Simplifying Sentences and Documents

Claire Gardent

Joint work with Liam Cripwell

CNRS / LORIA

Outline

Controllable Sentence Simplification

- Classifying sentences with simplification operations (copy, split, rephrase ..)

  Complex Sentence $\rightarrow$ *Simp_OP*

- Simplifying a sentence using a specific simplification operation

  Simp_OP, Complex Sentence $\rightarrow$ *Simple Sentence*

Document Simplification

- Predicting a sequence of simplification operations for the input document

$$c1, \ldots, c_n \rightarrow \hat{o}1, \ldots, \hat{o}_n$$

- Simplifying a document based on this sequence of simplification operations

$$c1, \ldots, c_n + \hat{o}1, \ldots, \hat{o}_n \rightarrow s1, \ldots, s_k$$

Simplification

Example

C: Historical research indicates that the Zibelemarit originated in 1850 with marmettes, farmer's wives from around Murten, coming to Bern at around St. Martin's Day to sell their produce.

⇓

S: The Zibelemarit started around 150 years ago with marmettes, farmer's wives. They came to Bern at around St. Martin's Day to sell their produce.

Simplification Operations

Deleting

C: ~~Historical research indicates~~ that the Zibelemarit originated in 1850 with marmettes, farmer's wives ~~from around Murten~~, coming to Bern at around St. Martin's Day to sell their produce.

S:The Zibelemarit started around 150 years ago with marmettes, farmer's wives. They came to Bern at around St. Martin's Day to sell their produce.

Simplification Operations

Rephrasing

C: ~~Historical research indicates~~ that the Zibelemarit originated in 1850 with marmettes, farmer's wives ~~from around Murten~~, coming to Bern at around St. Martin's Day to sell their produce.

S:The Zibelemarit started around 150 years ago with marmettes, farmer's wives. They came to Bern at around St. Martin's Day to sell their produce.

Simplification Operations

Splitting

C: ~~Historical research indicates~~ that the Zibelemarit originated in 1850 with marmettes, farmer's wives ~~from around Murten~~, coming to Bern at around St. Martin's Day to sell their produce.

S:The Zibelemarit started around 150 years ago with marmettes, farmer's wives. They came to Bern at around St. Martin's Day to sell their produce.

Why Simplify ?

- To aid reader comprehension (Mason, 1978; Williams et al., 2003; Kajiwara et al., 2013)
    - Adult vs children
    - Native vs non Native
    - Reading disability
    - Expert vs non-Expert
- Useful preprocessing step for downstream NLP tasks such as
    - relation extraction (Miwa et al., 2010; Niklaus et al., 2016)
    - machine translation (Chandrasekar et al., 1996; Mishra et al., 2014; Li and Nenkova, 2015; Mishra et al., 2014; Štajner and Popovic, 2016).

Controllable Sentence Simplification

Previous Neural Approaches to Sentence Simplification

Two main types of approaches

End-to-end

> *Complex Sentence* $\rightarrow$ *Simple Sentence*

Controllable

> *CONTROL, Complex Sentence* $\rightarrow$ *Simple Sentence*

End-to-end

- Encoder-Decoder Models
- Trained on parallel Corpora of (C,S) pairs
  - Wiki: English Wikipedia / Simple English Wikipedia
  - Newsela
- Implicitly learn simplification operations from the training data

End-to-end

- Encoder-Decoder Models
- Trained on parallel Corpora of (C,S) pairs
- Implicitly learn simplification operations from the training data

Shortcomings

- Noisy training data
- Some simplification operations are rare
  (Jiang et al., 2020)
- Overly conservative models
  (Alva-Manchego et al., 2017; Maddela et al., 2021)
- Limited capacity for controllability
- Unable to generate alternative variants of the simplified text
  (Alva-Manchego et al., 2017; Cripwell et al., 2021).

Controllable Simplification

CONTROL, Complex Sentence $\rightarrow$ *Simple Sentence*

Constrains

- the *shape* (length, amount of paraphrasing, lexical and syntactic complexity) of the output (Martin et al., 2020)
- or the *type of simplification operation* to be applied (e.g., copy, split, merge, rewrite, etc.) (Scarton and Specia, 2018; Dong et al., 2019; Scarton et al., 2020; Garbacea et al., 2021; Maddela et al., 2021).

