Sentence Simplification with Deep Reinforcement Learning

Created	@Jun 08, 2019 8:06 PM
Tags	
Updated	@Jun 17, 2019 5:13 PM

[Sentence Simplification with Deep Reinforcement Learning](https://arxiv.org/abs/1703.10931)
https://arxiv.org/abs/1703.10931

Neural Encoder-Decoder Model

- encoder transforms the source sentence X into a sequence of hidden states with a LSTM
- decoder uses another LSTM to generate one word yt+1 at a time in the target Y.
- dynamic context vector ct is the weighted sum of the hidden states of the source sentence
- weights αti are determined by an attention mechanism

$$P(Y|X) = \prod_{t=1}^{|Y|} P(y_t|y_{1:t-1}, X) \qquad (1)$$

$$P(y_{t+1}|y_{1:t}, X) = \text{softmax}(g(\mathbf{h}_t^T, \mathbf{c}_t)) \qquad (2)$$

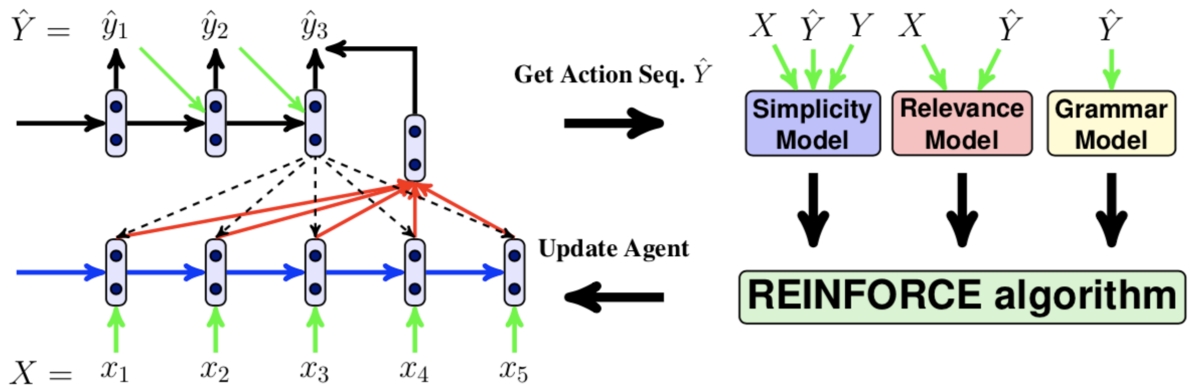
$$g(\mathbf{h}_t^T, \mathbf{c}_t) = \mathbf{W}_o \tanh(\mathbf{U}_h \mathbf{h}_t^T + \mathbf{W}_h \mathbf{c}_t) \qquad (3)$$

$$\mathbf{h}_t^T = \text{LSTM}(y_t, \mathbf{h}_{t-1}^T) \qquad (4)$$

$$\mathbf{c}_t = \sum_{i=1}^{|X|} \alpha_{ti} \mathbf{h}_i^S \tag{5}$$

$$\alpha_{ti} = \frac{\exp(\mathbf{h}_t^T \cdot \mathbf{h}_i^S)}{\sum_i \exp(\mathbf{h}_t^T \cdot \mathbf{h}_i^S)} \tag{6}$$

Reinforcement Learning for Sentence Simplification



Reward

The reward r(Yˆ) for system output Yˆ is the weighted sum of the three components

- Fluency(grammar) (is the output grammatical and well formed?)
- Adequacy(relevance) (to what extent is the meaning expressed in the original sentence preserved in the output?)
- Simplicity (is the output simpler than the original sentence?)
- X: the source
- Y: the reference (or target)
- Yˆ: the system output

$$r(\hat{Y}) = \lambda^S r^S + \lambda^R r^R + \lambda^F r^F \tag{7}$$

Simplicity (encourage the model to make changes)

SARI

- compares System output(Y^) Against References(Y) and against the Input sentence(X)
- arithmetic average of n-gram precision and recall of three rewrite operations: addition, copying, and deletion
- It rewards addition operations where system output was not in the input but occurred in the references.
- it rewards words retained/deleted in both the system output and the references.

