Reading Like HER: Human Reading Inspired Extractive Summarization

Ling Luo[1,3], **Xiang Ao**[1,3]*, **Yan Song**[2], **Feiyang Pan**[1,3], **Min Yang**[4], **Qing He**[1,3]
[1]Key Lab of Intelligent Information Processing, Institute of Computing Technology,
Chinese Academy of Sciences, Beijing, China
[2]Sinovation Ventures
[3]University of Chinese Academy of Sciences
[4]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
{luoling18s, aoxiang, panfeiyang, heqing}@ict.ac.cn, clksong@gmail.com

Abstract

In this work, we re-examine the problem of extractive text summarization for long documents. We observe that the process of extracting summarization of human can be divided into two stages: 1) a *rough reading* stage to look for sketched information, and 2) a subsequent *careful reading* stage to select key sentences to form the summary. By simulating such a two-stage process, we propose a novel approach for extractive summarization. We formulate the problem as a contextual-bandit problem and solve it with policy gradient. We adopt a convolutional neural network to encode gist of paragraphs for rough reading, and a decision making policy with an adapted termination mechanism for careful reading. Experiments on the CNN and Daily-Mail datasets show that our proposed method can provide high-quality summaries with varied length, and significantly outperform the state-of-the-art extractive methods in terms of ROUGE metrics.

1 Introduction

Automatic text summarization has wide popularity in NLP applications such as producing digests, headlines and reports. Among the supervised methods, two main types are usually explored, namely abstractive and extractive summarizations (Nenkova et al., 2011). Compared with abstractive approaches, extractive methods are more practical and applicable as they are faster, simpler and more reliable on grammar as well as semantic information (Yao et al., 2018).

Recent studies (Cheng and Lapata, 2016; Nallapati et al., 2017; Yasunaga et al., 2017; Feng et al., 2018) consider extractive summarization as a sequence labeling task, where each sentence is individually processed and determined whether it should be extracted or not. Various neural networks are used to label each sentence and trained using cross-entropy loss to maximize the likelihood of the ground-truth labeled sequences, which may derive the mismatch between the cross-entropy objective function and the evaluation criterion. On the other hand, some reinforcement learning based methods (Wu and Hu, 2018; Narayan et al., 2018; Yao et al., 2018) directly optimize the evaluation metric by combining cross-entropy loss with rewards and train model with policy gradient reinforcement learning. Note that the rewards usually reflect the quality of extracted summary and measured by standard evaluation protocol. However, they still sequentially process text and tend to extract earlier sentences over later ones due to the sequential nature of selection (Dong et al., 2018).

Although great efforts have been devoted to this field, most of the existing approaches neglect how human being reads and forms summaries. Human beings are very good at refining the main idea of a given text based on their reading cognitive process. Note that the reading habits of native speakers are varied and hard to modeled, so we adopt the three reading phases of second-language readers where there are potential behavior patterns. Such reading process of second-language readers generally includes *pre-reading*, *reading* and *post-reading* (Avery and Graves, 1997; Saricoban, 2002; Toprak and Almacıoğlu, 2009; Pressley and Afflerbach, 2012). In the *pre-reading* stage, they roughly preview the whole text to form an initial cognition and extract general but coarse-grained information at the meantime. Based on such prior knowledge, the subsequent *reading* stage is a conscious process that focuses on target-specific purposes to search fine-grained details through repeated skimming and scanning. For *post-reading*, re-reading is performed to check whether there are

*Corresponding author.

Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pages 3033–3043,
Hong Kong, China, November 3–7, 2019.