Proposal

A simplification model controlled for four operations

- copy (no simplification needed)
- rephrase
- split based on syntax
- split based on discourse structure

Contributions

- Novel simplification dataset labeled with these four operations
- Classification Model to predict these operations
- Controllable simplification based on these four operations

Cripwell et al. NAACL Findings 2022

Two types of Sentence Splitting

Discourse-Based Splits

The Masovians were caught by surprise, since virtually without any defense the capital, Płock, fell and after this Mindaugas crossed the Vistula river and captured the fortress of Jazdów.

The Masovians were caught by surprise, since virtually without any defense the capital, Płock, fell. Afterwards, Mindaugas crossed the Vistula river and captured the fortress of Jazdów.

Mindaugas crossed the Vistula river and captured the fortress of Jazdów. Before this, the Masovians were caught by surprise, since virtually without any defense the capital, Płock, fell.

Syntax-Based Splits

He settled in London, devoting himself chiefly to practical teaching.

He settled in London. He devoted himself chiefly to practical teaching.

Predicting Simplification Operations

Training Data

Wiki-Auto, Newsela-Auto
Automatically aligned (C, S) pairs extracted from Wikipedia and Newsela.

MUSS
2.7M (C, S) pairs mined from Common Crawl web data which are estimated paraphrases based on embedding distance.

WikiSplit
1M split (C, S) pairs mined from Wikipedia edit history,

D-CC-News
Discourse-split pairs mined from Common Crawl web data. Two subsets:

- D-CCNews-C which contains single Cs
- D-CCNews-S which contains pairs of organic Ss and synthetic Cs

Training Data

Class	Source						Total
	WikiSplit	MUSS	Wiki-auto	Newsela-auto	D-CCNews-C	D-CCNews-S	
Identity (0)	-	-	513,436	338,798	-	-	**852,234**
Rephrase (1)	-	461,702	366,382	171,508	-	-	**999,592**
Syntax-Split (2)	633,900	53,008	68,357	88,669	-	-	**843,934**
Discourse-Split (3)	269,666	1,002	5,277	2,060	250,062	249,958	**778,025**
Total	**903,566**	**515,712**	**953,452**	**601,035**	**250,062**	**249,958**	**3,473,785**

The data is balanced for the four simplification operations

IRSD$_4^C$ -- this dataset

IRSD$_3^C$ -- A 3-class subset which excludes the identity class to explore how results change when models are trained to always simplify

Silver Labels

Copy
Simplified sentences (Ss) from Wiki/Newsela-auto rephrase and syntax-split sets.

Rephrase
Complex sentences (Cs) from MUSS, Wiki-auto and Newsela-auto such that

- there is no split in the output S and
- sim(C,S) $\leq$ 1 standard deviation above the mean
  (Levenshtein similarity).

Syntax-split
Complex sentences (Cs) from WikiSplit, MUSS, Wikiauto and Newsela-auto whose S exhibits a split and does not contain an identifiable discourse marker.

Discourse-split
Complex sentences (Cs) from all datasets whose S contains a split and a discourse adverbial

Test Data

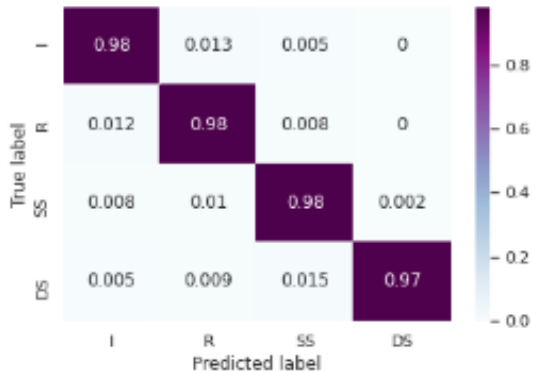
Silver Data
A random sample of 1% of the training data (34K examples)

Gold Data
A random sample of 100 items from each of the 4 classes in the silver test set and manually classified by 3 annotators + adjudication.

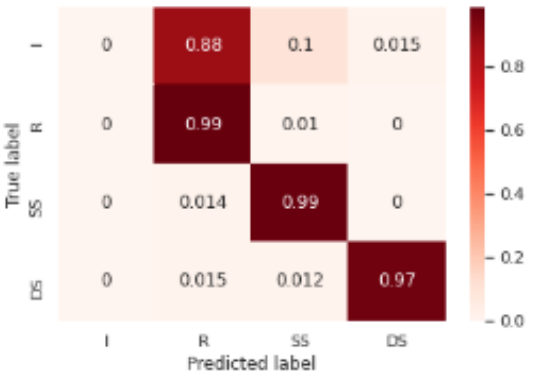
Classification Model

RoBERTa model with classification heads fine-tuned on IRSD_4^C and IRSD_3^C.

