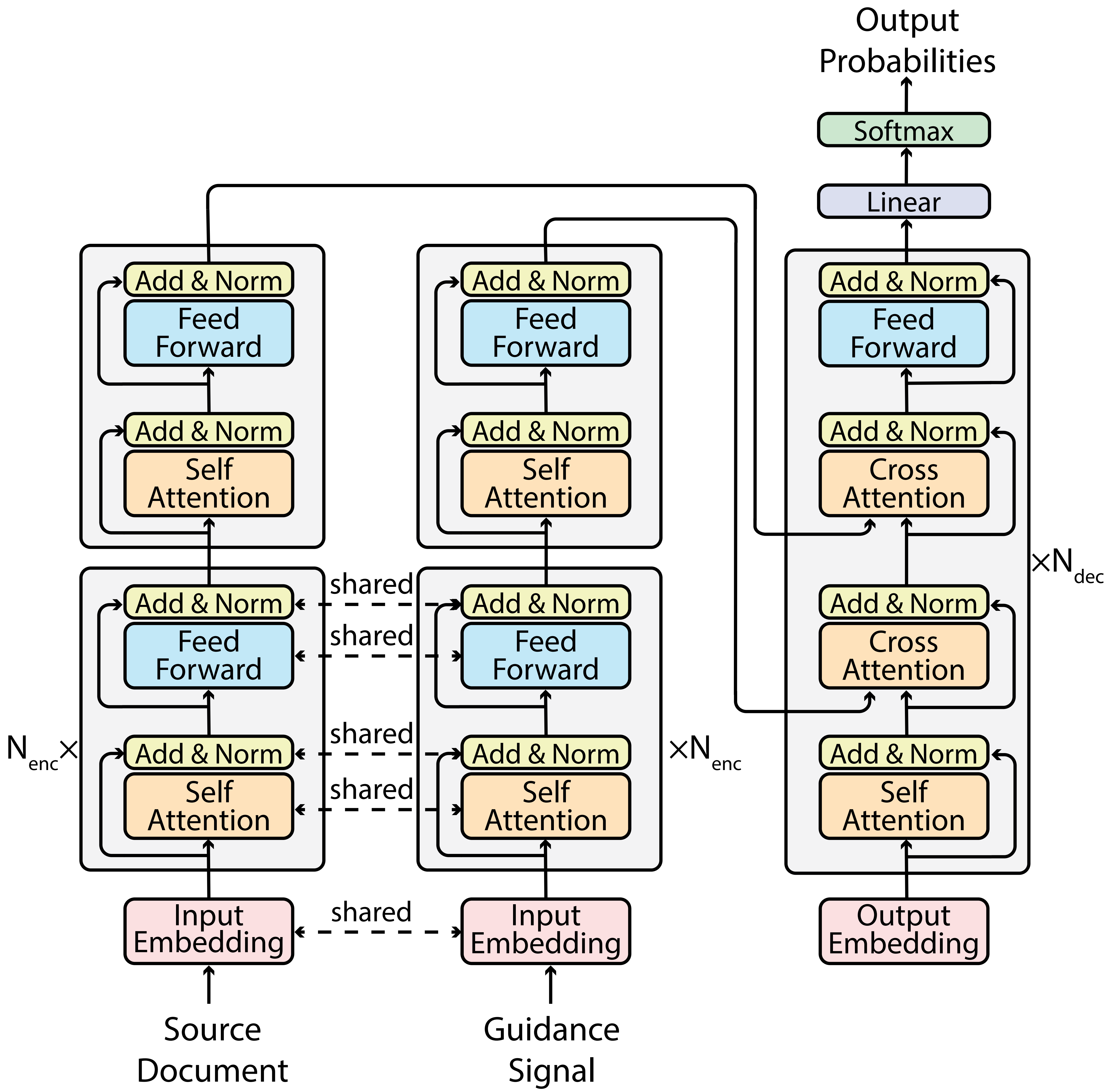
are high-level, will capture at top layers

(Capture at top layers)



Methods

Guidances choices:



Experiment

Datasets:

Dataset	Source	#Pairs			#Tokens		#Ext
		Train	Valid	Test	Doc.	Sum.	
Reddit	Social Media	41,675	645	645	482.2	28.0	2
XSum	News	203,028	11,273	11,332	430.2	23.3	2
CNN/DM	News	287,084	13,367	11,489	766.1	58.2	3
WikiHow	Knowledge Base	168,126	6,000	6,000	580.8	62.6	4
NYT	News	44,382	5,523	6,495	1183.2	110.8	4
PubMed	Scientific Paper	83,233	4,946	5,025	444.0	209.5	6

Experiment

Performance of different Guidance signals:

(On CNN/DM、on test time）

1，the model performance has the potential to be further improved given a better guidance prediction model；

2，the model does learn to depend on the guidance signals

Model	**Guide**	**R-1**	**R-2**	**R-L**
BertExt* (Base)	-	43.25	20.24	39.63
BertAbs*	-	41.72	19.39	38.76
BertAbs (Ours)	-	41.58	18.99	38.56
Ours				
BertAbs + Sentence	Auto.	43.78	20.66	40.66
	Oracle	55.18	32.54	52.06
BertAbs + Keyword	Auto.	42.21	19.36	39.23
	Oracle	45.08	22.22	42.07
BertAbs + Relation	Auto.	41.40	18.66	38.40
	Oracle	45.96	23.09	42.92
BertAbs + Retrieve	Auto.	40.88	18.24	37.99
	Oracle	43.69	20.53	40.71

Experiment

Model	**R-1**	**R-2**	**R-L**
Oracle	55.76	33.22	51.83
Extractive			
BertExt (Base)*	43.25	20.24	39.63
BertExt (Large)*	43.85	20.34	39.90
MatchSum†	44.41	20.86	40.55
Abstractive			
BertAbs*	41.72	19.39	38.76
BertAbs (Ours)	41.58	18.99	38.56
BertExtAbs*	42.13	19.60	39.18
BART ‡	44.16	21.28	40.90
BART (Ours)	44.66	21.53	41.35
Ours			
BertAbs + BertExt	43.78	20.66	40.66
BART + MatchSum	**45.94**	**22.32**	**42.48**

Table 4: Comparisons with state-of-the-art models on CNN/DM. The highest numbers are in **bold**. Marked results are from Liu and Lapata (2019)*, Zhong et al. (2020)†, Lewis et al. (2019)‡.

