Sentence-Level Content Planning and Style Specification for Neural Text Generation
(Supplementary Material)

Xinyu Hua and **Lu Wang**
Khoury College of Computer Sciences
Northeastern University
Boston, MA 02115
`hua.x@husky.neu.edu`   `luwang@ccs.neu.edu`

A   Data Collection Details

Rule-based Argumentative Style Label Construction.   In §4, we mention a set of rules to automatically label sentence styles for argument generation. The goal is to capture the argumentative discourse function using common patterns. The complete version of the rules for CLAIM and PREMISE are listed in Table 1.

Wikipedia Sentence Length Distribution.   As is described in §4.2, we assign the sentence style labels for Wikipedia data based on the sentence length. In Figure 1 we show the distribution of sentence lengths for normal and simple Wikipedia. For both versions, the majority fall into the $[10, 30]$ range. The normal version tends to contain more longer sentences and fewer shorter sentences than its simple counterpart. Based on these observations, we choose the four intervals with a step size of 10 words for length categories, to ensure the balance of data samples across categories.
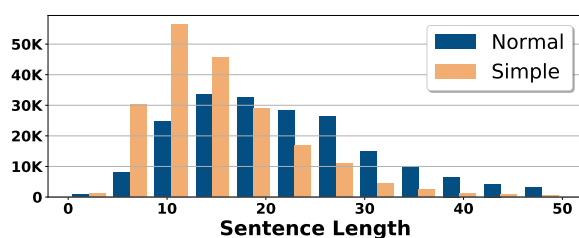

Figure 1: The distribution of sentence length in normal and simple Wikipedia. The normal Wikipedia contains more sentences in the longer range, and the opposite is true for its simple counterpart.

B   Human Evaluation Guidelines

We conduct human evaluation on the generated arguments and Wikipedia paragraphs. For argument generation, we randomly choose 33 topics from the test set, from which the first 3 topics are used merely for the human judges to calibrate among their own standards. The remaining 30 topics are used for the final evaluation. We show the guidelines for evaluation on argument data in Table 2.

On Wikipedia data, we randomly select 16 topics from the test set, and present both the normal and simple output during the evaluation. The first topic (2 samples) are used for rater calibration, while the rest 15 topics (30 samples) are kept for analysis. The guidelines are listed in Table 3.

For both tasks, we consider our model and an ablated model where style specification is disabled. We present these two system outputs alongside human constructed ones, and shuffle them for each sample to eliminate the biases associated with the order.

C   Sample Output

We show more sample outputs on all three tasks from Figure 2 to 7. We highlight our model generation among the human constructed texts and oracle plan guided generation.

Rule	Patterns
	CLAIM
Belief	`i (don't)? (believe|agree|concede|suspect|doubt|see|feel|understand)`
Imperative	`(any|anyone|anybody|every|everyone|everybody|all|most|few|no|no one` `|nobody|it|we|you|they|there) \w{0,10} (could|should|might|need|must)`
Sense	`(it|this|that) make (no|zero)? sense`
Chance	`(chance|likelihood|possibility|probability) .* (slim|zero|negligible)`

Evaluation	`(be|seem) (necessary|unnecessary|moral|immoral|right|wrong|stupid`
	`|unconstitutional|costly|inefficient|efficient|reasonable|beneficial`
	`|important|unfair|harmful|justified|jeopardized|meaningless|flawed`
	`|justifiable|unacceptable|impossible|irrational|foolish)`

Miscellaneous	`(in my opinion|imo|my view|i be try to say|have nothing to do` `with|tldr)`
	PREMISE
Affect	`(help|improve|reduce|deter|increase|decrease|promote)`
Example	`(for example|for instance|e.g.)`

Table 1: Patterns for sentence style label construction on CLAIM and PREMISE for argument generation.

