



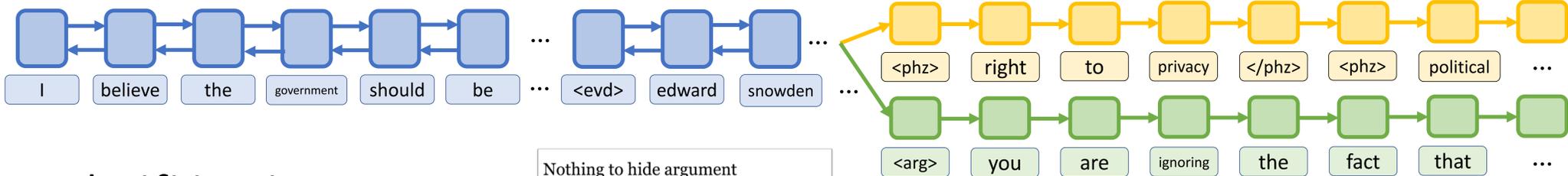
# Neural Argument Generation Augmented with Externally Retrieved Evidence



Xinyu Hua and Lu Wang  
College of Computer and Information Science  
Northeastern University



Github repo:  
Xinyuhua/neural-argument-generation/



## >>> Input Statement

I believe the government should be allowed to view my emails for national security concerns. CMV.

I have nothing to hide. I don't break the law...



Nothing to hide argument  
From Wikipedia, the free encyclopedia

The **nothing to hide argument** states that government surveillance programs commit these activities does not have the right to keep them secret. Hence that a p

**Political corruption**  
From Wikipedia, the free encyclopedia

**Political corruption** is the use of powers by government

**Surveillance**  
From Wikipedia, the free encyclopedia  
(Redirected from Surveillance)

This article is about observing person of interest; "Electronic surveillance" redirects here. For surve

**Surveillance** (/sərˈveɪ.əns/ or /sərˈveɪtiŋ/)<sup>[1]</sup> is the m

## >>> Ranked Evidence Sentences

1. Edward Snowden: "Arguing that you don't care about **right to privacy** because..."
2. **Political corruption** is the use of powers by government officials for illegitimate private gain.

## MOTIVATION

- Argumentation is crucial in human reasoning and decision making.
- Constructing high quality arguments is challenging in that it needs to be both informative and persuasive.
- Previous work is mostly template or rule based, lacking language flexibility.

## FRAMEWORK

1. **Document Retrieval**: using topic signatures in OP to retrieve Wikipedia articles.
2. **Sentence Reranking**: breaking articles into sentences and rerank them base on TF-IDF similarity.
3. **Encoding**: using BiLSTM to encode both the OP and retrieved evidence sentences.
4. **Keyphrase Decoding**: generating keyphrases as talking points.
5. **Argument Decoding**: generating the final argument.

## CONTRIBUTIONS

1. We study the novel task of generating arguments of a different stance for a given statement.
2. Based on r/ChangeMyView, we construct a large-scale dataset for argument generation research.
3. We propose an encoder-decoder style neural model with external information retrieved from Wikipedia to generate both keyphrases and arguments.

## SAMPLE OUTPUTS

Original Post	Generated Argument
<p>CMV: Putin is trying to re-form a "Soviet Union" with his past actions in Georgia and current actions in Ukraine.</p> <p>...I firmly believe that Putin and the Russian Federation (RF) are trying to re-form a Soviet Union type regime. The Russian Army invaded certain regions of Georgia and claimed that Georgia was being aggressive towards these regions...</p>	<p>There are two reasons you are so far beyond. There is no reason to see the <i>military army</i>. You can infer what they assume it, so they tend to <i>protect</i> up. There. If we assume it were bad, they can not be controlled. So we're talking going to ignore the <i>wars of the world</i>. The fact of the matter is not bad. I would also assume it. However, the <i>government</i> can not be reminded of the world.</p>

## DATA



↑	<b>[OP]</b> CMV: Putin is trying to re-form a "Soviet Union" with his past actions in Georgia and current actions in Ukraine ...I firmly believe that Putin and the Russian Federation (RF) are trying to re-form a Soviet Union type regime...
↑ ↓	<b>[U1]</b> There is a very large difference between taking land that belonged to Imperial Russia/Soviet union in the past and actually reforming the Soviet union and turning back to communism...
↑ ↓	<b>[U2]</b> So far Putin has only really targeted areas where the local population is majority Russian...I don't think there is yet any compelling reason to believe he would.

**Data source:** subreddit r/ChangeMyView – for open discussion and debate  
**Collection:** Jan 2013 – Jun 2017, about 27K threads in total  
**Domain:** Politics and policy, about 13K threads  
**Filtering:** only high quality root replies are considered

**Delta:** acknowledgement of persuasion  
**Karma:** upvote - downvote

	Input statement	Human argument
Count	12,549	117,960
Avg number of sentences	16.1	7.7
Avg number of tokens	356.4	161.1

## EVALUATION

### Experiment setup:

- RETRIEVAL: concatenates retrieved evidence sentences.
- *System* vs. *Oracle* retrieval: whether use OP or gold-standard argument to construct query.

### Evaluation metrics:

1. BLEU-2: precision based up to bigram
2. METEOR: unigram precision and recall based on alignment.
3. Multireference: if more than one gold-standard arguments exist, we pick the best aligned one.

### Human judgement:

- Three trained judges
- *Grammaticality*
- *Informativeness*
- *Relevance*

	w/ System Retrieval			w/ Oracle Retrieval		
	BLEU	MTR	Len	BLEU	MTR	Len
<b>Baseline</b>						
RETRIEVAL	15.32	<b>12.19</b>	151.2	10.24	<b>16.22</b>	132.7
<b>Comparisons</b>						
SEQ2SEQ	10.21	5.74	34.9	7.44	5.25	31.1
+ encode evd	18.03	7.32	67.0	13.79	10.06	68.1
+ encode KP	21.94	8.63	74.4	12.96	10.50	78.2
<b>Our Models</b>						
DEC-SHARED	21.22	8.91	69.1	15.78	11.52	68.2
+ attend KP	<b>24.71</b>	10.05	74.8	11.48	10.08	40.5
DEC-SEPARATE	24.24	10.63	88.6	17.48	13.15	86.9
+ attend KP	24.52	11.27	88.3	<b>17.80</b>	13.67	86.8

System	Grammaticality	Informativeness	Relevance
RETRIEVAL-BASED	4.5 ± 0.6	3.7 ± 0.9	3.3 ± 1.1
SEQ2SEQ	3.3 ± 1.1	1.2 ± 0.5	1.4 ± 0.7
OUR MODEL	2.5 ± 0.8	1.6 ± 0.8	1.8 ± 0.8