BART + MatchSum：

BART model + Guidance signal

（chose Highlight sentences as Guidance with MatchSum）

Experiment

Performance on Other Datasets:

Model	**Reddit**			**XSum**			**WikiHow**			**PubMed**			**NYT**		
	R-1	**R-2**	**R-L**	**R-1**	**R-2**	**R-L**	**R-1**	**R-2**	**R-L**	**R-1**	**R-2**	**R-L**	**R-1**	**R-2**	**R-L**
Oracle	36.21	13.74	28.93	29.79	8.81	22.66	35.59	12.98	32.68	45.12	20.33	40.19	58.44	38.39	50.00
Extractive															
BertExt (Base)	23.86	5.85	19.11	22.86	4.48	17.16	30.40	8.67	28.32	40.29	14.37	35.88	45.98	25.29	42.46
MatchSum	25.09	6.17	20.13	24.86	4.66	18.41	31.85	8.98	29.58	41.21	14.91	36.75	46.98	26.67	43.62
Bert-Based															
BertAbs	**26.92**	6.35	19.81	38.76	**16.33**	**31.15**	38.16	15.06	34.71	36.04	12.16	29.02	49.94	31.44	46.67
Ours (BertAbs + MatchSum)	26.89	**6.75**	**20.35**	**38.77**	16.14	30.96	**38.29**	**15.10**	**34.80**	**37.82**	**12.32**	**30.53**	**50.50**	**31.57**	**47.24**
BART-Based															
BART	**35.00**	**12.89**	**27.96**	**45.51**	**21.94**	**36.75**	41.46	**17.80**	39.89	44.72	16.48	41.00	54.13	35.15	47.00
Ours (BART + MatchSum)	34.52	12.71	27.58	45.40	21.89	36.67	**41.74**	17.73	**40.09**	**45.09**	**16.72**	**41.32**	**54.27**	**35.37**	**47.63**

Table 5: Results of our model guided with highlighted sentences on five datasets. Highest numbers in each section are in **bold**. We use MatchSum to predict the guidance at test time. Extractive results are from Zhong et al. (2020).

Experiment

Analysis Novel n-grams on CNN/DM:

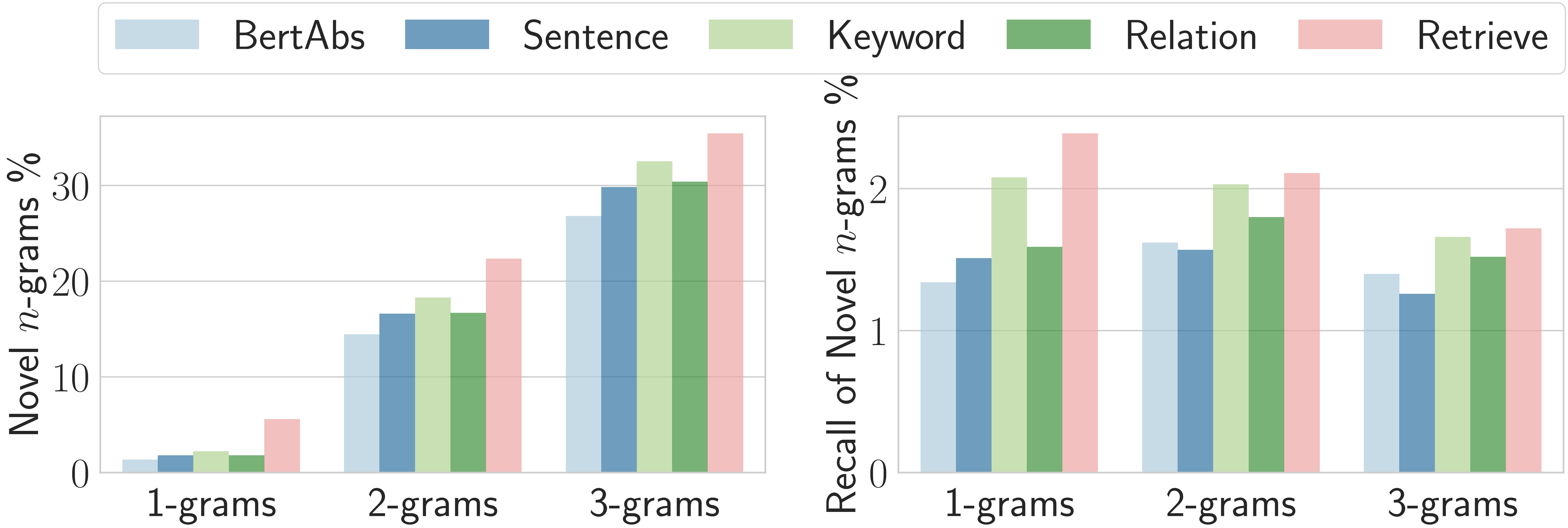


Figure 3: Our model can generate more novel words and achieve higher recall of novel words in the gold reference compared with baseline.

Experiment

Complementarity of Different Guidance Signals:

Win [%]				Combined
Sentence	**Keyword**	**Relation**	**Retrieve**	**R-1/R-2/R-L**
39.28	19.55	21.12	20.05	48.30/25.25/45.15

Table 6: No guidance signals can outperform all the other ones for all the test data, and aggregating the best outputs of the four guided models achieves significant improvements over the best single guided model (43.78/20.66/40.66 R-1/R-2/R-L scores).

	Sentence	**Keyword**	**Relation**	**Retrieve**
Sentence	43.78	46.11	46.20	46.27
Keyword	-	42.21	44.39	44.35
Relation	-	-	41.40	44.60
Retrieve	-	-	-	40.88

Table 7: Combining the best outputs of each pair of guidance signals leads to improvements (in terms of ROUGE-1), indicating every pair of guidance complements each other. The underlined results are the model performance without combinations.

Experiment

Controllability and Faithfulness of Generated Summaries:

BertAbs	Ours			
	Sentence	**Keyword**	**Relation**	**Retrieve**
2.117	2.393*	2.347*	2.303*	2.310*

Table 9: Human evaluation of the faithfulness of different model outputs. * indicates significant improvements ($p < 0.001$) over baseline with using bootstrap.

Train	**Test**	**R-1**	**R-2**	**R-L**
Oracle	Auto	43.78	20.66	40.66
	Oracle	55.18	32.54	52.06
Auto	Auto	41.61	19.04	38.65
	Oracle	43.07	20.79	40.13

Table 10: Using automatically constructed guidance during training degrades the performance significantly.

Conclusion

1, Propose a general framework for guided neural summarization

2, Investigate four types of guidance signals and achieve SOTA performance

3, Demonstrate the complementarity of the four guidance signals

StructSum: Summarization via Structured Representations

(ArXiv-21)

Vidhisha Balachandran Artidoro Pagnoni Jay Yoon Lee

Dheeraj Rajagopal Jaime Carbonell Yulia Tsvetkov

School of Computer Science

Carnegie Mellon University, Pittsburgh, USA

{vbalacha,apagnoni,jaylee,dheeraj,jgc,ytsvetko}@cs.cmu.edu

Motivation

1， Problem on Abs summarization model: layout bias； limited abstractiveness； lack of transparency

2， Augmenting standard encoder-decoder summarization models with rich structure-aware document representations based on **implicitly learned (latent) structures** and **externally-derived linguistic (explicit) structures**.