In the following survey, you will read 33 short argumentative text prompts and evaluate 3 counter-arguments for each of them. Please rate each counter-argument on a scale of 1-5 (the higher the better), based on the following three aspects:

- **Grammaticality:** whether the counter-argument is fluent and has no grammar errors
  - 1. *the way the way etc. 'm not 's important*
  - 3. *is a good example. i don't think should be the case. i're not going to talk whether or not it's bad.*
  - 5. *i agree that the problem lies in the fact that too many representatives do n't understand the issues or have money influencing their decisions.*

- **Correctness:** whether the counter-argument is relevant to the topic and of correct stance
  - 1. *i don't think it 's fair to say that people should n't be able to care for their children*
  - 3. *i don't agree with you and i think legislative bodies do need to explain why they vote that way*
  - 5. *there are hundreds of votes a year . how do you decide which ones are worth explaining ? so many votes are bipartisan if not nearly unanimous . do those all need explanations ? they only have two years right now and i do n't want them spending less time legislating .*

- **Content richness:** whether the counter-argument covers many talking points
  - 1. *i do n't agree with your point about legislation but i 'm not going to change your view.*
  - 3. *i agree that this is a problem for congress term because currently it is too short.*
  - 5. *congressional terms are too short and us house reps have to spend half of their time campaigning and securing campaign funds. they really have like a year worth of time to do policy and another year to meet with donors and do favors.*

Table 2: Evaluation guidelines on argument data and representative examples on rating scales.

In the following survey, you will read 32 samples. Each of them contains a topic, and 3 text snippets explaining the topic. For each of the explanation, please rate it on a scale of 1-5 (the higher the better) based on the following three aspects:

- **Grammaticality:** whether the explanation is fluent and has no grammar errors

  - 1. *android android operating system mobile kindle is a modified operating*
  - 3. *android is an operating system for mobile mobile mobile and other manufactures like htc and the*
  - 5. *android is an operating system for mobile devices . it is also used by other manufactures like htc and samsung .*

- **Correctness:** whether the explanation contains obvious semantic mistakes or contradictions. Please note that this is NOT intended for fact checking, so you should not find other resources to determine if the concrete information (such as years, locations) are wrong, instead please apply commonsense level knowledge to judge the correctness

  - 1. *android is used for tablets such as amazon.com as well as other phone such as linux and amazon*
  - 3. *android is an operating system for android and devices .*
  - 5. *android is an operating system for mobile devices .*

- **Content richness:** whether the explanation covers the amount of information that is necessary to explain the topic

  - 1. *modified mobile mobile android*
  - 3. *android is an operating system used for mobile devices .*
  - 5. *android is an operating system for mobile devices , it is mostly used for, like google 's own google pixel, as well as by other phone manufacturers like htc and samsung .*

Table 3: Evaluation guidelines on Wikipedia data and representative examples on rating scales.

Topic: CMV: All alleged crimes or infractions involving a police officer should be investigated by a higher-level agency.
Human: if it was required to take this seriously , people who want to protest the police would just file complaints against all the police in a department constantly . are you going to make it a crime to file a complaint if the review board determines there is n't enough evidence to convict the cop ? they could just keep the complaint vague to avoid charges of filing false reports . every person who has a cop car pass them could file a complaint that the cop is stalking them.
Our model: jail sentences for example are a result of negligence . if you do n't like it , you 're going to have to deal with a lot of bad lawyers , and you 'll have to prove you 're guilty of a crime in the first place and you 're not going to be able to sue for damages if you 're in jail , you should n't have to
Our model (Oracle Plan): additionally , there is a difference between the two cases in the united states and between the federal government and the department of justice . in the case of a police officer , the fbi has a duty to report a crime , and it is not required to report the crime . this is not the case , it is the victim 's job to determine whether or not

Figure 2: Sample outputs on argument generation.

Topic: CMV:High Heeled shoes are just as "oppressive" as the face veil/niqab
Human: the niqab has the unique property of covering a person 's face , which is where we tend to ascribe people 's personality and emotion . in any media critique , calling someone faceless means that they lack humanization . it 's usually the default way to mark someone as either a villain or an acceptable target for consequence-free violence .
Our model: i live in a small town in the united states . there are a lot of things i do n't want to do , but it 's not a bad thing for me to think about it as a matter of personal experience , and i think it 's important to keep in mind that it 's something that can be seen as a good thing in the long run
Our model (Oracle Plan): lastly , a woman 's right to bodily autonomy is not the same as being a woman . it 's not a matter of whether or not a woman has a right to her body , it 's about her ability to make decisions about her own body . she has the right to do whatever she wants with her body and her body is her right to use her

Figure 3: Sample outputs on argument generation.

