

Data-to-Text Generation with Iterative Text Editing: Supplementary Material

A Hyperparameter Setup

Examples in the original datasets can have multiple reference lexicalizations. We introduce three strategies for dealing with this fact during the construction of the training dataset for the sentence fusion model:

- “*best*”: select the best lexicalizations for both the source and the target using LMSCORER
- “*best_tgt*”: select the best lexicalization for the target using LMSCORER and use all lexicalizations for the source
- “*all*”: use all lexicalizations for both the source and the target

Note that the training dataset is further filtered by the limited vocabulary of LASERTAGGER, which helps to filter out the outliers. We experiment with vocabulary sizes $V \in \{100, 500, 1000, 5000\}$. Table 4 shows the results on the development sets of both datasets. Based on these results, we select $V = 100$ and the strategy *all* for our final experiments.

WebNLG												
vocab. size	<i>best</i>				<i>best_tgt</i>				<i>all</i>			
	100	500	1000	5000	100	500	1000	5000	100	500	1000	5000
BLEU	0.373	0.370	0.370	0.335	0.382	0.382	0.375	0.342	0.397	0.389	0.391	0.370
NIST	7.610	7.478	7.411	6.713	7.673	7.596	7.470	6.831	7.912	7.676	7.679	7.307
METEOR	0.398	0.399	0.397	0.396	0.401	0.399	0.396	0.393	0.400	0.401	0.400	0.399
ROUGE_L	0.566	0.569	0.568	0.553	0.569	0.570	0.569	0.556	0.574	0.577	0.576	0.568
CIDER	2.586	2.573	2.466	2.023	2.594	2.525	2.466	2.133	2.639	2.570	2.557	2.385

E2E												
vocab. size	<i>best</i>				<i>best_tgt</i>				<i>all</i>			
	100	500	1000	5000	100	500	1000	5000	100	500	1000	5000
BLEU	0.252	0.254	0.249	0.255	0.269	0.258	0.260	0.256	0.293	0.277	0.273	0.268
NIST	4.168	4.180	4.049	4.077	4.435	4.167	4.154	4.097	4.762	4.461	4.357	4.238
METEOR	0.345	0.346	0.348	0.351	0.351	0.352	0.351	0.350	0.353	0.350	0.352	0.355
ROUGE_L	0.426	0.435	0.429	0.429	0.441	0.434	0.435	0.430	0.460	0.448	0.447	0.441
CIDEr	0.739	0.759	0.647	0.634	0.929	0.728	0.693	0.678	1.128	0.967	0.881	0.799

Table 4: Results of automatic metrics on the WebNLG and E2E development sets with different reference strategies and vocabulary sizes.

B Discourse Types

The list of different discourse types available in the DISCOFUSE dataset, with an indication whether they were selected for our zero-shot training, is shown in Table 5.

type	selected	type	selected
PAIR_ANAPHORA	yes	SINGLE_CONN_INNER_ANAPHORA	no
PAIR_CONN	no	SINGLE_CONN_START	no
PAIR_CONN_ANAPHORA	no	SINGLE_RELATIVE	yes
PAIR_NONE	yes	SINGLE_S_COORD	yes*
SINGLE_APPPOSITION	yes	SINGLE_S_COORD_ANAPHORA	yes*
SINGLE_CATAPHORA	no	SINGLE_VP_COORD	yes*
SINGLE_CONN_INNER	no		

Table 5: A list of available discourse types in the DISCOFUSE dataset. For our zero-shot experiments, we select a subset of DISCOFUSE, omitting the phenomena which mostly do not occur in our datasets. The asterisk (*) symbolizes that only the examples with the connectives “and” or “, and” were selected.

C Output Examples

Tables 7–8 show examples of outputs of our iterative sentence fusion method (with in-domain training) on both the E2E and WebNLG datasets. We show both instances that produce flawless output (Tables 6 and 7) and instances where our approach makes an error (Table 8 and 9). Table 10 then illustrates the behavior of the zero-shot approach (without in-domain training data).