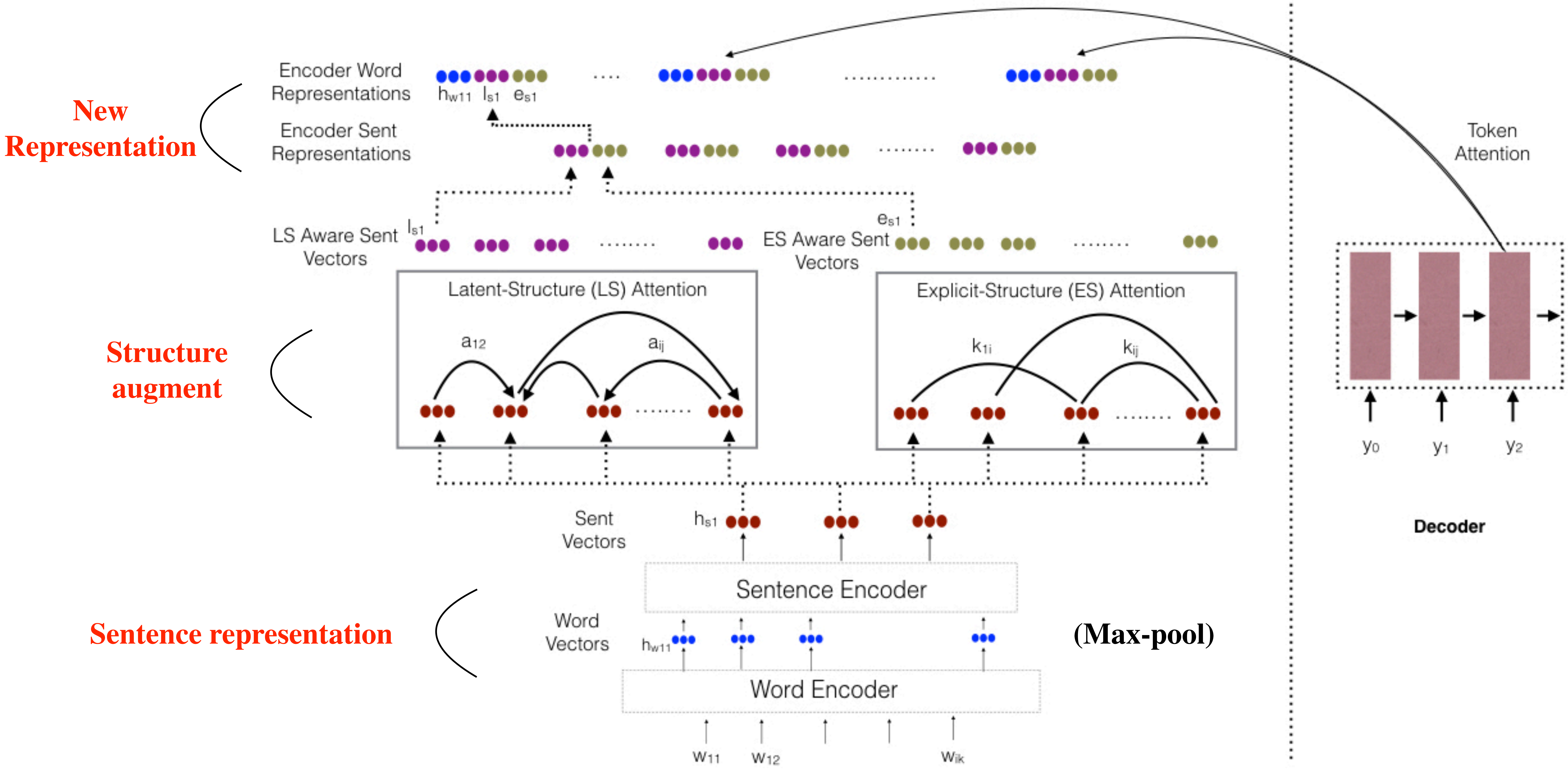
Solving

Solve problems:

Layout bias (with improving coverage of source document sentences);

Limited abstractiveness (by reducing bias of copying large sequences from the source and generate more novel n-grams);

Lack of transparency (to show examples of the learned interpretable sentence dependency structure)


New Representation
Structure augment
Sentence representation
Encoder Word Representations
Encoder Sent Representations
LS Aware Sent Vectors
ES Aware Sent Vectors
Latent-Structure (LS) Attention
Explicit-Structure (ES) Attention
Sent Vectors
Sentence Encoder
Word Vectors
Word Encoder
(Max-pool)
Token Attention
Decoder