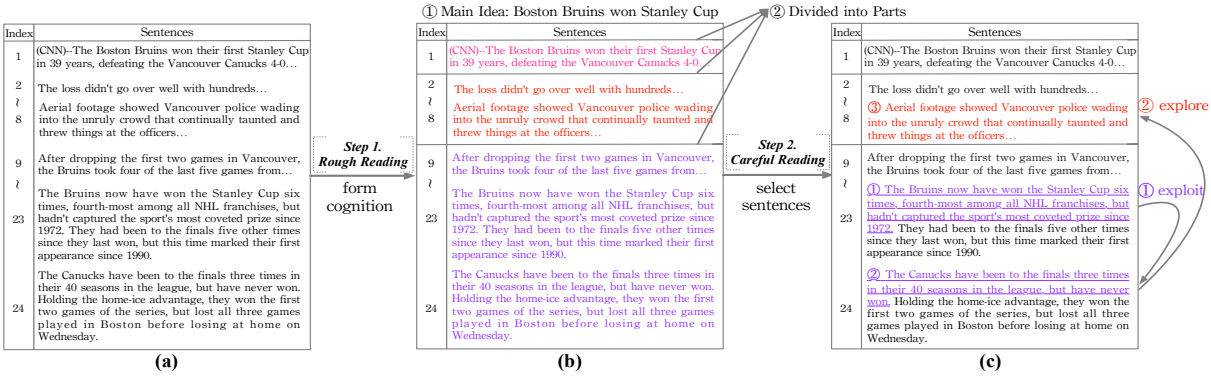

Figure 1: An example of how human beings extract summary. The article is from CNN/DailyMail dataset.

any missed details. The three-stage reading process makes it effective in capturing essential sentences of text as the extracted summarization.

Inspired by such human's reading cognitive process, in this paper, we re-examine the problem of extractive summarization and propose a new approach **HER** (**H**uman-b**E**ing-**R**eading inspired extractive summarization). We simplify the three-stage reading process to two subsequent stages called *rough reading* and *careful reading*. In *rough reading*, coarse-grained information of the original context is identified to form a general cognition. A detailed case is shown in Figure 1. In Figure 1 (a) and (b), after browsing on the whole article, the main idea is outlined and the text is roughly divided into three parts based on the gist of paragraphs at the meantime. Each part describes related but not the same contents. To implement the *rough reading* process, we use a hierarchical neural network to encode sentence vectors and derive a document representation as global feature for the main idea. Meanwhile, a convolutional neural network (CNN) is utilized to encode local features from different paragraphs.

During *careful reading*, the model searches for specific but important details through re-readings to cover the content and extract essential fine-grained information as the final summary. For instance, as shown in Figure 1 (c), after rough reading, two sentences close to the main idea "Boston Bruins won Stanley Cup" may be selected firstly. Then an earlier and more detailed sentence about "fans rioting" is appended to the summary by performing re-reading. It is a combination of people's *reading* and *post-reading* process. To accomplish this, we train a neural network to score each sentence. A multi-armed bandit policy with an adapted termination mechanism is then used to form the final summary.

In our HER model, the whole process is formulated as a contextual bandit problem. We train a reinforcement learning agent to solve it using the policy gradient method (Sutton et al., 2000). At each step, the agent takes an action which is a to-be-selected sentence set, and then receives a reward based on the correlation between extractive summary and gold-standard reference summary.

Our main contributions are as follows:

- We propose an extractive summarization method that simulates human being's reading cognitive process. We formulate it as a contextual bandit problem in which two stages including rough reading and careful reading are devised.
- We use a hierarchical neural network for rough reading which consists of a neural net to encode the whole document and another one to capture features in paragraphs. Then we use a contextual-bandit agent to flexibly select sentences during careful reading, with an adapted termination mechanism to select various but proper numbers of sentences.
- We conducted experiments on the CNN and DailyMail datasets and showed that our proposed model can outperform state-of-the-art methods and provide high-quality summaries.

2 The HER Model

In this section, we introduce the overall framework of our model, HER. We formulate extractive summarization as a contextual bandit (Langford and Zhang, 2007) trained using policy gradient reinforcement learning.