Results on the Silver Data



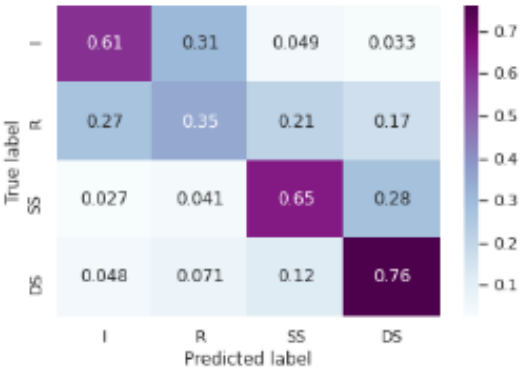
(a) 4-class



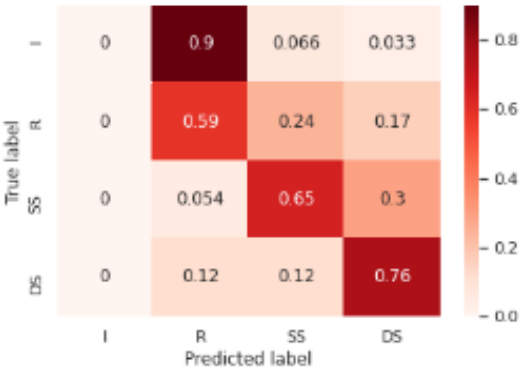
(b) 3class

- Accuracy (98%) is much higher than previous works
  - Scarton et al. (2018, 2020): 51% and 70% for a similar 4 classification task.
  - Garbacea et al. (2021): 81% for a binary simp/nosimp classification.

Results on the Gold Data



(a) 4-class



(b) 3-class

Lower than on the silver data.

- Identity examples are often predicted as rephrase
- Syntax split as discourse-split
- Rephrase examples regularly receive predictions across all four classes.

Controlling Sentence Simplification

Training Data

(C,o,S) triplets from IRSD_4^C and IRSD_3^C

Test Data

Subset of $\text{IRSD}_{3/4}^S$

Newsela-auto-M

- test set introduced by Maddela et al. (2021)
- 24,035 rephrases, 9,208 syntax-splits, and 148 discourse-splits.
- for comparison with pre-existing system outputs from past works

ASSET corpus (Alva-Manchego et al., 2020)

- 359 examples with 10 human-written references per input.
- All test examples have at least one ***rephrase*** reference
- 248 have at least one ***syntax-split*** reference
- 12 have at least one ***discourse-split*** reference
- 0 have an ***identity*** reference

Previous Work

Hybrid (Narayan and Gardent, 2014)

- statistical system combining a probabilistic splitting component with an MT-based lexical paraphraser

BERT (Jiang et al., 2020)

- finetuned on simplification data

EditNTS (Dong et al., 2019)

- a model using operation prediction

MadExP (Maddela et al., 2021)

- state-of-the-art controllable system

Models

Baseline End-to-end Model

BART: C $\rightarrow$ S

- BART$_{3/4}$: fine tuned on IRSD$^{S}_{3/4}$
- BART$_{W}$: fine tuned on Wiki-auto
- BART$_{N}$: fine tuned on Newsela-auto

- 4-class classifier $\rightarrow$ more conservative outputs
- 3-class generator $\rightarrow$ more model capacity to focus on simplification.
- 4-class generator: more training data could improve general performance.