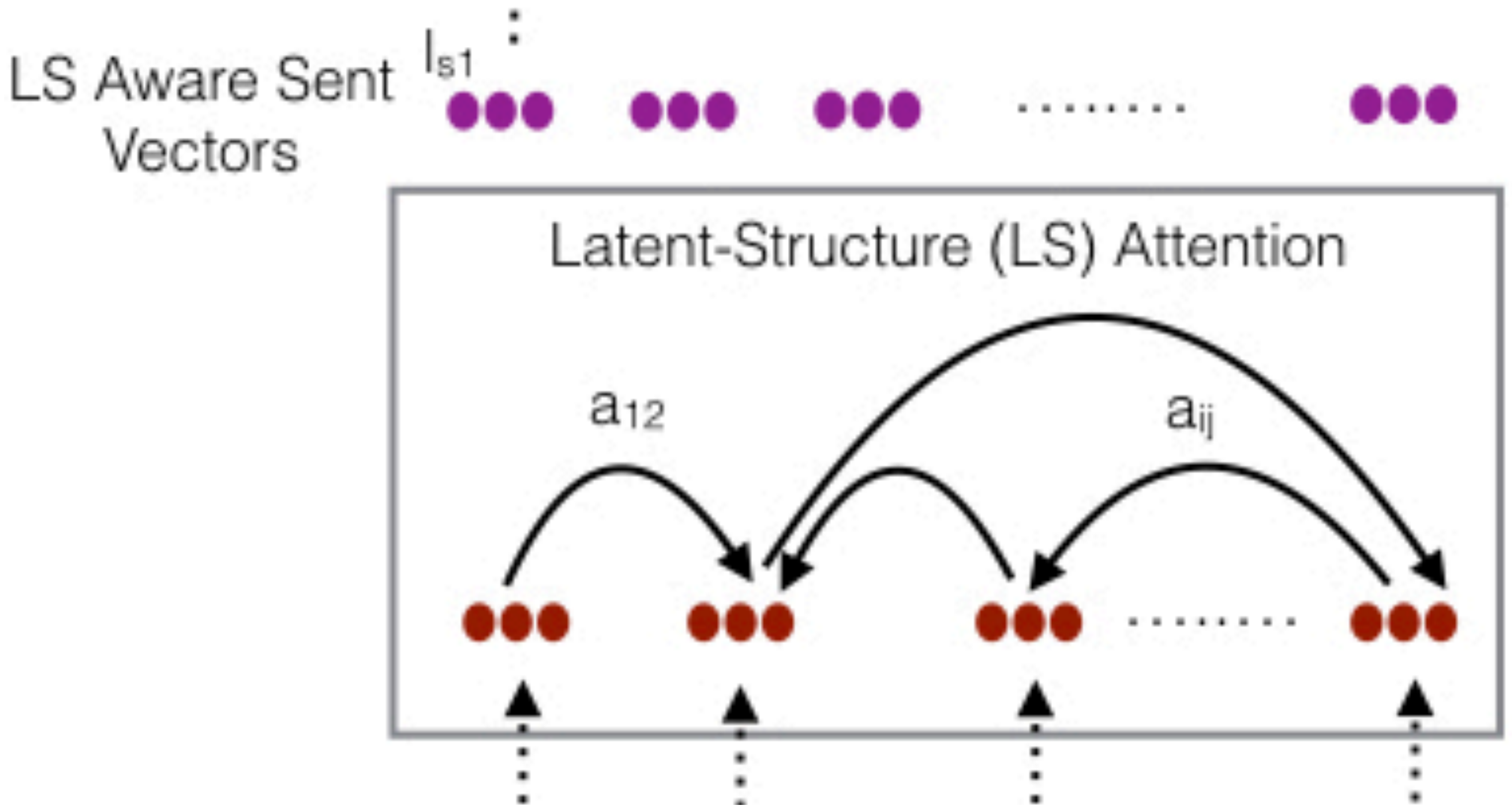
Latent Structure

non-projective dependency tree

pairwise attention： $a_{ij} = p(z_{ij} = 1)$

where z_{ij} is the latent variable representing

the edge from sentence i to sentence j.



use the **Kirchoff’s matrix tree theorem**（基尔霍夫） to compute

the marginal probability P of a dependency edge.

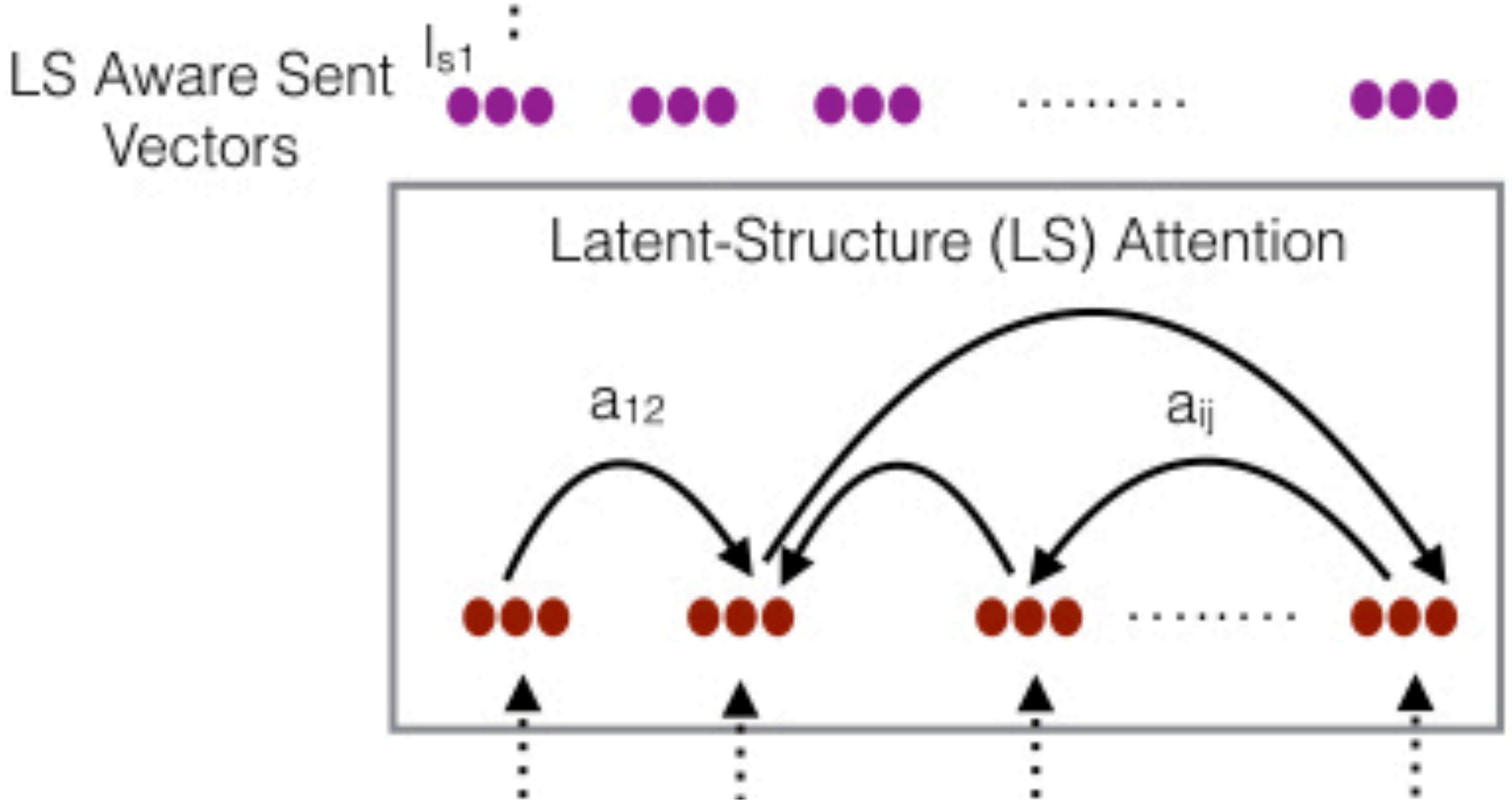
Latent Structure

Compute the sentence pairs score:

h_{si} = [g_{si} ; d_{si}]: representation of a sentence s_i into a semantic vector g_{si} and structure vector d_{si} .

$$f_{ij} = F_p(\mathbf{d}_{\boldsymbol{s}_i})^T W_a F_c(\mathbf{d}_{\boldsymbol{s}_j}) \text{ and } r_i = F_r(\mathbf{d}_{\boldsymbol{s}_i})$$