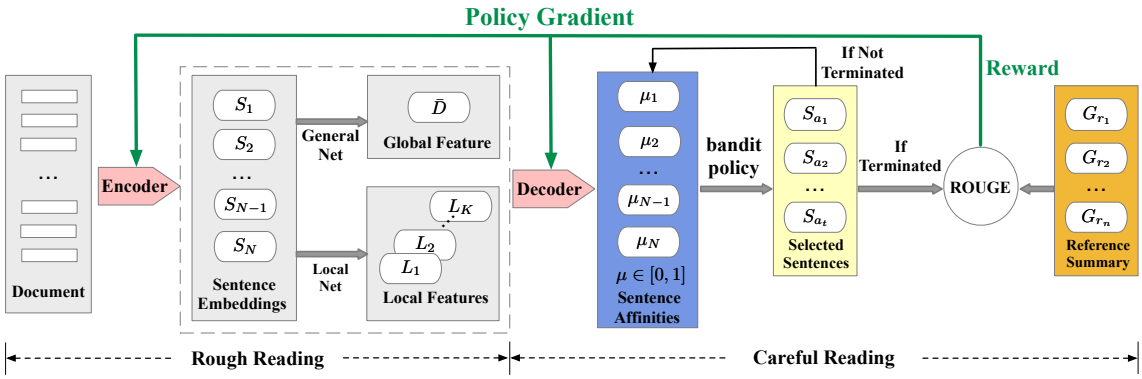

Figure 2: The overall framework of HER is formulated as a contextual bandit and can be divided into a two-stage process containing rough reading and careful reading.

As illustrated in Figure 2, the framework can be divided into two stages: rough reading and careful reading. During rough reading, a document with N sentences is encoded into sentence vectors $\{S_1, S_2, \ldots, S_N\}$ as well as a set of features denoted as F, which includes one global feature $\bar{S}$ representing the whole documentary information and K local features $\{L_1, L_2, \ldots, L_K\}$ depicting paragraphical contentsIn careful reading, sentence vectors are decoded into real-valued scores called sentence affinities $\{\mu_1, \mu_2, \ldots, \mu_N\}$, which can be considered as an estimation of sentence correlation to cover the context. Then a bandit policy is used to repeatedly choose unique sentence until the termination mechanism is triggered.

We will detail the preliminaries in Sec. 2.1, the rough reading stage in Sec. 2.2, and the careful reading stage in Sec. 2.3. The training process is illustrated in Sec. 2.4.

2.1 Summarization as Contextual-Bandit

Contextual-bandit (Langford and Zhang, 2008; Li et al., 2010; Pan et al., 2019a) is the multi-armed bandit (Auer et al., 2002) problem with featured contexts. At each step, the agent observes a context, selects an action based on the context, and then receives a reward. The agent's goal is to quickly find a decision making policy to maximize its return.

In extractive summarization, the goal of the task is to extract an undetermined number of sentences from the original document as summary. We show that it can be formulated as a contextual bandit problem if we select the sentences sequentially. Specifically, for each document, the *context* includes its documentary representation learned through rough reading. At each step $t = 1, 2, \ldots,$ we define the *action* as the index of the next sentence to select. The agent keeps selecting sentences until it reaches the termination condition (detailed in Sec. 2.3.3). Finally, the selected sentences SUM $= (S_{a_1}, \ldots, S_{a_M})$ corresponding to the selected actions $(a_1, \ldots, a_M)$ form an extractive summary. The agent will receive a reward $R(\text{SUM}; G)$. Note that G is the manually-labeled gold-standard summary of the document D, and the reward $R(\text{SUM}; G)$ measures the correlation between G and the predicted summary SUM.

2.2 Rough Reading

In rough reading, we aim to form a general cognition on a given document which encodes the document into sentence embeddings as well as produces the feature set F including global and local features.