Controllable Models

BART: (C, o) $\rightarrow$ S

- o is an operation label prepended to the input sequence for C

Ctrl$_{Oracle}$

- End-to-end
- o is the silver label from the dataset

Ctrl$_{i,j}$

- Pipeline
- o is predicted
- i,j is the number of classes the classifier/generator is trained on

Results

Model	IRSD_4^S				IRSD_3^S		Newsela-auto			
	P_{BERT}	SARI	R_{Split}	P_{Split}	P_{BERT}	SARI	P_{BERT}	SARI	R_{Split}	P_{Split}
Input	0.83	27.4	0.00	0.00	0.77	25.4	0.53	15.9	0.00	0.00
Reference	0.99	80.1	1.00	1.00	0.99	95.3	0.99	94.1	1.00	1.00
End-to-End Models										
BART_W	0.81	35.0	0.18	0.85	0.76	34.7	0.54	24.6	0.05	**0.64**
BART_N	0.77	38.9	0.64	0.81	0.74	42.0	**0.56**	**35.9**	0.46	0.59
BART_3	0.85	50.6	0.82	0.94	0.81	54.9	0.55	27.3	0.27	0.59
BART_4	0.86	51.2	0.85	0.93	0.82	55.7	**0.56**	26.9	0.21	0.62
Controllable Models with predicted control-tokens										
$\text{Ctrl}_{3,3}$	0.83	50.6	**0.99**	0.93	0.82	58.5	0.54	33.6	0.48	0.54
$\text{Ctrl}_{3,4}$	0.84	51.2	**0.99**	0.93	**0.83**	59.4	0.55	**35.9**	**0.49**	0.54
$\text{Ctrl}_{4,3}$	0.86	52.9	**0.99**	**0.98**	**0.83**	59.5	0.55	30.7	0.45	0.56
$\text{Ctrl}_{4,4}$	**0.87**	**55.1**	**0.99**	**0.98**	**0.83**	**60.4**	**0.56**	32.4	0.45	0.56
Controllable Models with Oracle control-tokens										
Ctrl_{Oracle}	**0.87**	**55.5**	**1.00**	**1.00**	**0.83**	**60.7**	**0.57**	**38.3**	**0.99**	**1.00**

- **Data:** Models trained on IRSD^S outperform those trained on Newsela and Wiki-auto
- **Control:** On Newsela test set, $\text{Ctrl}_{3,*}$ yields ***higher SARI and R Split*** than $\text{Ctrl}_{4,*}$ showing that the ***operation label helps ensuring that the input is simplified*** .

Results

Model	IRSD_4^S				IRSD_3^S		Newsela-auto			
	P_{BERT}	SARI	R_{Split}	P_{Split}	P_{BERT}	SARI	P_{BERT}	SARI	R_{Split}	P_{Split}
Input	0.83	27.4	0.00	0.00	0.77	25.4	0.53	15.9	0.00	0.00
Reference	0.99	80.1	1.00	1.00	0.99	95.3	0.99	94.1	1.00	1.00
End-to-End Models										
BART_W	0.81	35.0	0.18	0.85	0.76	34.7	0.54	24.6	0.05	**0.64**
BART_N	0.77	38.9	0.64	0.81	0.74	42.0	**0.56**	**35.9**	0.46	0.59
BART_3	0.85	50.6	0.82	0.94	0.81	54.9	0.55	27.3	0.27	0.59
BART_4	0.86	51.2	0.85	0.93	0.82	55.7	**0.56**	26.9	0.21	0.62
Controllable Models with predicted control-tokens										
$\text{Ctrl}_{3,3}$	0.83	50.6	**0.99**	0.93	0.82	58.5	0.54	33.6	0.48	0.54
$\text{Ctrl}_{3,4}$	0.84	51.2	**0.99**	0.93	**0.83**	59.4	0.55	**35.9**	**0.49**	0.54
$\text{Ctrl}_{4,3}$	0.86	52.9	**0.99**	**0.98**	**0.83**	59.5	0.55	30.7	0.45	0.56
$\text{Ctrl}_{4,4}$	**0.87**	**55.1**	**0.99**	**0.98**	**0.83**	**60.4**	**0.56**	32.4	0.45	0.56
Controllable Models with Oracle control-tokens										
Ctrl_{Oracle}	**0.87**	**55.5**	**1.00**	**1.00**	**0.83**	**60.7**	**0.57**	**38.3**	**0.99**	**1.00**

- Controllable systems outperform their end-to-end counterpart
- Large increase in R Split
- Ctrl_{Oracle} shows highest scores indicating that operation labels help improve simplification

Human Evaluation

- 25 I/O examples from each of the 4 classes
- C, Reference S, Model Output
- fluency, adequacy, simplicity
- Models
  - Previous work
    EditNTS, MadExp vs Ctrl$_{4,4}$, BART$_4$ ...
  - IRSD vs Newsela Training data
    BART$_N$ vs. BART$_4$
  - End-to-end vs. Controllable
    BART$_4$ vs. Ctrl$_{4,4}$

