Negative Lexically Constrained Decoding for Paraphrase Generation

Tomoyuki Kajiwara (Osaka University) kajiwara@jnlp.org

The Problem of Conservative Rewriting in Paraphrase Generation Tasks

Paraphrase Generation = Monolingual Machine Translation ?
MT: all words appearing in an input sentence must be rewritten in the target language.
PG: does not require rewriting of all words.

Owing to the characteristics of the task where only a limited portion of an input sentence needs to be rewritten, previous methods based on MT often perform conservatively and fail to produce necessary rewrites.

Example of Formality Transfer (Informal → Formal). Bolded word is identified as source style.

Source	**mama** so ugly, she scares buzzards off of a meat wagon.
Reference	Your mother is so unattractive she scared buzzards off of a meat wagon.
Baseline	**mama** is so ugly, she scares buzzards off of a meat wagon.
Ours	The mother is so unattractive that she scares buzzards off of a meat wagon.

Identification of Words to be Paraphrased → Constrained Beam Search

1. We extract words w strongly related to the source style x included in the input sentence s_i as vocabulary V_i to be paraphrased.

$$V_i = \{w | w \in s_i \land \mathrm{PMI}(w, x) \geq \theta\}$$

2. We force to not include words contained in the vocabulary V_i in output sentence.
This is implemented by avoiding certain words in the beam search. [Post and Vilar, 2018]
= Negative Lexically Constrained Decoding

Matt Post and David Vilar (2018) Fast Lexically Constrained Decoding with Dynamic Beam Allocation for Neural Machine Translation. In Proc. of NAACL.

Performance on the Text Simplification and Formality Transfer Tasks

Our method consistently improves the quality of paraphrase generation across styles or domains.

	Newsela					GYAFC-E&M				GYAFC-F&R			
	Add	Keep	Del	BLEU	SARI	Add	Keep	Del	BLEU	Add	Keep	Del	BLEU
RNN-Base	1.8	60.8	22.3	24.1	17.4	31.9	90.0	57.5	71.2	32.9	90.5	61.1	74.7
RNN-PMI	**2.8**	**61.1**	**36.5**	**24.7**	**22.8**	**33.5**	**90.0**	**59.9**	**71.7**	**34.3**	**90.9**	**63.1**	**75.9**
RNN-Oracle	10.4	82.9	89.9	36.4	40.0	34.8	92.7	72.4	75.2	35.7	93.2	74.6	79.3
SAN-Base	1.8	60.9	23.8	24.0	17.8	34.4	90.0	59.9	71.8	34.5	91.1	63.2	76.7
SAN-PMI	**2.5**	**61.3**	**38.0**	**24.6**	**23.3**	**35.2**	**90.0**	**61.2**	**72.1**	**35.3**	**91.1**	**64.0**	**77.0**
SAN-Oracle	10.1	82.0	89.4	35.9	39.9	36.6	92.4	71.4	75.1	36.6	92.9	73.7	79.8

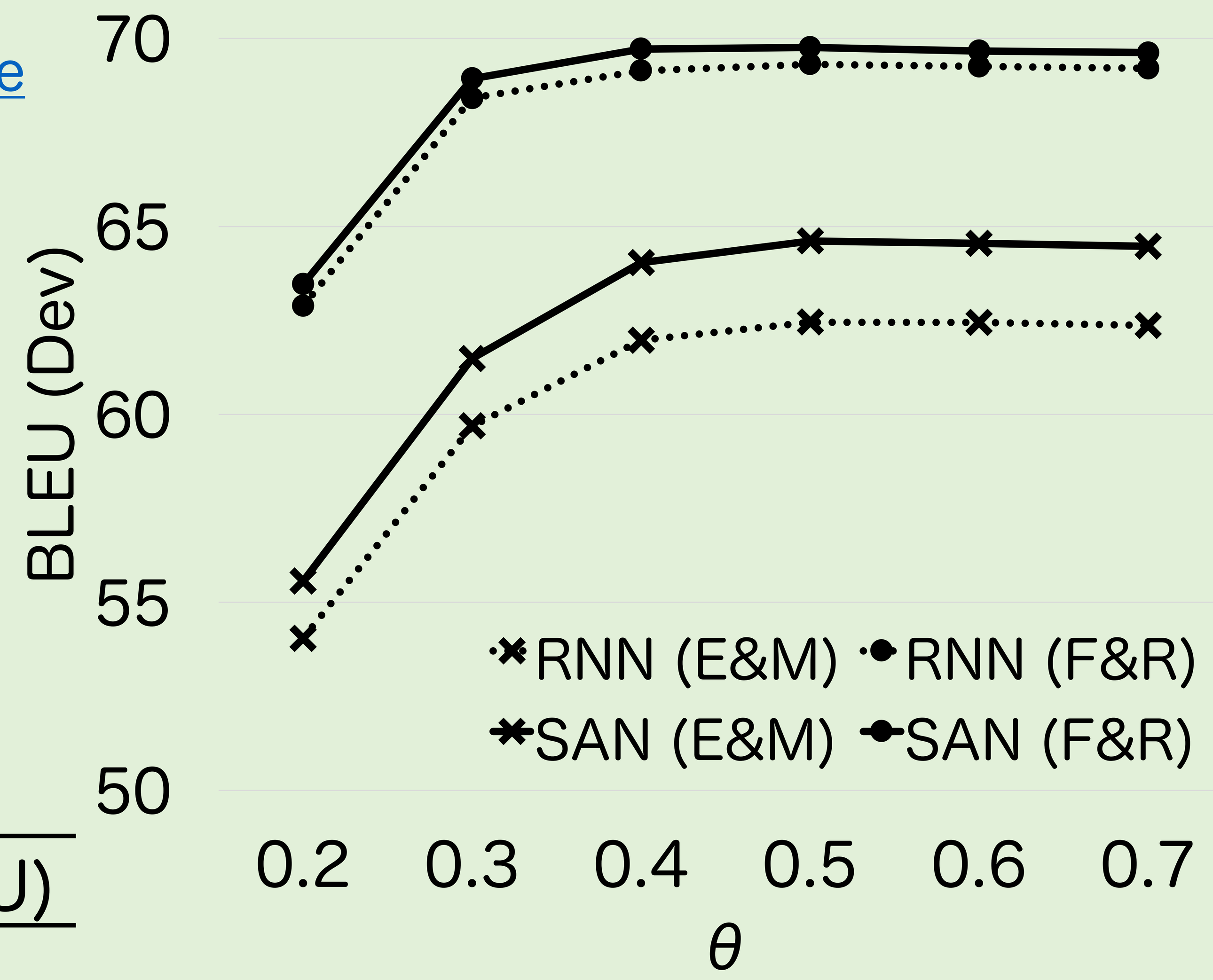
Models are based on Sockeye toolkit https://github.com/awslabs/sockeye
- RNN: a single Bi-LSTM with a layer size of 512.
- SAN: a six-layer Transformer with a model size of 512.

Oracle: Upper bound of identification. Identify all words that do not appear in the reference sentence.

In addition to BLEU, F1 scores of the word correctly added / kept / deleted from the input sentence are also evaluated.

sentence pairs for each dataset. GYAFC has multiple references.

	Train	Dev	Test	SOTA (BLEU)
Text Simplification: Newsela	94,208	1,129	1,077	24.3
Formality Transfer: GYAFC-E&M	52,595	2,877	1,416	71.4
Formality Transfer: GYAFC-F&R	51,967	2,788	1,332	74.5



A threshold that is not too low brings high performance.