To countenance the effect of noise SARI(X, Yˆ , Y)

- in the expected direction, with X as the source, Yˆ the system output, and Y the reference
- in the reverse direction with Y as the system output and Yˆ as the reference.
- reward is the weighted sum of SARI and reverse SARI:

$$r^S = \beta \, \textsc{Sari}(X, \hat{Y}, Y) + (1-\beta) \, \textsc{Sari}(X, Y, \hat{Y}) \quad (8)$$

Relevance (generated sentences preserve the meaning of the source)

The relevance reward is simply the cosine similarity between these two vectors

- qx = SAE(X)
- qy = SAE(Y)

$$r^R = \cos(\mathbf{q}_X, \mathbf{q}_{\hat{Y}}) = \frac{\mathbf{q}_X \cdot \mathbf{q}_{\hat{Y}}}{||\mathbf{q}_X|| \, ||\mathbf{q}_{\hat{Y}}||} \qquad (9)$$

- SAE(sequence auto-encoder):
  - encoder LSTM converts X into a sequence of hidden states (h1, . . . , h|X|)

- use h|X| to initialize the hidden state of the decoder LSTM
- recover/generate X one word

Fluency (well-formedness of the generated sentences)

$$r^F = \exp\left(\frac{1}{|\hat{Y}|}\sum_{i=1}^{|\hat{Y}|}\log P_{LM}(\hat{y}_i|\hat{y}_{0:i-1})\right) \quad (10)$$

The REINFORCE Algorithm (DRESS)

The training loss for one sequence is its negative expected reward:

- Prl (Policy) Equation(2): the distribution produced by the encoder-decoder model
- r(.): the reward function

$$\mathcal{L}(\theta) = -\mathbb{E}_{(\hat{y}_1,\ldots,\hat{y}_{|\hat{Y}|})\sim P_{RL}(\cdot|X)}[r(\hat{y}_1,\ldots,\hat{y}_{|\hat{Y}|})]$$

The gradient of L(θ) is:

- approximate this expectation with samples
- bt: reduce the variance of gradients use baseline linear regression model
  - concatenation of hTt and ct
  - the parameters of the regressor are trained by minimizing mean squared error
  - do not back-propagate this error to hTt or ct during training

$$\nabla\mathcal{L}(\theta) \approx \sum_{t=1}^{|\hat{Y}|}\nabla\log P_{RL}(\hat{y}_t|\hat{y}_{1:t-1},X)[r(\hat{y}_{1:|\hat{Y}|})-b_t]$$

Learning (curriculum learning strategy)

- model training challenging for generation tasks like ours with large vocabularies
- In the beginning of training, we give little freedom to our agent allowing it to predict the last few words for each target sentence.
- For every target sequence, we use negative log-likelihood to train the first L (initially, L = 24) tokens and apply the reinforcement learning algorithm to the (L + 1)th tokens onwards.
- Every two epochs, we set L = L − 3 and the training terminates when L is 0.

Lexical Simplification (DRESS-LS)

Lexical substitution, the replacement of complex words with simpler alternatives, is an integral part of sentence simplification.

We use an pre-trained encoder-decoder model (which is trained on a parallel corpus of complex and simple sentences) to obtain probabilistic word alignments, aka attention scores.

- source sentence X=(x1,x2,...,x|X|)
- target sentence Y = (y1,y2,...,y|Y |)
- hidden states with LSTM V = (v1, v2, . . . , v|X|)
- alignment(attention) scores αt1 , αt2 , . . . , αt|X |

$$P_{LS}(y_t|X, \alpha_t) = \text{softmax}(\mathbf{W}_l\, \mathbf{s}_t) \qquad (11)$$

$$\mathbf{s}_t = \sum_{i=1}^{|X|} \alpha_{ti} \mathbf{v}_i \qquad (12)$$

$$\begin{aligned} P(y_t|y_{1:t-1}, X) = (1 - \eta)\, P_{RL}(y_t|y_{1:t-1}, X) \\ + \eta\, P_{LS}(y_t|X, \alpha_t) \end{aligned} \qquad (13)$$

Experimental Setup

Datasets

We conducted experiments on three simplification datasets.

- WikiSmall
- WikiLarge
- Newsela

Training Details

1. trained an encoder-decoder model
    - uniformly initialized to $[-0.1, 0.1]$
    - We used Adam to optimize the model with learning rate 0.001
        - the first momentum coefficient was set to 0.9
        - the second momentum coefficient to 0.999.
    - The gradient was rescaled when the norm exceeded 5.
    - Both encoder and decoder LSTMs have two layers with 256 hidden neurons in each layer.
        - regularized all LSTMs with a dropout rate of 0.2
        - We initialized the encoder and decoder word embedding matrices with 300 dimensional Glove vectors.
2. second performed reinforcement learning training
    - learning rate 0.01
    - $\beta = 0.1$
    - $\lambda S = 1$
    - $\lambda R = 0.25$
    - $\lambda F = 0.5$.
3. trained the lexical simplification model
    - $\eta = 0.1$.