Human Evaluation

System	Fluency	Adequacy	Simplicity	Mean
Ref.	4.65**	3.95**	4.37*	4.32
EditNTS	3.81**	3.83**	3.91**	3.85
MadExP	3.74**	3.52**	3.97**	3.75
BART_N	4.68	4.26**	**4.38***	4.44
BART_4	4.71	**4.74**	4.14	4.53
$\text{Ctrl}_{4,4}$	**4.77**	**4.74**	4.20	**4.57**

Our systems (BART_N, BART_4, $\text{Ctrl}_{Ctrl\_\{3, *\}\{4, 4\}}$) are rated highly accross all criteria and receive better average scores than even the reference.

The relatively low adequacy rating given to the reference can partly be attributed to cases where S makes reference to terms mentioned earlier in their article that are not explicit in C.

Summary

- Our ***new dataset*** helps learning ***classifiers of much higher accuracy than previous work***
- Our controllable Sentence Simplification models
  - ***outperform end-to-end baselines and previous work***
  - receive high ratings in fluency, adequacy and simplicity from human evaluators
  - have ***lower simplicity ratings compared to reference texts***

Document Simplification

Previous work

Mostly on sentence simplification

Document simplification

- Sentence-level techniques iteratively applied over a document (Woodsend and Lapata, 2011a; Alva-Manchego et al., 2019b)

  Insufficient for maintaining the discourse coherence of the document (Siddharthan, 2003; Alva-Manchego et al., 2019b).

- Sub-problems of simplification
  - paraphrasing and sentence re-ordering (Lin et al., 2021)
  - insertion (Srikanth and Li, 2021) or
  - deletion (Zhong et al., 2020; Zhang et al., 2022).

  Only consider a limited set of operations

- A sentence-level model that uses context information to influence document simplification (Sun et al. 2020)

  Unable to outperform the baseline (Sun et al. 2021)

Hypothesis

We hypothesize that a document- level simplification model that is based on a plan specifying a simplification operation for each input sentence should fare better than a simplification model that directly simplifies an entire document

Simplification decomposed into a two-stage process

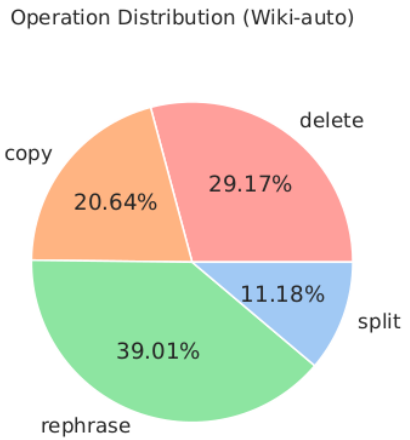
$$p(S \mid C) = p(S \mid C, P)p(P \mid C)$$

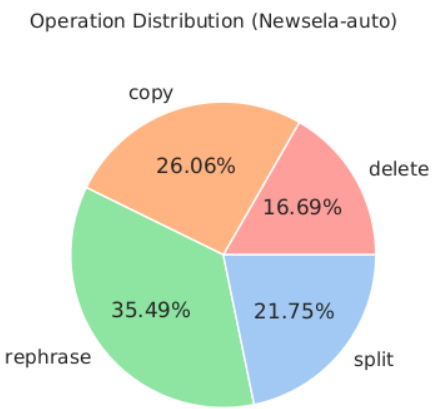
- $C = c_1 \dots c_n$
  the input document, a sequence of complex sentences c_i
- $S = s_1 \dots s_k$
  a sequence of simplified sentences s_i
- $P = o_1 \dots o_n$
  A sequence of sentence-level simplification operations o_i for C (a simplification ***plan***).
- Four sentence-level simplification operations
  copy, rephrase, split, **delete**

Cripwell et al. EACL 2023

Data

Complex/Simplified Documents with sentence alignment





	Wiki-auto	Newsela-auto
# Doc Pairs	85,123	18,319
# Sent Pairs	461,852	707,776
Avg. $\|C\|$	155.51	868.98
Avg. $\|S\|$	97.72	674.94
Avg. $\|c_i\|$	28.64	22.49
Avg. $\|s_i\|$	21.57	15.84
Avg. n	5.43	38.64
Avg. k	4.53	42.60

- n: the number of sentences in C
- k: the number of sentences in S

Labeling the data

Delete

- c_i is not aligned to any s_j .