Specifically, bidirectional LSTMs (BiLSTMs) on word- and sentence-level are first used to encode a document with N sentences into d_s-dimentional sentence embeddings $\{S_1, S_2, \ldots, S_N\}$, $S_i \in \mathbb{R}^{d_s}$. Second, a global feature $\bar{S} \in \mathbb{R}^{d_s}$ is computed as an average of all the sentence vectors. Third, we use a convolutional neural network to refine gist of different paragraphs and generate multiple local features on the sentence level, which is different from previous methods (Kim, 2014; Narayan et al., 2017; Yao et al., 2018) processing on the word level. In detail, a stacking of N sentence vectors is represented as,

$$S_{1:N} = [S_1, S_2, \cdots, S_N] \in \mathbb{R}^{N \times d_s}, \quad (1)$$

We apply 1-D convolutional neural networks on $S_{1:N}$ followed with a max-over-time pooling so that a final document-level representation can be

extracted. Specifically, we altogether used K convolutional nets with K different window sizes to summarize different gists of paragraphs. Finally, by stacking the outputs together, we get the final document-level representation for local features $L_{1:K} \in \mathbb{R}^{K \times d_s}$.

2.3 Careful Reading

Now that we have the sentence vectors and document-level features given by rough reading, we can perform careful reading to select the sentences one by one to form the summary.

2.3.1 Sentence Decoder

In order to extract high-quality summaries, we first compute the sentence affinities by a *sentence decoder*, which is observed effective in Dong et al. (2018). The sentence affinities are calculate by the following principles: (1) *Salience* (The sentences whose meanings are close to the central idea should be emphasized); (2) *Coverage* (The sentences that match different paragraphical information should be encouraged); (3) *Redundancy* (The unselected sentences which are similar to already extracted ones should be inhibited).

As we need to learn the relations between each sentence and the rest of the document, we update the sentence representations by,

$$S'_t = S_t \oplus \bar{S} \oplus L_{1:K}, t = 1, \ldots, N. \quad (2)$$

Then, we utilize a decoder Dec_1 to score the Salience and Coverage for each sentences, and a secondary score function Dec_2 to screen the sentences that might have Redundancy. Specifically,

$$\boldsymbol{\mu}_1 = (\mu_{11}, \ldots, \mu_{1N})^\top = \mathrm{Dec}_1(S'_{1:N}), \quad (3)$$
$$\boldsymbol{\mu}_2 = \mathrm{Dec}_2(S'_{1:N} \circ (1 - \boldsymbol{\mu}_1)), \quad (4)$$

We implement Dec_1 and Dec_2 as multi-layered perceptrons. Finally, we average the two scores as the final sentence affinities.

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2. \quad (5)$$

2.3.2 Bandit Policy

The overall decision making process goes as follows. The agent selects one sentence at each step based on the contextual information provided by the rough reading and sentence affinities computed by the sentence decoder. It stops taking actions once the termination mechanism is triggered. After that, all the selected sentences are formed as a summary, and a final reward can be calculated by comparing to the labeled golden summary. As there is no intermediate reward before termination, the goal of the agent is to find a policy to maximize its expected long-term return.

It is an intuitive choice to select sentences with the highest affinities as summary, which is similar to training data selection (Song et al., 2012; Liu et al., 2019). However, such an *argmax* policy is prone to only learn the easy patterns since it lacks exploration. We will show an example in Section 4.3. They should be explored to form the summary as well. Since the search space for summarization is extremely large, we must explicitly address the tradeoff between exploration and exploitation for fast learning, which is an active research area in reinforcement learning and applications (Pan et al., 2019a,b). In our work, we find the use of ϵ-greedy with stochastic policy works well enough to encourage exploration. Specifically, with a probability of $1 - \epsilon$, the agent chooses the sentence following the current policy, i.e., to sample an index $a_t \in [1, N]$ from the multinomial distribution with sentence affinities $\{\mu_1, \mu_2, \ldots, \mu_N\}$ as probabilities. Otherwise, with a small probability of ϵ, the agent randomly picks one available sentence as an exploration. Note that such exploration is only used during training.

