

Improved Lexically Constrained Decoding for Translation and Monolingual Rewriting

J. Edward Hu, Huda Khayrallah, Ryan Culkin,
Patrick Xia, Tongfei Chen, Matt Post, Benjamin Van Durme



Automatically
paraphrase your data!



Rewriting lots of sentences quickly,
with constraints on the lexical items

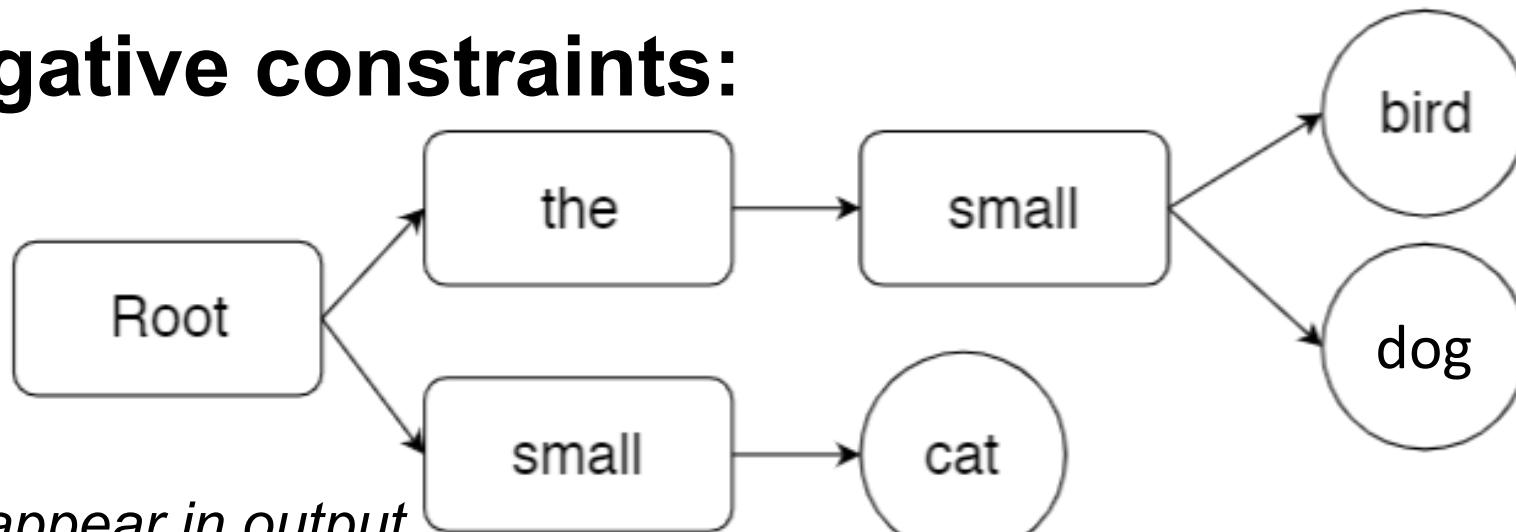
We propose a new algorithm and show positive results when used to augment NLP tasks.

Improved Constrained Decoding

Better algorithm:
Multi-state Trie (MST)

Representing 3 negative constraints:

- “the small dog”
- “the small bird”
- “small cat”



None of the phrases should appear in output.

Without tracking multiple states,
“Here comes the small cat.” becomes a possible output!

Better algorithm:
Vectorized Dynamic Beam Allocation (VDBA)^[1]

sentno	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
unmet	0	0	1	1	1	1	2	2	2	3	4	4	5	5	5
step	0	1	0	1	2	3	0	1	2	0	0	1	0	1	2
score	1.23	1.07	1.13	1.01	0.89	0.87	1.08	0.92	0.77	0.83	0.69	0.54	0.71	0.46	0.41
hypo #	4	2	1	0	6	3	2	0	1	5	1	0	1	0	0
vocab ID	4241	923	135	712	2122	14	805	4209	212	324	712	441	5243	853	5071