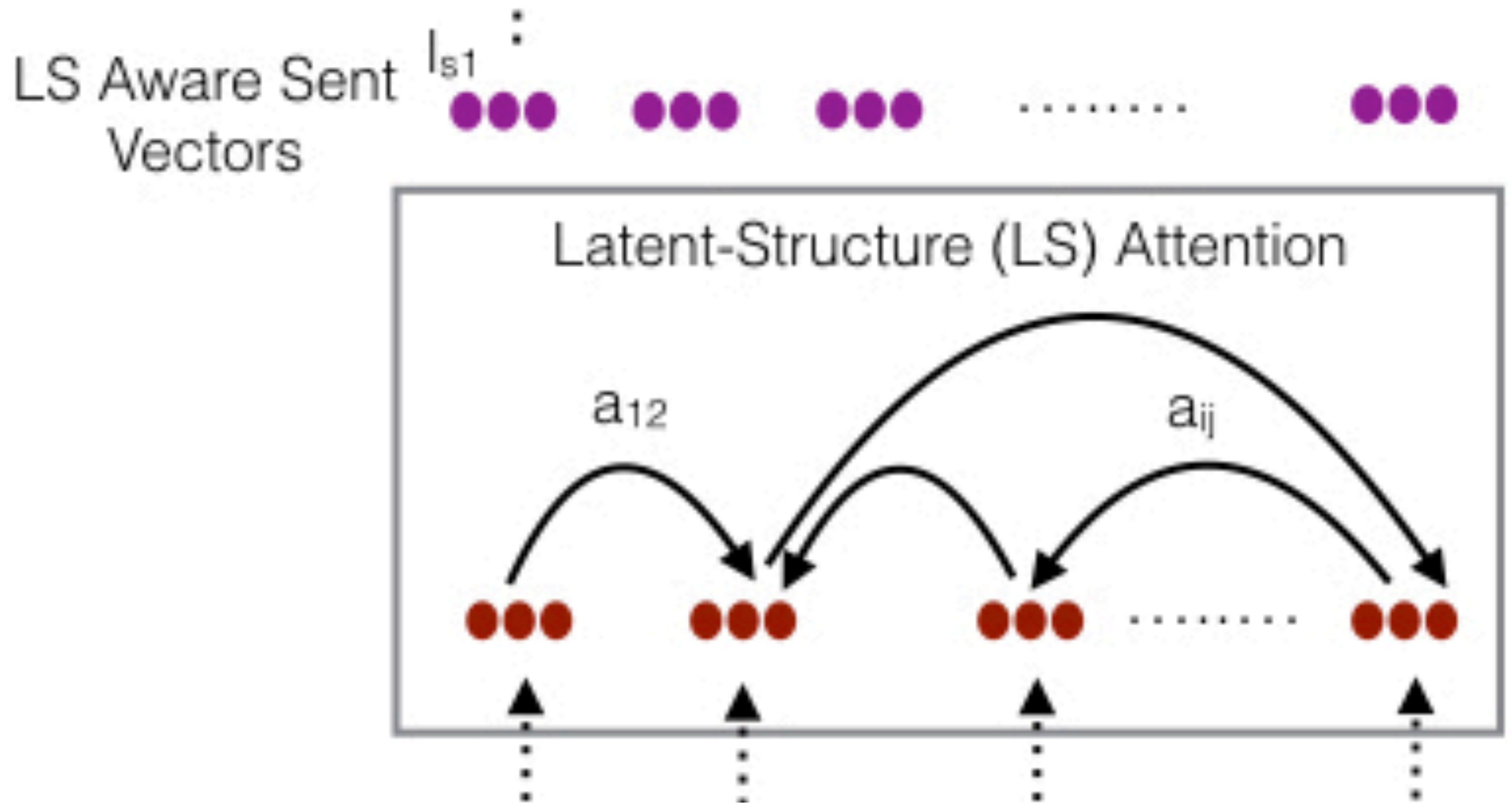
(r_i : when s_i is the root node)



Latent Structure

Compute the attended sentence representations:

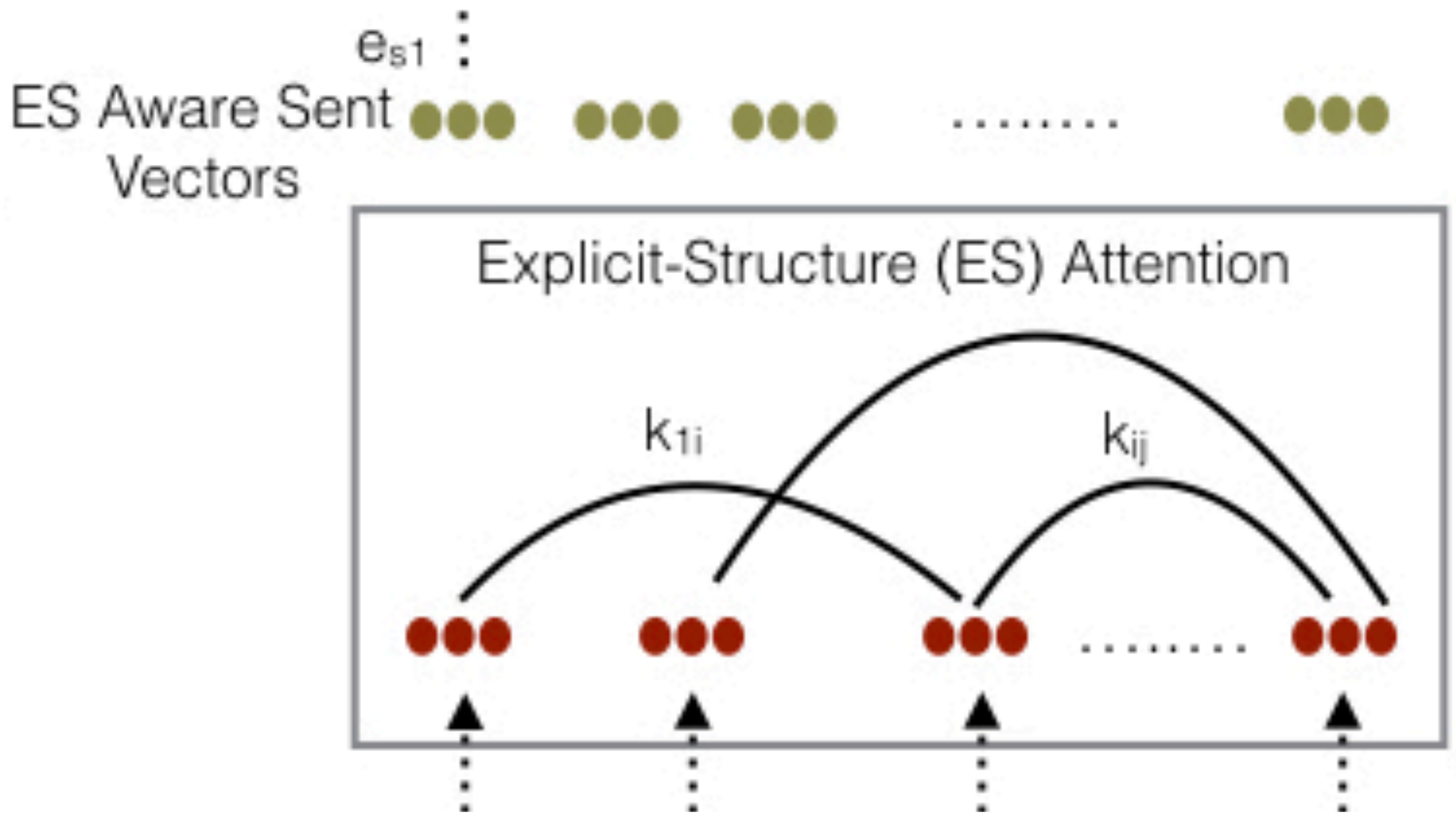
$$\mathbf{p}_{\boldsymbol{s}_i} = \sum_{j=1}^{n} a_{ji}\mathbf{g}_{\boldsymbol{s}_j} + a_i^r \mathbf{g}_{root}$$
$$\mathbf{c}_{\boldsymbol{s}_i} = \sum_{j=1}^{n} a_{ij}\mathbf{g}_{\boldsymbol{s}_i}$$
$$\mathbf{l}_{\boldsymbol{s}_i} = \tanh(W_r[\mathbf{g}_{\boldsymbol{s}_i}, \mathbf{p}_{\boldsymbol{s}_i}, \mathbf{c}_{\boldsymbol{s}_i}])$$