2.3.3 Termination Mechanism

In HER, we propose a termination mechanism that is independent on future rewards to make our model flexible in extracting summary with various numbers of sentences. This mechanism is defined as follows

$$\text{Done} \sim \text{Bernoulli}(\min(\frac{\mu_{\min}}{\mu_{\max}}, 1 - \mu_{\max})), \quad (6)$$

where $\mu_{\max}$ and $\mu_{\min}$ denotes the maximal and minimal value in $\boldsymbol{\mu}$ for all the remained sentences, respectively. Thus Done $= 1$ terminates the sentence extraction when there is no key sentences left. With this mechanism, the agent will stop extraction with high probability as long as the differences among remaining affinities are small enough or the remaining sentence affinities are very low.

2.4 Training

After the agent sequentially takes an action a_t until terminated, we can derive an summary induced by a out of a document D. Then the agent would receive a reward $R(\text{SUM}; G)$ where G is the gold-standard summary of D. $R(\text{SUM}; G)$ is computed

by the average of three variants of ROUGE (Lin, 2004). To balance precision and recall, we use F-score here,

$$R(\text{SUM}; G) = \frac{1}{3}(\text{ROUGE-1}_f(\text{SUM}; G) + \text{ROUGE-2}_f(\text{SUM}; G) + \text{ROUGE-L}_f(\text{SUM}; G)). \quad (7)$$

We represent the whole extractive neural network as $p_\theta(\cdot|D)$ containing the encoder in rough reading and the decoder in careful reading. The goal of our model is to find parameters θ of p_θ to produce high-quality summary and maximize the rewards (c.f. Eq. (8)). But we cannot obtain gradient to maximize Eq. (8) with gradient ascent as it is discretely sampled. So we use the likelihood ratio gradient estimator, also known as REINFORCE (Williams, 1992; Sutton et al., 2000), to acquire the gradient by Eq. (9).

We use $Q(D)$ in Eq. (10) to construct $p_\theta(a|D)$ following Dong et al. (2018), where $z(D) = \sum_t \mu_t(D)$ and ϵ is the exploration probability of the ϵ-greedy denoted in Sec. 2.3.2. M is the number of extracted sentences this is determined jointly by the termination mechanism and the document context. $Q(D)^{\frac{1}{M}}$ is adopted to present $p_\theta(a|D)$ to avoid extracting fewer or more sentences when maximizing the objective function. Hence, $\bigtriangledown_\theta \text{log} p_\theta(a|D)$ in Eq. (9) can be easily computed.

$$J(\theta) = E[R(\text{SUM}; G)] \quad (8)$$

$$\bigtriangledown_\theta J(\theta) = E[\bigtriangledown_\theta \text{log} p_\theta(a|D) R(\text{SUM}; G)] \quad (9)$$

$$Q(D) = \prod_{j=1}^{M} \left(\frac{\epsilon}{N-j+1} + \frac{(1-\epsilon)\mu_{a_j}(D)}{z(D) - \sum_{k=1}^{j-1} \mu_{a_k}(D)}\right) \quad (10)$$

However, the exact document distribution is unknown and we cannot evaluate the expected value in Eq. (9). So we use sampling to estimate it instead. Given a document-summary pair (D, G), our HER samples B summaries induced by $a^1, \ldots, a^B$ from $p_\theta(\cdot|D)$ and obtain all the gradients, then the average value is considered as the estimation. As sample-based gradient estimate may have high variance, we use a baseline for variance reduction. The gradient of the objective function is finally represented as,

$$\bigtriangledown_\theta J(\theta) \approx \frac{1}{B} \sum_{b=1}^{B} \bigtriangledown_\theta \text{log} p_\theta(a^b|D)(R(a^b, G) - \bar{r}) \quad (11)$$

where we choose self-critical reinforcement learning to acquire the baseline $\bar{r}$ following Ranzato et al. (2015); Rennie et al. (2017); Paulus et al. (2017); Dong et al. (2018) computed by greedy encoding $\bar{r} = R(a_{greedy}; G)$. More concretely, $a_{greedy} = \text{argmax} p_\theta(a|D)$ and this baseline satisfies that the probability of a sampled sequence would be increased when the summary it induces is better than what is obtained by greedy decoding. The procedure of HER is shown in Algorithm 1.