Sorted by num of unmet constraints

sentno	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
unmet	0	1	2	3	4	5	0	1	2	4	5	1	2	5	1
step	0	0	0	0	0	0	1	1	1	1	1	2	2	2	3
score	1.23	1.13	1.08	0.83	0.69	0.71	1.07	1.01	0.92	0.54	0.46	0.89	0.77	0.41	0.87
hypo #	4	1	2	5	1	1	2	0	0	0	0	6	1	0	3
vocab ID	4241	135	805	324	712	5243	923	712	4209	441	853	2122	212	5071	14

Sorted by the step row, creating a round-robin

Constraints	Speed (Sent/Sec)						BLEU					
	Batch Size 1		Batch Size 20		Batch Size 1		Batch Size 20		Batch Size 1		Batch Size 20	
	Baseline	+MST,VDBA	Baseline	+MST,VDBA	Baseline	+MST,VDBA	Baseline	+MST,VDBA	Baseline	+MST,VDBA	Baseline	+MST,VDBA
none	4.44	4.44	9.35	9.10	40.8	40.8						
rand3	0.51	1.29	0.50	4.13	43.7	43.8						
phr4	0.46	1.25	0.57	3.93	48.6	50.1						

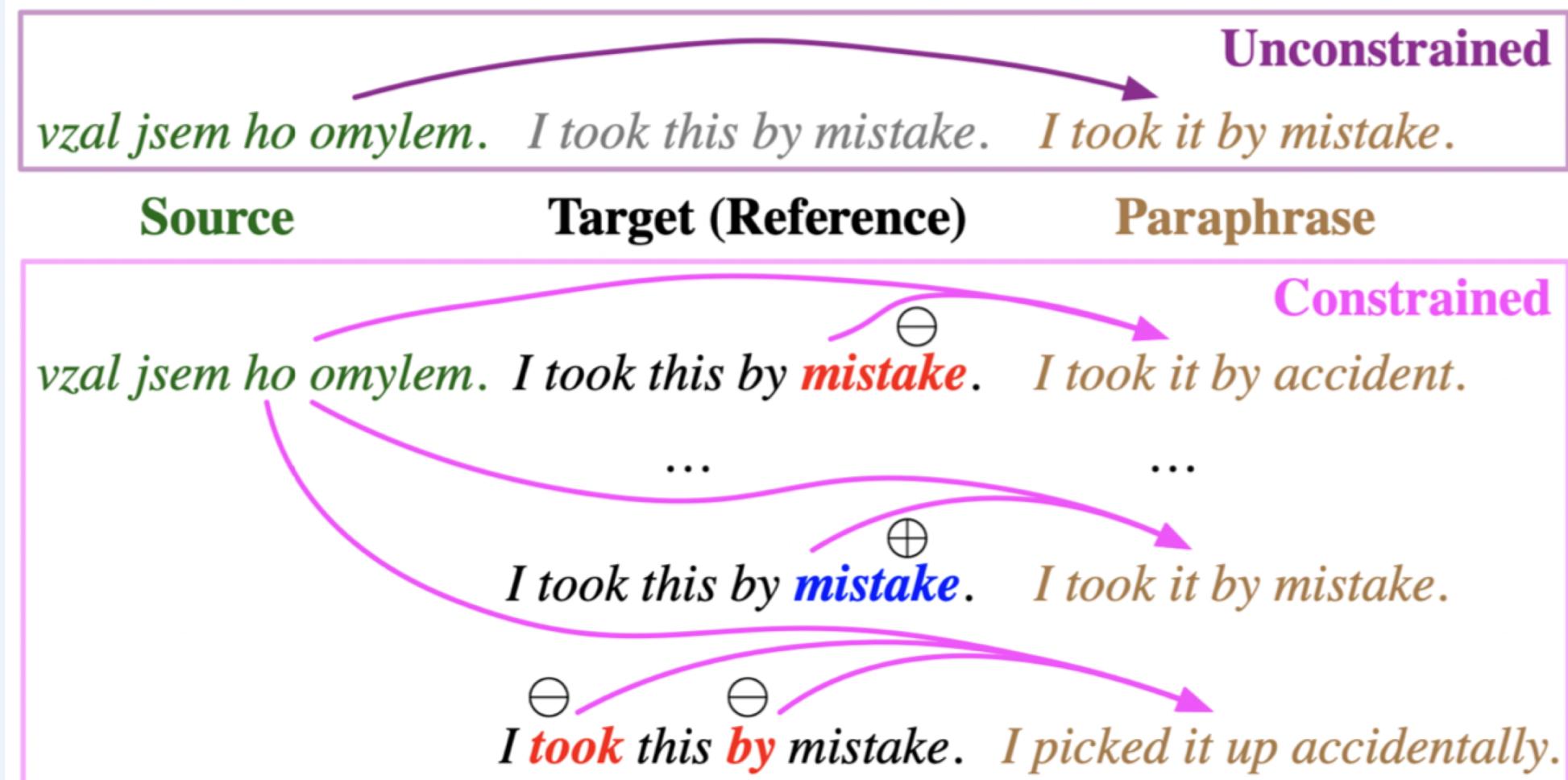
100% faster for positively constrained decoding;
600% faster for positively constrained decoding with batching
Better constraint placements (by BLEU)