Topic: Breaking Bad
Simple Wikipedia: Breaking Bad is an American television series set in Albuquerque, New Mexico. It started in January 2008. The show was broadcast across Canada and the United States on cable channel AMC. It has won 10 Emmy Awards. Breaking Bad ended in September 2013. Bryan Cranston plays the main role, Walter White. There are five seasons and 62 episodes. It is about a chemistry teacher who is told he has lung cancer and starts making the illegal drug methamphetamine to pay for his family's needs after he dies. Breaking Bad was made by Vince Gilligan.
Our model: bad breaking is an american television series set in albuquerque , new mexico . it started in january 2008 and on january 20 , 2011 .
Our model (Oracle Plan): breaking bad is an american television series set in albuquerque , new mexico . it started in january 2008 . the show has been broadcast across canada and the united states on cable channel amc . it has won emmy awards 10 and 3 respectively since it ended in september 2013 ended on september 2013 after breaking bad breaking itself in 2013 and 2013 respectively cable channel channel 2 ended in 2013) cable cable channel cable 10 .
Normal Wikipedia: Breaking Bad is an American neo-Western crime drama television series created and produced by Vince Gilligan. The show originally aired on AMC for five seasons, from January 20, 2008 to September 29, 2013. Set and filmed in Albuquerque, New Mexico, the series tells the story of Walter White, a struggling and depressed high school chemistry teacher who is diagnosed with stage-3 lung cancer. Together with his former student Jesse Pinkman, White turns to a life of crime by producing and selling crystallized methamphetamine to secure his family's financial future before he dies, while navigating the dangers of the criminal world. The title comes from the Southern colloquialism "breaking bad" which means to "raise hell" or turn to a life of crime.
Our model: bad breaking is an american an american neo-western crime drama television series an american neo-western crime drama television series drama series television written and produced by vince vince . the show aired on amc for five seasons from january 20 , 2011 to september 29 , 2013 .
Our model (Oracle Plan): bad breaking is an american an american neo-western crime drama television series an american neo-western crime drama television series drama series television written and produced by vince gilligan . the show originally aired on amc for five seasons from january 20 , 2011 to september 29 , 2013 . the series was filmed in filmed in albuquerque , new mexico and tells the story of walter walter , a highly high school depressed depressed chemistry school diagnosed with lung with lung cancer . the novel follows a life of student and former jesse SOS student , walter white methamphetamine , and navigating the navigating of his family in the world the the world his future financial and depressed high school chemistry teacher ' depressed high school chemistry teacher SOS ' " the breaking bad SOS " is about a life in the criminal world , with white and walter white methamphetamine and the navigating the dangers of financial financial and depressed high school in new mexico 's financial financial , an american neo-western crime drama television series .luca luca badoer was an formula one italian racing driver . he was born in born in montebelluna . he raced for the scuderia the minardi , forti corse corse , forti corse and teams ferrari and raced for his raced for italia raced for raced for the scuderia italia raced for them as well as for raced for the scuderia italia for the raced for both the raced for the scuderia italia in addition to his racing duties , he also raced for and raced internationally for the prix de forti corse and competed for the for the right and raced against the scuderia raced for racing with the for for his racing he was also active in the european prix for reserve reserve racing , qualifying for the grand prix de corse and michael michael 's ferrari teams as well .

Figure 4: Sample outputs on Wikipedia generation.

Topic: Kingdom of Iceland
Simple Wikipedia: The Kingdom of Iceland was a constitutional monarchy that had a personal union with Denmark. The kingdom was created on 1 December 1918. It lasted until 17 June 1944 when a national referendum created the Republic of Iceland.
Our model: the kingdom of iceland was a monarchy in the kingdom and union of iceland . it was created on 1 december created out of a personal union , the republic of denmark , in 1918 1945 .
Our model (Oracle Plan): the kingdom of iceland was a personal union of iceland and union with a monarchy . the kingdom was created on 1 december 1 , 1944 . it lasted until 1944 , when iceland officially created the republic of iceland in a national referendum . iceland created the created republic of denmark , which lasted until june until a national national referendum in 1944 . the name " iceland " lasted until 1945 , when it became a a a national a referendum .
Normal Wikipedia: The Kingdom of Iceland was a constitutional monarchy, a sovereign and independent country that was established by the Act of Union with Denmark signed on 1 December 1918. It lasted until 17 June 1944 when a national referendum established the Republic of Iceland in its place.
Our model: the kingdom of iceland , officially the kingdom " iceland " , was a constitutionally independent country and a sovereign of the kingdom . it was established by the act of 1918 and signed on 1 december 1 , 1918 , in iceland , union of denmark , and denmark and denmark , where it existed until it ceased to exist until it lasted until june 17 , 1945 , iceland was officially established under the act on 1 may 1949 .
Our model (Oracle Plan): the kingdom of iceland was a monarchy established by the act of union of a sovereign and independent country signed on 1 december 1 , 1918 . it lasted until 1944 , when iceland established the republic as a national referendum and took place in 1944 . the name " iceland " lasted until 12 june june , when it established the established republic of iceland . the " a national referendum " took place after a a a national national national a constitution that established the the republic that lasted until june 1944 . it existed until 2 june until a national a referendum took place outside iceland in may 1944 .