Algorithm 1 HER: training

Input $\{(D_i, G_i)\}$, a dataset of document-summary pairs
Output θ, the updated parameters in p_θ

for each document-summary pair (D, G) **do**
  $S_{1:N}, \bar{S}, L_{1:K} = \text{Encoder}(D)$ ▷ Rough Reading
  $\mu_{1:N} = \text{Decoder}(S_{1:N}, \bar{S}, L_{1:K})$ ▷ Careful Reading
  for each trail $b = 1, \ldots, B$ **do**
    Initialize the action set $A = \{1, \ldots, N\}$
    for time step $t = 1, \ldots, N$ **do**
      Sample $u \sim U(0, 1)$
      if $u < \epsilon$ **then**
        Uniformly sample a_t from A
      else
        $a_t \sim \text{Categorical}(\boldsymbol{\mu}_A)$
      Get termination flag by Eq.(6)
      Let $M = t$
      if Done **then**
        Break the inner loop
      else
        Update the action set $A = A \setminus \{a_t\}$
    Generate summary $\text{SUM}_b = (S_{a_1}, \ldots, S_{a_M})$
    $l_b = \frac{1}{M} \sum_{t=1}^{M} \text{log}(\frac{\epsilon}{N-t+1} + \frac{(1-\epsilon)\mu_{a_t}}{\sum_j \mu_j - \sum_{k=1}^{t-1} \mu_{a_k}})$
    Compute reward $r_b = R(\text{SUM}_b; G)$
  $\bar{r} = R(\text{SUM}_{greedy}; G)$
  $l = \frac{1}{B} \sum_{b=1}^{B} l_b(\bar{r} - r_b)$ ▷ Surrogate loss
  $\theta \leftarrow \theta - \lambda \bigtriangledown_\theta l$ ▷ Update

3 Experiment Settings

In this section we present our experimental setup for evaluating the performance of the proposed HER, including the datasets, evaluation protocol, baselines and implementation details.

Datasets: We evaluate our models on three datasets: the CNN, the DailyMail and the combined CNN/DailyMail (Hermann et al., 2015; Nallapati et al., 2016). We also use the standard splits of Hermann et al. (2015) for training, validation, and test ($90,266/1,220/1,093$ documents for CNN and $196,961/12,148/10,397$ for DailyMail) with the same setting in See et al. (2017).

Evaluation: We evaluate summarization quality using F_1 ROUGE (Lin, 2004) including unigram and bigram overlap (ROUGE-1 and ROUGE-2) to assess informativeness and the longest common subsequence (ROUGE-L) to assess fluency with the reference summaries. We obtain ROUGE scores using a faster python im-

plementation[1] for training and evaluation, and the standard pyrouge package[2] for test following Dong et al. (2018).

Baselines: We compare our proposed HER against four kinds of extractive methods: (1) Lead-3 model simply selects the first three sentences. (2) NN-SE (Cheng and Lapata, 2016) and SummaRuNNer (Nallapati et al., 2017) are sequence labeling task and trained with cross-entropy loss. (3) Refresh (Narayan et al., 2018), DQN (Yao et al., 2018) and RNES (Wu and Hu, 2018) extract summary via reinforcement learning. (4) BANDITSUM (Dong et al., 2018) considers the extractive summarization as a contextual bandit but fails to simulate human reading recognition process.