Copy

- c_i is aligned to a single s_j with a Levenshtein similarity above 0.92.

Rephrase

- c_i is aligned to a single s_j with a Levenshtein similarity below 0.92.

Split

- c_i is aligned to multiple s_j s.

Planning Simplification Operations

Given some input document

$$C = c_1, \ldots, c_n$$

the task of the planner is to ***predict a sequence of n simplification operations***

$$\hat{P} = \hat{o}, \ldots, \hat{o}_n$$

with $\hat{o}_i \in$ {copy, rephrase, split, delete}

Challenges

Operations have different requirements

- Splitting is mostly ***context independent*** as it is mainly determined by the ***input sentence's internal structure***

  The man who sleeps snores → The man sleeps. He snores.

  John went shopping after he left work → John left work. Afterwards he went shopping.

- Deletion and to a lesser extent, copy and rephrase are mostly ***context dependent*** . A sentence can only be omitted if it is either redundant with, or of minor semantic import relative to, other sentences in the document

Model

RoBERTa classifier with cross-attention over the context

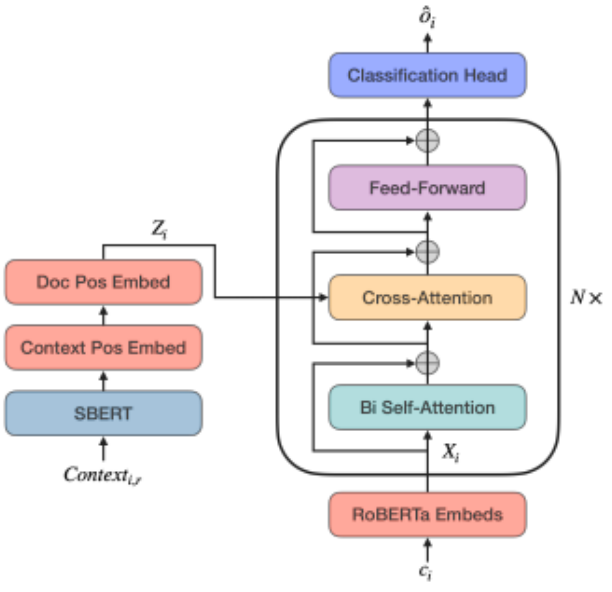
- layers initialised with weights from a context-independent classifier

Internal structure

- ***Token level*** encoder for c_i

Context

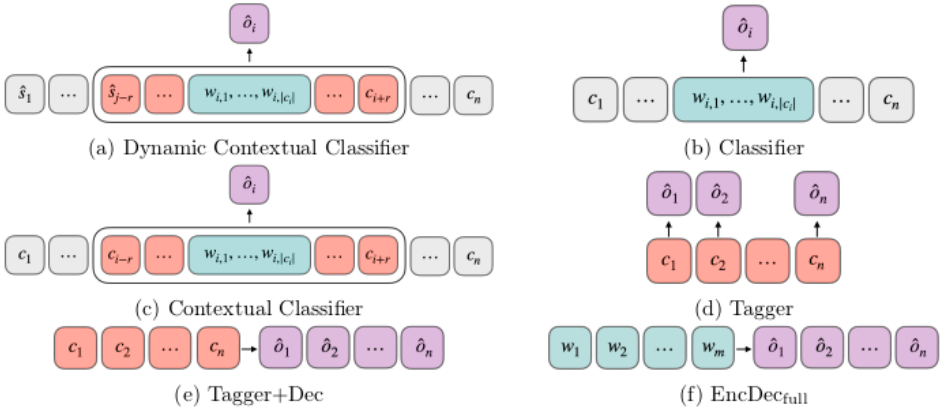
- fixed window of Sentence level embedding (SBERT) for ***surrounding sentences***
- The left context is ***dynamically*** updated with previously simplified sentences



Context positional embedding: relative distance of a given sentence from the input sentence c_i

Document positional embedding: the document quintile (1-5) that a given sentence falls into

Alternative Models



Dynamic Contextual Classifier: our model

Contextual Classifier: Static left context

Classifier: no context

Tagger: Sequence tagging on SBERT representations (no access to the internal structure of c_i)

Tagger-Decoder: Each prediction is conditioned on the input document and on the previously predicted operation tags. SBERT encodings.