Triples	(A Loyal Character Dancer, publisher, Soho Press); (Soho Press, country, United States); (United States, leaderName, Barack Obama)
Step #0	Soho Press is the publisher of A Loyal Character Dancer.
Step #1	Soho Press is the publisher of A Loyal Character Dancer which can be found in the <u>United States</u> .
Step #2	Soho Press is the publisher of A Loyal Character Dancer which can be found in the United States where <u>Barack Obama</u> is president.
Reference	A Loyal Character Dancer is published by Soho Press in the United States where Barack Obama is the president.

Table 6: An example of correct behavior of the algorithm on the WebNLG dataset (newly added entities are underlined).

Triples	(Giraffe, area, riverside); (Giraffe, eatType, pub); (Giraffe, familyFriendly, no); (Giraffe, food, Chinese); (Giraffe, near, Raja Indian Cuisine)
Step #0	Giraffe serves French food and is not family-friendly. ↳ A template for the pair of predicates "eatType" and "familyFriendly" is selected.
Step #1	Giraffe serves French food in the <u>riverside</u> area and is not family-friendly.
Step #2	Giraffe is a French <u>pub</u> in the riverside area that is not family-friendly.
Step #3	Giraffe is a French pub in riverside that is not family-friendly. It is located near <u>Raja Indian Cuisine</u> .
Reference	Giraffe is a not family-friendly French pub near Raja Indian Cuisine near the riverside.

Table 7: An example of correct behavior of the algorithm on the E2E dataset (newly added entities are underlined).

Triples	(Poland, language, Polish language); (Adam Koc, nationality, Poland); (Poland, ethnicGroup, Kashubians)
Step #0	Polish language is one of the languages that is spoken in Poland .
Step #1	Polish language is spoken in Poland, where Adam Koc <u>is spoken</u> . ↳ An incorrect expression is inserted.
Step #2	Polish language is spoken in Poland, where Adam Koc <u>is spoken</u> and Kashubians are an ethnic group.
Reference	The Polish language is used in Poland, where Adam koc was from. Poland has an ethnic group called Kashubians.

Table 8: An example of incorrect behavior of the algorithm on the WebNLG dataset (with the error underlined).

Triples	(The Phoenix, area, riverside); (The Phoenix, eatType, restaurant); (The Phoenix, familyFriendly, yes); (The Phoenix, near, Raja Indian Cuisine); (The Phoenix, priceRange, cheap)
Step #0	The Phoenix is a cheap place to eat. Yes it is family friendly. ↳ A template for the pair of predicates "price" and "familyFriendly" is selected.
Step #1	The Phoenix is a <u>cheap family friendly on the riverside</u> . ↳ A grammatical error is made.
Step #2	The Phoenix is a <u>cheap family friendly offering</u> restaurant in the riverside area. ↳ The grammar of the sentence is still not correct.
Step #3	The Phoenix is a cheap, family friendly restaurant in the riverside area, located near Raja Indian Cuisine. ↳ Grammatical errors are fixed in the last step of sentence fusion.
Reference	Cheap food and a family friendly atmosphere at The Phoenix restaurant. Situated riverside near the Raja Indian Cuisine.

Table 9: An example of behavior of the algorithm on the E2E dataset with several intermediate mistakes (underlined) and fixed output.

Triples	(Arrabbiata sauce, region, Rome); (Arrabbiata sauce, country, Italy); (Arrabbiata sauce, ingredient, olive oil)
Step #0	Arrabbiata sauce is a dish that comes from the Rome region. ↳ A template for the predicate "region" (suitable for food) is selected.
Step #1	Arrabbiata sauce is a dish that comes from the Rome region, <u>and it</u> is a dish that is popular in Italy. ↳ The sentences are correctly joined together.
Step #2	Arrabbiata sauce is a dish that comes from the Rome region, and it is a dish that is popular in Italy. Olive oil is one of the ingredients used to make Arrabbiata sauce. ↳ The text is left intact.
Reference	Arrabbiata sauce is a traditional dish from Rome, Italy. Olive oil is one of the ingredients in the sauce.

Table 10: An example of behavior of the zero-shot algorithm on the WebNLG dataset (with a single change made by the editing step underlined).