



NumHG: A Dataset for Number-Focused Headline Generation

Jian-Tao Huang,¹ Chung-Chi Chen,² Hen-Hsen Huang,³ Hsin-Hsi Chen¹

¹Department of Computer Science and Information Engineering, National Taiwan University, Taiwan

²AIST, Japan

³Institute of Information Science, Academia Sinica Taiwan



Numerical Reasoning + Headline Generation

News:

At least **30** gunmen burst into a drug rehabilitation center in a Mexican border state capital and opened fire, killing **19** men and wounding **four** people, police said. Gunmen also killed **16** people in another drug-plagued northern city. The killings in Chihuahua city and in Ciudad Madero marked one of the bloodiest weeks ever in Mexico and came just weeks after authorities discovered **55** bodies in an abandoned silver mine, presumably victims of the country's drug violence. More than **60** people have died in mass shootings at rehab clinics in a little less than **two** years. Police have said **two** of Mexico's **six** major drug cartels are exploiting the centers to recruit hit men and drug smugglers, ...

Headline (Question): Mexico Gunmen Kill ____

Answer: 35

Annotation: Add(19,16)

NumHG

Corpus	# Sents	# Words	# Nums
Dolphin18K (Huang et al., 2016)	2.6	30.6	4.4
AQUA-RAT (Ling et al., 2017)	2.2	32.5	4.2
Math23K (Wang et al., 2017)	1.6	28.0	3.1
MathQA (Amini et al., 2019)	2.0	37.9	4.5
SVAMP (Patel et al., 2021)	2.8	31.8	3.2
NumHG (Proposed)	9.4	191.8	13.7

Operator	Description	Ratio
Copy(v)	Copy v from the article	65.00%
Trans(e)	Covert e into a number	17.37%
Paraphrase(v_0, n)	Paraphrase the form of digits to other representations	8.27%
Round(v_0, c)	Hold c digits after the decimal point of v_0	3.10%
Subtract(v_0, v_1)	Subtract v_1 from v_0	2.15%
Add(v_0, v_1)	Add v_0 and v_1	1.73%
Span(s)	Select a span from the article	1.34%
Divide(v_0, v_1)	Divide v_0 by v_1	0.54%
Multiply(v_0, v_1)	Multiply v_0 and v_1	0.50%

Pretrained LMs

	Num Acc.			ROUGE			BERTScore			MoverScore
	Overall	Copy	Reasoning	1	2	L	P	R	F1	
BART	70.09	73.88	61.54	46.63	21.79	41.55	48.02	49.19	48.62	62.57
T5	67.84	71.42	59.74	47.82	23.10	42.89	50.23	49.64	49.94	62.98
Pegasus	66.45	70.25	57.86	48.08	23.40	43.25	50.97	49.99	50.49	63.11
Season	67.81	71.11	60.35	48.58	23.81	43.74	51.64	50.32	50.98	63.29
BRIO	66.56	70.43	60.07	48.93	24.09	44.12	52.17	50.84	51.43	63.50

	Num Acc.	Reasonableness	Readability
BART	59.2	43.9	53.7
T5	53.9	52.1	55.9
Pegasus	64.6	58.8	61.2
Season	62.7	63.6	60.7
BRIO	79.1	65.2	63.5

Seasos (Wang et al., 2022) adopts a dual approach, learning to predict the informativeness of each sentence and using this predicted information to guide abstractive summarization.

BRIO (Liu et al., 2022) employs pre-trained abstractive models to generate candidate summaries, assigning each a probability mass according to their quality and defining a contrastive loss across the candidates. By considering both token-level prediction accuracy and sequence-level coordination, BRIO combines cross-entropy loss and contrastive loss for abstractive summarization.

LLMs

Team	Method	Num Accuracy			ROUGE			BERTScore			MoverScore
		Overall	Copy	Reasoning	1	2	L	P	R	F1	
NP-Problem	Finetuned Mistral-7B	73.49	76.91	67.26	39.82	17.58	34.34	27.80	48.56	37.82	57.02
Challenges	BART + Contrastive Learning	72.96	82.17	56.18	31.22	12.24	26.86	19.53	47.56	33.13	55.36
YNU-HPCC	Flan-T5 + Instruction Tuning + Retrieved Similar Example	69.04	73.02	61.81	48.85	24.68	44.18	51.55	50.10	50.38	60.55
Infrd.ai	Llama 2-7B + RAG	65.84	68.35	61.26	46.79	22.36	42.10	51.01	47.26	49.13	59.73
hinoki	T5-Based Title Generator	62.35	66.28	55.18	43.07	19.72	39.00	47.22	43.44	45.34	58.71
NCL_NLP	Mistral-7B + CoT + Finetune	62.12	65.54	55.90	43.51	19.39	38.88	46.40	45.04	45.73	58.86
NoNameTeam	-	55.72	57.68	52.13	40.65	17.26	35.75	44.26	40.39	42.32	57.74
Noot Noot	GPT-3.5	38.39	57.48	3.63	31.47	11.14	27.28	25.39	43.98	34.54	55.56
ClusterCore	Few-Shot Llama	38.23	51.57	13.94	33.47	11.84	28.93	31.88	42.23	37.03	56.41