EncDec $_{full}$: Same as Tagger-Decoder but with token encodings

Planning Accuracy Results

	Wiki-auto						Newsela-auto					
Model	C	R	S	D	Micro	**Macro**	C	R	S	D	Micro	**Macro**
EncDec$_{full}$	26.9	42.2	36.0	51.8	43.2	40.8	26.1	10.8	11.7	9.0	12.2	11.5
EncDec	29.3	54.5	30.0	51.8	47.7	41.4	72.2	73.9	75.9	79.7	75.0	75.4
Tagger	38.6	54.2	31.7	**58.5**	50.6	45.8	71.4	72.7	74.1	78.4	73.7	74.1
Classifier	42.1	52.9	42.6	49.0	48.4	46.7	77.0	75.6	80.0	78.5	77.4	77.8
Dyn. Context	**44.8**	**57.9**	42.4	54.8	**52.8**	**50.0**	79.3	77.3	82.8	81.4	79.7	80.2
+ docpos	43.7	55.4	**43.6**	56.7	52.3	49.9	**80.0**	**78.1**	**83.6**	**82.0**	**80.3**	**80.8**

- Our model consistently shows best results on both datasets.
- The ***context-free classifier under-performes for deletions***, which confirms the intuition that context modeling particularly matters for this operation.
- ***EncDec full performs worst*** presumably because the very long input (the whole context is modelled at the token level) challenges the attention mechanism
- The encoder-decoder and the tagger, which both use a ***sentence level encoding of the complex sentence*** to be classified perform worse than the classifier - this highlights the importance of having a ***token-level modeling of the input sentence*** .

Ablations

Model	Copy	Rephrase	Split	Delete	Micro	**Macro**
(a) Ablation on Best Model						
Dyn, $r = 13$, +init, +docpos	80.0	78.1	83.6	82.0	80.3	80.8
-docpos	79.3	77.3	82.8	81.4	79.7	80.2
-init	74.9	72.1	77.8	75.2	74.6	75.0
-init, -docpos	75.6	72.0	77.7	77.1	75.1	75.6
(b) Dynamic vs. Static Context						
Stat, $r = 9$	71.3	69.5	75.4	73.3	72.0	72.4
Stat, $r = 13$	72.2	65.3	69.9	68.3	68.5	68.9
Dyn, $r = 9$	73.1	70.1	75.5	75.9	73.1	73.6
Dyn, $r = 13$	75.6	72.0	77.7	77.1	75.1	75.6
(c) With vs without Initialisation						
Dyn, $r = 9$	73.1	70.1	75.5	75.9	73.1	73.6
Dyn, $r = 9$ +init	79.3	78.0	82.7	79.8	79.7	80.0
Dyn, $r = 13$	75.6	72.0	77.7	77.1	75.1	75.6
Dyn, $r = 13$ +init	79.3	77.3	82.8	81.4	79.7	80.2
(d) Window Size						
Stat, $r = 9$	71.3	69.5	75.4	73.3	72.0	72.4
Stat, $r = 13$	72.2	65.3	69.9	68.3	68.5	68.9
Dyn, $r = 9$	73.1	70.1	75.5	75.9	73.1	73.6
Dyn, $r = 13$	75.6	72.0	77.7	77.1	75.1	75.6
Dyn, $r = 9$ +docpos	73.8	72.9	77.2	75.8	74.6	74.9
Dyn, $r = 13$ +docpos	74.9	72.1	77.8	75.2	74.6	75.0
Dyn, $r = 9$ +init +docpos	79.4	78.0	83.1	82.0	80.1	80.6
Dyn, $r = 13$ +init +docpos	80.0	78.1	83.6	82.0	80.3	80.8

Simplifying Documents

Models

Doc-BART

- fine-tuned on full document pairs

Sent-BART

- fine-tuned on sentence pairs and iteratively applied to each input sentence

PG (plan-guided)

- fine-tuned on sentence pairs and iteratively applied to each input sentence
- simplification is conditioned on both the input complex sentence and a predicted simplification operation