	Dev.	Test. (m/mm)
Baseline	74.8	74.7 (74.8/74.6)
+Agg.	75.0	74.9 (74.9/74.8)
+Train	75.4	75.2 (75.0/75.3)
+Train+Agg.	75.6	75.4 (75.1/75.7)
+ELMo	75.8	75.0 (75.1/75.0)
+Agg.	75.9	75.2 (75.3/75.1)
+Train	76.4	75.6 (75.6/75.6)
+Train+Agg.	76.7	75.8 (75.9/75.7)

Table 3: F1 scores on MNLI. +Train denotes training on augmented data; +Agg. denotes using a weighted aggregation. Scores on the development set are a weighted average between the matched (m) and mismatched (mm) portions of the dataset, while the test set scores are additionally broken down into each category.

	MAP	MRR
Baseline	71.42	75.16
+Voting	72.94	77.17
+Train	71.57	74.63
+Train + Voting	73.96	80.77
+ELMo	77.49	81.86
+Voting	80.61	85.65
+Train	75.58	80.30
+Train + Voting	77.86	84.34

Table 5: Experimental results on QA selection.
TREC-QA^[4] Answer Sentence Selection task of Wang et al. EMNLP '07



Our prior work, ParaBank (AAAI'19)^[2]: sentential paraphrase generation via backtranslation with lexical constraints. Resulted in millions of Eng:Eng pairs, suitable for training rewriting systems

Data Augmentation via Rewriting

Given a rewriter and an NLP dataset you wish was larger than it was: perhaps generate various paraphrases of what you have?

P: I had rejected it as absurd , nevertheless it persisted
H: It persisted even after I rejected it as an absurdity
H': It went on even after I turned it down as an absurdity

Brittle NLI models can break under paraphrasing.
We release pMNLI, a paraphrastic expansion of MNLI^[3].

Gold: Entailed
P, H: Predict Entailed
P, H': Predict Contradiction

QA models benefit from more diverse training data.
We release pTREC-QA, expanded from TREC-QA^[4].

Q: What is the brightest star visible from Earth ?
A: Voyager will be headed toward Sirius , the brightest star in the heavens , after it leaves our solar system .
A': The Voyager will be heading for Sirius , the brightest star in the sky , after leaving our solar system .

Tr: En azından kağıt üzerinde, harika bir fikir gibi görünüyor.
En: On paper at least, it looks like a great idea.
En': It looks like a good idea on paper.

Low-resource MT benefits from augmentation for either side;
En-Tr (12.4 + 1.0 BLEU) Tr-En (15.6 + 0.5 BLEU).

Contributions

Better rewriting system (publicly available) for lexically-constrained decoding that is about 600% faster with batching;
Demonstrated improvements in QA, low resource MT, and in MNLI on top of ELMo

Code is based on extensions to AWS Sockeye, an open source, enterprise NMT toolkit. Trained models and data are available at:
<http://nlp.jhu.edu/parabank>

[1] Post and Vilar. 2018, Fast Lexically Constrained Decoding with Dynamic Beam Allocation for Neural Machine Translation

[2] Hu et al. 2019, ParaBank: Monolingual Bitext Generation and Sentential Paraphrasing via Lexically-constrained Neural Machine Translation

[3] Williams et al. 2018, A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference

[4] Wang et al. 2007, What is the Jeopardy model? A quasi-synchronous grammar for QA