Team	Numerical Accuracy	Optimal
Infrd.ai	1.81	22
NCL_NLP	1.73	16
Challenges	1.70	10
YNU-HPCC	1.69	15
Noot Noot	1.68	11
hinoki	1.67	16
ClusterCore	1.60	31
NoNameTeam	1.59	12
NP_Problem	1.57	14
Ground Truth	-	28

SemEval-2024

SemEval-2024 Task 7: Numeral-Aware Language Understanding and Generation

Numerical Reasoning

Team	Method	Accuracy
CTYUN-AI	Qwen-72B-Chat + Task Classification + Data Augmentation	0.95
ZXQ	Finetuned GPT-3.5	0.94
YNU-HPCC	Flan-T5 + CoT + Calculator	0.94
NCL_NLP	Mistral-7B + CoT + Finetune	0.94
NumDecoders	Ensemble (Flan T5 + GPT-3.5)	0.91
Infrd.ai	Llama 2-7B + CoT	0.90
hc	Flan-T5-LaMini	0.88
NP-Problem	Finetuned Mistral-7B	0.86
AIRah	-	0.83
Noot Noot	GPT-3.5	0.77
Sina Alinejad	-	0.74
StFX-NLP	-	0.60

Statistics of the operators present in the error sets of the top four systems for numerical reasoning

Operator	Ratio
Copy	23.42%
Trans	9.91%
Paraphrase	11.71%
Round	21.62%
Subtract	7.21%
Add	11.71%
Span	4.50%
Divide	4.50%
Multiply	5.41%

Mother of ____ Gives Huge Gift to Dying Friend

When Beth Laitkep's breast cancer spread to her brain and spine, doctors realized she had limited time left. The concern arose about the future of her six children. 'If a miracle doesn't occur and I do not survive, could you take my children as your own?' she inquired of her friend Stephanie Culley, as recounted to People magazine. Culley agreed without hesitation. Consequently, Ace (aged 2), Lily (5), Dallas (10), Jaxson (11), Selena (14), and Will (15) moved in with Culley, her husband Donnie, and **their three children** following Laitkep's demise in May at 39. Fortuitously, Donnie, a construction worker, had constructed their home in Alton, Virginia, with ample bedrooms to accommodate everyone. 'She is exceedingly humble and refrains from seeking assistance,' a friend of Stephanie's informed WSET. 'She's an angel.' (This family adopts children who are facing terminal conditions.)

Three out of four models filled the blank with 6, while one model suggested 7.

Open Research Question

Evaluation: Which one is the best headline?

2. Dow Falls 64 Points, Comes Within Half a Point of 20K
2. Dow Stocks Soar but Fail to Reach 20,000 Mark

1. Dow Nears 20K, But Loses Momentum
1. Dow Comes Within Half a Point of 20K
1. Dow Closes Below 20K
1. Dow Falls Short of 20K

NBA Season Cancellations Likely to Extend Through November 28 Due to Salary
NBA Season in Jeopardy as Owners Push for 50-50 Revenue Split
NBA Season Could Be Canceled Through Nov. 28
NBA May Cancel 2 More Weeks of Season
NBA to Cancel 2 More Weeks of Season
NBA Canceling 2 More Weeks of Games? 102 More Games Gone
NBA Planned to Ax 102 More Games

Upcoming Events

Temporal Inference of Financial Arguments (FinArg-2)@NTCIR-2025
• Assessment of the Premise's Influence Period
• Detection of Argument Temporal References
• Assessment of the Claim's Validity Period
• <http://finarg.nlpfin.com/>

FinNLP Workshop@IJCAI-2024
• Shared Tasks: FinLLM & Comment Generation
• <http://finnlp.nlpfin.com/>

INLG 2024@Tokyo
• Submission Deadline: May 31
• <https://inlg2024.github.io/>