Figure 5: Sample outputs on Wikipedia generation.

Topic: written-domain language modeling for automatic speech recognition .
Human: language modeling for automatic speech recognition systems has been traditionally in the verbal domain . in this paper , we present finite-state modeling techniques that we developed for language modeling in the written domain . the first finite-state modeling techniques we describe is for the verbalization of written-domain vocabulary items , which include lexical and non-lexical entities . the second finite-state modeling techniques is the decomposition – recomposition approach to address the out-of-vocabulary and the data sparsity problems with non-lexical entities such as urls , e-mail addresses , phone numbers , and dollar amounts . we evaluate the proposed written-domain language modeling approaches on a very large vocabulary speech recognition system for en-glish . we show that the written-domain language modeling approaches improves the speech recognition and the asr transcript rendering accuracy in the written domain over a baseline system using a verbal-domain language model . in addition , the written-domain language modeling approaches is much simpler since written-domain language modeling approaches does not require complex and error-prone text normalization and denormalization rules , which are generally required for verbal-domain language modeling .
Our model: in this paper , we propose an approach to automatic speech recognition systems based on verbal verbal domain verbal finite-state finite-state modeling . our approach is motivated by the observation that a modeling language can be represented in a written written written in a verbalization of written-domain vocabulary items vocabulary of verbalization of written-domain vocabulary items vocabulary . this is achieved by introducing a recomposition recomposition – approach to represent the non-lexical and lexical entities . we also present a approach to learn a set of decomposition recomposition approach . this approach can be applied to sparsity sparsity problems in the written domain . the approach is tested on the task of e-mail e-mail addresses . it is shown that it is possible to achieve accuracy comparable to that of a state-of-the-art approach . in addition , the
Our model (Oracle Plan): in this paper , we present automatic speech recognition systems based on verbal verbal domain . our approach is based on finite-state modeling of the modeling language in the form of a domain written domain . the finite-state finite-state finite-state techniques is designed to capture the verbalization of written-domain vocabulary items verbalization of written-domain vocabulary items verbalization of written-domain vocabulary items of verbalization of written-domain vocabulary items items , and exploit the relationship between entities and lexical entities . a approach based on recomposition recomposition recomposition approach , called out-of-vocabulary , is applied to the problems of out-of-vocabulary out-of-vocabulary out-of-vocabulary problems , and a approach to e-mail e-mail e-mail addresses sparsity problems with entities , entities , urls , urls and urls , and urls . experimental results show that the proposed recomposition – – recomposition – recomposition recomposition – approach can be effective for addresses e-mail addresses and asr transcript rendering rendering rendering accuracy .

Figure 6: Sample outputs on abstract generation.

Topic: optimal linear feature transformations for semi-continuous hidden markov models .
Human: linear discriminant or karhunen-lo eve transforms are established techniques for mapping features into a lower dimensional subspace . this paper introduces a uniform statistical framework , where the computation of the optimal feature reduction is formalized as a maximum-likelihood estimation problem . the experimental evaluation of this suggested extension of linear selection methods shows a slight improvement of the recognition accuracy.
Our model: in this paper , we address the problem of estimating the number of mapping features in a lower dimensional lower dimensional space . we propose a new statistical framework based on reduction of optimal feature optimal optimal optimal feature reduction . we show that the proposed method can be applied to a variety of mapping mapping and features , and demonstrate its superior performance in terms of both speed and recognition recognition recognition performance , as well as improving the accuracy of the proposed method to the best of our knowledge this is the first attempt to improve the accuracy of the mapping mapping by up to a factor of three orders of
Our model (Oracle Plan): in this paper , we address the problem of estimating the number of mapping features in a lower dimensional lower dimensional space , given a set of dimensional points of a given matrix , and a sparse set of points in a large number of variables.

Figure 7: Sample outputs on abstract generation.