Results

System	BARTScore ↑				SMART ↑			FKGL ↓	SARI ↑	Length	
	Faith. ($s \rightarrow h$)	P ($r \rightarrow h$)	R ($h \rightarrow r$)	F1	P	R	F1			Tokens	Sents
Input	-0.93	-2.47	-1.99	-2.23	63.2	62.7	62.8	8.44	20.52	866.9	38.6
Reference	-1.99	-0.93	-0.93	-0.93	100	100	100	4.93	99.99	671.5	42.6
Doc-BART	-2.48	-2.68	-2.76	-2.72	61.9	43.9	50.6	10.01	47.07	600.8	20.7
Sent-BART	**-1.86**	-1.63	-1.56	-1.60	78.9	80.1	79.3	5.03	73.02	666.4	42.6
PG_{Tag}	-1.95	-2.22	-2.18	-2.20	5.07	62.0	62.6	61.6	56.13	657.4	41.8
PG_{EncDec}	-1.94	-2.22	-2.18	-2.20	62.2	62.5	61.6	5.09	56.06	654.2	41.4
PG_{Clf}	-1.91	-1.68	**-1.53**	-1.60	77.8	**81.2**	79.3	**4.95**	73.83	688.8	44.5
PG_{Dyn}	-1.91	**-1.60**	-1.54	**-1.57**	**80.2**	81.0	**80.5**	4.98	**75.00**	667.2	42.6
PG_{Oracle}	-1.93	**-1.39**	**-1.40**	**-1.40**	**85.5**	**85.0**	**85.3**	**4.91**	**80.74**	655.6	42.1

PG Dyn achieves the highest results of all systems.

PG Oracle leads to a substantial further increase in performance highlighting the impact of planning

Doc-BART is by far the worst performing system, presumably a result of it failing to properly handle the long document lengths.

Example output

Complex	Silvano "Nano" Campeggi (January 23, 1923 – August 29, 2018) was an Italian artist who designed and produced the artwork for the posters of many classic Hollywood films. His iconic images are associated with the golden era of Hollywood and Campeggi is now generally regarded as the most important graphic artist and poster designer in the history of American cinema. In the following decades, Campeggi designed and produced the poster and advertising graphics for over 3000 films, working not only under contract with the MGM studios, but also with Warner Brothers, Paramount, Universal, Columbia Pictures, United Artists, RKO, Twentieth-Century Fox and several other movie studios. Sixty-four of the films he illustrated won Oscars, including "Casablanca", "Ben-Hur", "Singin' in the Rain", "An American in Paris", "West Side Story", "Exodus", "Breakfast at Tiffany's", and "Gigi". Campeggi died on 29 August 2018, at the age of 95.
Simple	Silvano "Nano" Campeggi (January 23, 1923 – August 29, 2018) was an Italian artist. He designed and produced the artwork for the posters of many classic Hollywood movies. Campeggi was known for his poster designs for "Casablanca", "Singin' in the Rain", and "Breakfast at Tiffany's". Campeggi died on August 29, 2018 in Florence from respiratory failure at the age of 95.
Doc-BART	Silvano "Nano" Campeggi (January 23, 1923 – August 29, 2018) was an Italian artist. He designed and produced the artwork for the posters of many classic Hollywood movies, including "Casablanca", "Ben-Hur", "Singin' in the Rain", "An American in Paris", "West Side Story", "Exodus", "Breakfast at Tiffany's", "Gigi", and "".
Sent-BART	Silvano "Nano" Campeggi (23 January 1923 – 29 August 2018) was an Italian artist. He designed and produced the artwork for the posters of many classic Hollywood movies. Campeggi is generally regarded as the most important graphic artist and poster designer in the history of American cinema. Campeggi designed and produced the poster and advertising graphics for over 3000 movies, working not only under contract with the MGM studios, but also with Warner Brothers, Paramount, Universal, Columbia Pictures, United Artists, RKO, Twentieth Century Fox and several other movie studios. Sixty-four of the movies he illustrated won Oscars, including "Casablanca", "Ben-Hur", "Singin' in the Rain", "An American in Paris", "West Side Story", "Exodus", "Breakfast at Tiffany's", and "Gigi". Campeggi died on 29 August 2018, at the age of 95.
PG_{Dyn}	Silvano "Nano" Campeggi (January 23, 1923 – August 29, 2018) was an Italian artist. He designed and produced the artwork for the posters of many classic Hollywood movies. Sixty-four of the movies he illustrated won Oscars, including "Casablanca", "Ben-Hur", "Singin' in the Rain", "An American in Paris", "West Side Story", "Exodus", "Breakfast at Tiffany's", and "Gigi". Campeggi died on 29 August 2018 at the age of 95.

Questions ?