Explicit Structure

Coreference score:

$$k_{ij} = P(z_{ij} = 1)$$
$$= \frac{count(m_i \bigcap m_j) + \epsilon}{\sum_{v=1}^{n} count(m_i \bigcap m_v)}$$



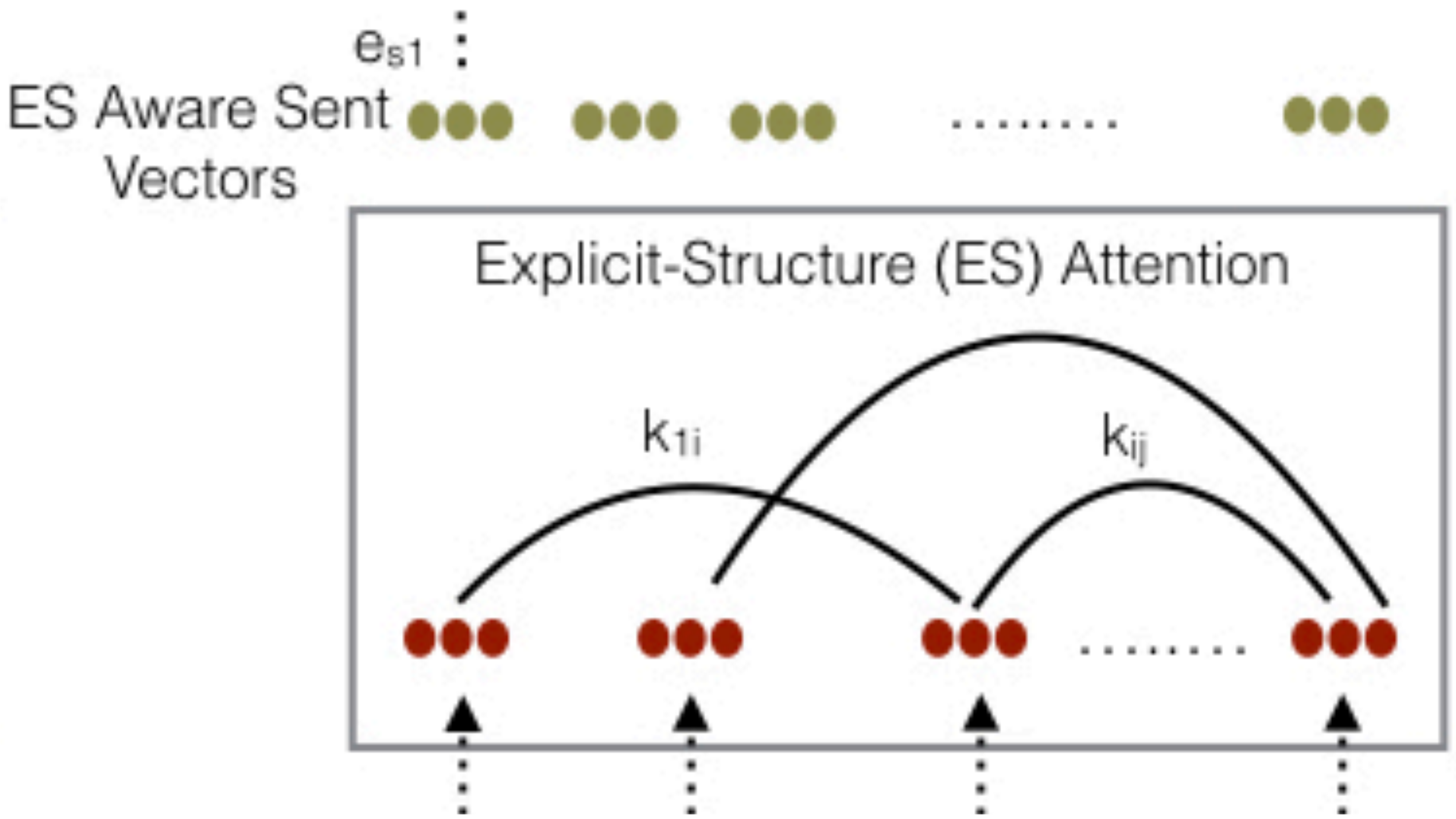
Explicit Structure

Learn an explicit structure-aware representation:

$$\mathbf{u}_{\boldsymbol{s}_i} = \tanh(F_u(\mathbf{h}_{\boldsymbol{s}_i}))$$

$$\mathbf{t}_{\boldsymbol{s}_i} = \sum_{j=1}^{p} k_{ij}\mathbf{u}_{\boldsymbol{s}_j}$$

$$\mathbf{e}_{\boldsymbol{s}_i} = \tanh(F_e(\mathbf{t}_{\boldsymbol{s}_i}))$$



Experiments

Evaluation on CNN/DM dataset:

Model	**ROUGE 1**	**ROUGE 2**	**ROUGE L**
Pointer-Generator (See et al., 2017)	36.44	15.66	33.42
Pointer-Generator + Coverage (See et al., 2017)	39.53	**17.28**	36.38
Graph Attention (Tan et al., 2017)	38.10	13.90	34.00
Pointer-Generator + DiffMask (Gehrmann et al., 2018)	38.45	16.88	35.81
Pointer-Generator (Re-Implementation)	35.55	15.29	32.05
Pointer-Generator + Coverage (Re-Implementation)	39.07	16.97	35.87
Latent-Structure (LS) Attention	39.52	16.94	36.71
Explicit-Structure (ES) Attention	**39.63**	16.98	36.72
LS + ES Attention	39.62	17.00	**36.95**

Experiments

Evaluation novel n-grams and coverage:

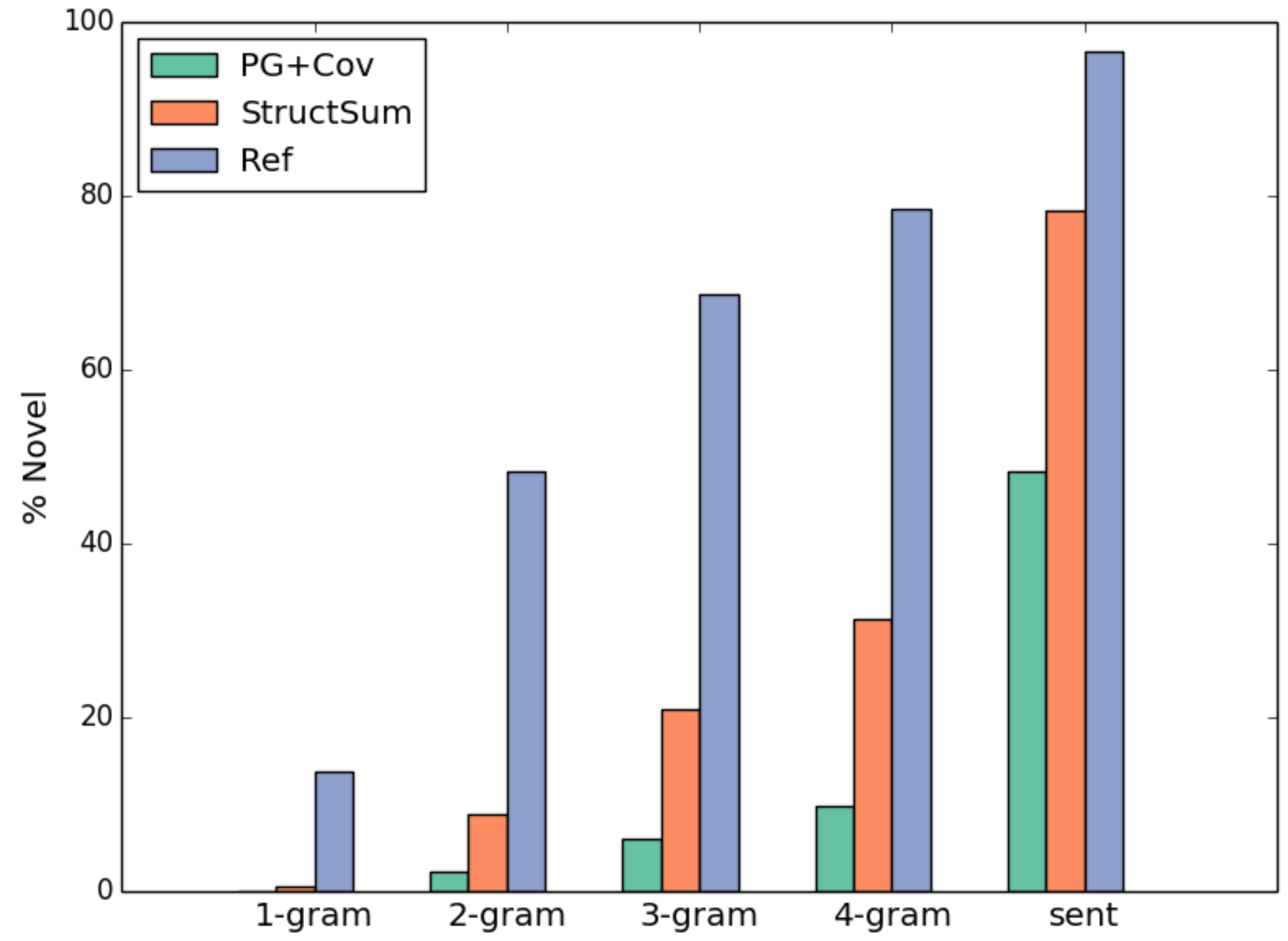


Figure 2: Comparison of % Novel n-grams between StructSum, Pointer-Generator+Coverage and the Reference. Here, "sent" indicates full novel sentences.

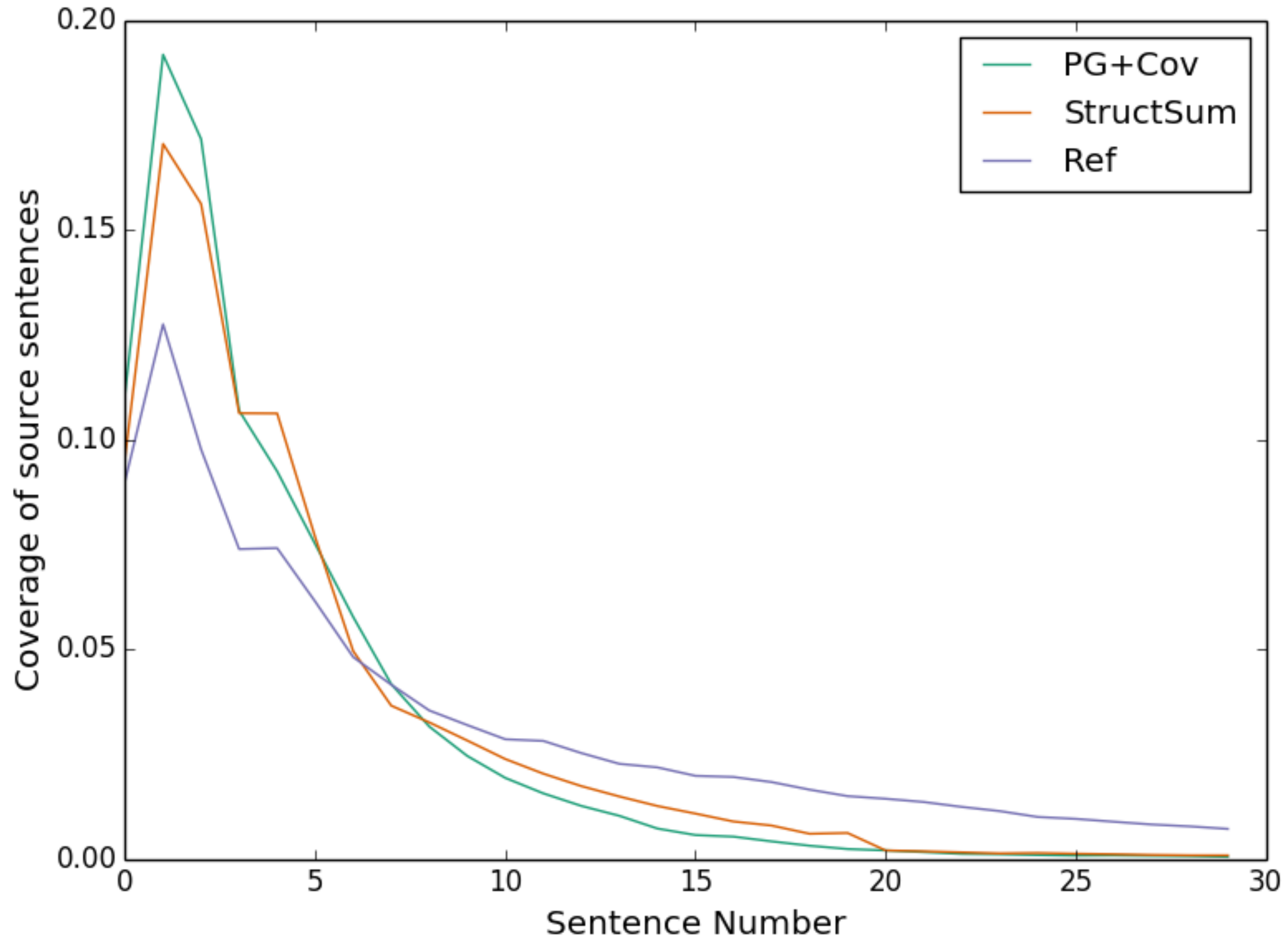


Figure 3: Coverage of source sentences in summary. Here the x-axis is the sentence position in the source article and y-axis shows the normalized count of sentences in that position copied to the summary.