- To reduce vocabulary size
    - named entities were tagged with the Stanford CoreNLP and anonymized with a NE@N token

- where NE ∈ {PER, LOC, ORG, MISC} and N indicates NE@N is the Nth distinct NE typed entity.
- For example, "John and Bob are . . . " becomes "PER@1 and PER@2 are . . . ".

Evaluation

- BLEU: to assess the degree to which generated simplifications differed from gold standard references
- FKGL: to measure the readability of the output.
- SARI: evaluates the quality of the output by comparing it against the source and reference simplifications.

Comparison Systems

- PBMT-R: a mono-lingual phrase-based machine translation system with a reranking post-processing step
- Hybrid: a model which first performs sentence splitting and deletion operations over discourse representation structures and then further simplifies sentences with PBMT-R.
- SBMT-SARI: a syntax-based translation model trained on PPDB and tuned with SARI
- EncDecA: basic attention-based encoder-decoder model

Results

- indicates that the model has indeed learned to optimize the reward function

Newsela	BLEU	FKGL	SARI
PBMT-R	18.19	7.59	15.77
Hybrid	14.46	**4.01**	**30.00**
EncDecA	21.70	5.11	24.12
DRESS	23.21	4.13	27.37
DRESS-LS	**24.30**	4.21	26.63

WikiSmall	BLEU	FKGL	SARI
PBMT-R	46.31	11.42	15.97
Hybrid	**53.94**	9.20	**30.46**
EncDecA	47.93	11.35	13.61
DRESS	34.53	**7.48**	27.48
DRESS-LS	36.32	7.55	27.24

WikiLarge	BLEU	FKGL	SARI
PBMT-R	81.11	8.33	38.56
Hybrid	48.97	**4.56**	31.40
SBMT-SARI	73.08	7.29	**39.96**
EncDecA	**88.85**	8.41	35.66
DRESS	77.18	6.58	37.08
DRESS-LS	80.12	6.62	37.27

- We report results for Fluency, Adequacy, and Simplicity individually and in combination (All is the average rating of the three dimensions).

Newsela	Fluency	Adequacy	Simplicity	All
PBMT-R	3.56	**3.58****	2.09**	3.08**
Hybrid	2.70**	2.51**	2.99	2.73**
EncDecA	3.63	2.99	2.56**	3.06**
DRESS	3.65	2.94	**3.10**	3.23
DRESS-LS	**3.71**	3.07	3.04	**3.28**
Reference	3.90	2.81**	3.42**	3.38

WikiSmall	Fluency	Adequacy	Simplicity	All
PBMT-R	3.91	**3.74****	2.80**	3.48*
Hybrid	3.26**	3.42	2.82**	3.17**
DRESS-LS	**3.92**	3.36	**3.55**	**3.61**
Reference	3.74*	3.34	3.13**	3.41**

WikiLarge	Fluency	Adequacy	Simplicity	All
PBMT-R	3.68	**3.63***	2.70**	3.34*
Hybrid	2.60**	2.42**	**3.52**	2.85**
SBMT-SARI	3.34**	3.51*	2.77**	3.21**
DRESS-LS	**3.70**	3.28	3.42	**3.46**
Reference	3.79	3.72**	2.86**	3.46

Complex	There's just one major hitch: the primary purpose of education is to develop citizens with a wide variety of skills.
Reference	The purpose of education is to develop a wide range of skills.
PBMT-R	It's just one major hitch: the purpose of education is to **make people** with a wide variety of skills.
Hybrid	one hitch the purpose is to develop citizens.
EncDecA	The **key** of education is to develop **people** with a wide variety of skills.
DRESS	There's just one major hitch: the **main goal** of education is to develop **people** with **lots of** skills.
DRESS-LS	There's just one major hitch: the **main goal** of education is to develop citizens with **lots of** skills.

Complex	"They were so burdened by the past they couldn't think about the future," said Barnet, 62, who was president of Columbia Records, the No.1 record label in the United States, before joining Capitol.
Reference	Capitol was stuck in the past. It could not think about the future, Barnett said.
PBMT-R	"They were so **affected** by the past they couldn't think about the future," said Barnett, 62, was president of Columbia Records, before joining Capitol **building**.
Hybrid	'They were so burdened by the past they couldn't think about the future," said Barnett, 62, who was Columbia Records, president of the No.1 record label in the united states, before joining Capitol.
EncDecA	"They were so burdened by the past they couldn't think about the future," said Barnett, who was president of Columbia Records, the No.1 record labels in the United States.
DRESS	"They were so **sicker** by the past they couldn't think about the future," said Barnett, who was president of Columbia Records.
DRESS-LS	"They were so burdened by the past they couldn't think about the future," said Barnett, who was president of Columbia Records.