Experiments

Evaluation the abstractiveness and coverage:

	Copy Len	Coverage
PG+Cov	16.61	12.1 %
StructSum	9.13	24.0 %
Reference	5.07	16.7 %

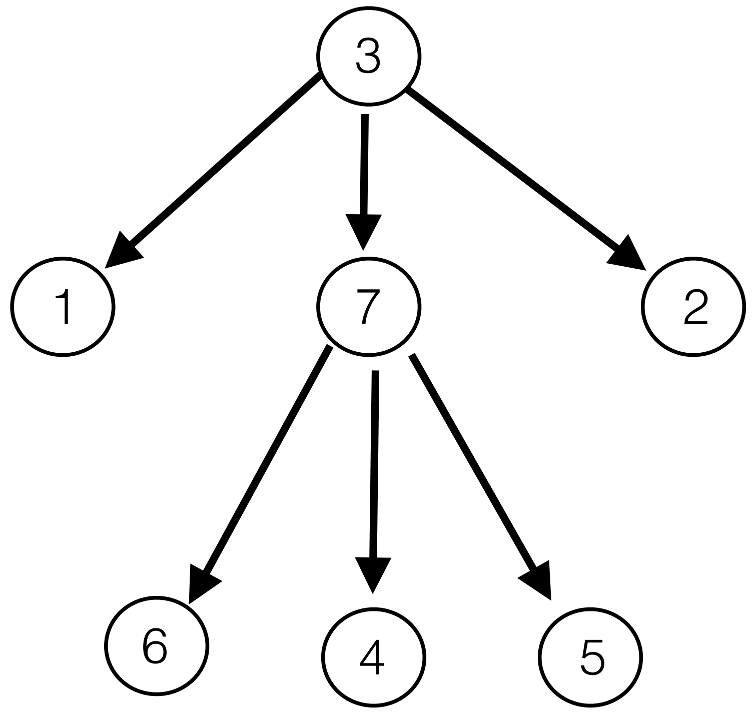
Table 2: Results of analysis of copying and coverage distribution over the source sentences on CNN/DM test set. Copy Len denotes the average length of copied sequences; Coverage – coverage of source sentences.

Depth	2	3	4	5+
StructSum	29.3%	53.7%	14.4%	2.6%

Table 4: Distribution of latent tree depth.

Document | **Latent Structures** | **Summaries**

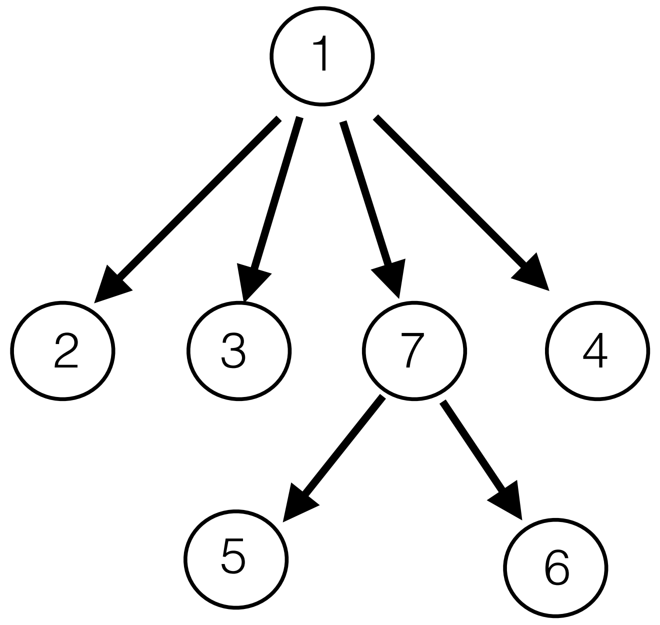
1. leicester city have rejected approaches for striker tom lawrence from an astonishing nine clubs .
2. the former manchester united forward has barely played for leicester since arriving from old trafford in the summer but manager nigel pearson wants to have all options available as he battles against the odds to keep leicester in the premier league .
3. lawrence , 21 , is poised to make his full international debut for wales in their european championship qualifier with israel on saturday but has only figured in four games for leicester this season and three as a substitute .
4. leicester city have rejected approaches for striker tom lawrence from an astonishing nine clubs .
5. championship promotion chasers bournemouth , ipswich and wolves have all asked about lawrence .
6. blackburn , charlton , leeds , bolton , rotherham and wigan have also made contact .
7. however , they are now looking at other options in a last-gasp bid to bolster their squad .



Reference:
bournemouth , ipswich , wolves , blackburn , charlton , leeds , bolton , rotherham and wigan have all asked about tom lawrence . the 21-year-old is poised to make his full international debut for wales . leicester manager nigel pearson wants to have options available .

StructSum:
leicester city have rejected approaches for tom lawrence . lawrence is poised to make his debut for wales in their european championship qualifier with israel on saturday . leicester city are looking at other options in last-gasp bid to bolster their squad . lawrence from old trafford has only figured in four games for leicester this season and three as a substitute . the former manchester united star has barely

1. andrew henderson celebrated landing the london broncos coaching job on a permanent basis as halifax were beaten 22-18 .
2. henderson was given the nod by the london hierarchy this week after a mixed spell in caretaker charge since the departure of joey grima .
3. his weakened side put on a fine show to crown his appointment , though , scoring four tries through daniel harrison , matt garside , iliess macani and brad dwyer , whose score was the winning one .
4. iliess macani , pictured last year , scored one of london broncos ' four tries in the 22-18 win over halifax .
5. james saltonstall , ben heaton and mitch cahalane scored for halifax .
6. henderson had spoken earlier in the week about how he felt broncos were moving in the right direction , and their narrow victory put some substance to his words .
7. the win was just their third in six in the kingstone press championship having been relegated from super league at the end of last season .



Reference:
andrew henderson won his first game as broncos full-time coach . daniel harrison , matt garside , iliess macani and brad dwyer all scored . james saltonstall , ben heaton and mitch cahalane scored for halifax .

StructSum:
andrew henderson celebrated landing the coaching job on a permanent basis .henderson was given the nod by london hierarchy this week after a mixed spell in the 22-18 win over halifax . his weakened side put on fine show to crown his appointment , though he felt broncos were moving in the right direction . the win was their third in six in the press championship having been relegated

Figure 4: Examples of induced structures and generated summaries.

Conclusion

1, We propose the framework StructSum for incorporating latent and explicit document structure in neural abstractive summarization

2, Our framework improves the abstractiveness and coverage of generated summaries, and helps mitigate layout biases associated with prior models

3, Improve the Interpretability