Implementation Details: We initialize word embeddings with 100-dimension Glove embeddings (Pennington et al., 2014). In rough reading, the encoder is hierarchical and each layer is a two-stacked BiLSTM with a hidden size of 200. Therefore, sentence vectors and the document representation $\bar{S}$ have a dimension of 400. For the variant CNN, we adopt filter windows H in $\{1, 2, 3\}$ with 100 feature maps each and generate $K = 3$ local representations for each document. In careful reading, we set $\epsilon = 0.1$ for bandit policy. We also bound the minimum and maximum number of selected sentence to be 1 and 10 for termination mechanism. During training, we use the optimizer Adam (Kingma and Ba, 2014) with a learning rate of 10^{-5}, beta parameters as $(0, 0.999)$ and a weight decay of 10^{-6} to maximize the objective function following Dong et al. (2018). We employ gradient clipping of 1 for regularization and sample $B = 20$ times for each document. We train our model within two epochs. Note that we choose the whole document as the final summary if the document length is less than 3 sentences as the local features cannot be obtained through the CNN-based network.

4 Experimental Results

4.1 Quantitative Analysis

We first report the ROUGE metrics on the combined CNN/DailyMail test sets in Table 1 and the separate results in Table 2. We can get several observations from these two tables.

[1] https://github.com/pltrdy/rouge

[2] Pyrouge is a Python package. We compute all ROUGE scores with parameters "-a -c 95 -m -n 4 -w 1.2." Refer to https://pypi.python.org/pypi/pyrouge/0.1.3

Model	ROUGE		
	R1	R2	RL
Lead-3	40.0	17.5	36.2
SummaRuNNer	39.6	16.2	35.3
DQN	39.4	16.1	35.6
Refresh	40.0	18.2	36.6
RNES	41.3	**18.9**	37.6
BANDITSUM	41.5	18.7	37.6
HER	**42.3**	**18.9**	**37.9**

Table 1: Results on the combined CNN/DailyMail test set. We report F1 scores of ROUGE-1 (R1), ROUGE-2 (R2), and ROUGE-L (RL). The result of Lead-3 is taken from Dong et al. (2018).

Model	CNN			DailyMail		
	R1	R2	RL	R1	R2	RL
Lead-3	28.8	11.0	25.5	41.2	18.2	37.3
NN-SE	28.4	10.0	25.0	36.2	15.2	32.9
Refresh	30.4	11.7	26.9	41.0	18.8	37.7
BANDITSUM	**30.7**	**11.6**	27.4	42.1	18.9	38.3
HER	**30.7**	11.5	**27.5**	**42.7**	**19.0**	**38.5**

Table 2: Results of the test sets on the CNN and DailyMail datasets separately.

Firstly, our model generally performs the best and even surpasses 42 on ROUGE-1 score on the combined CNN/DailyMail dataset. It also shows better results on the separate datasets. We argue that global and local features from rough reading can help extract summaries by capturing deep contextual relations, and the designed structure in careful reading makes it more flexible in selecting sentence sets. Hence a two-stage framework based on the human's reading cognition is more appropriate for extractive summarization.

Secondly, directly optimizing the evaluation metric by combining cross-entropy loss with rewards may improve the extractive results. RL-based methods, Refresh (Narayan et al., 2018) and RNES (Wu and Hu, 2018), perform better than the sequence labeling methods like SummaRuNNer (Nallapati et al., 2017). BANDITSUM (Dong et al., 2018) generally performs better than the other baselines, and it reports that framing the extractive summarization based on contextual bandit is more suitable than sequential labeling setting and also has more search space than other RL-based methods (Narayan et al., 2018; Yao et al., 2018; Wu and Hu, 2018).

4.2 Ablation Test

Next, we conduct ablation test by removing the modules of the proposed HER step by step. Firstly, we replace the automatic termination mechanism with a fixed extracting strategy that always selects

Model	ROUGE		
	R1	R2	RL
HER	**42.3**	**18.9**	**37.9**
HER-3	42.0	18.5	37.6
HER-3 w/o policy	41.7	18.3	37.1
HER-3 w/o policy&L	41.2	18.4	37.0
HER-3 w/o policy&F	40.6	18.2	36.9

Table 3: The results of ablation test on the test split of the combined CNN/DailyMail dataset. L and F are short for local net and rough reading.

three sentences for every document and we present the model as HER-3. Based on HER-3, we also remove bandit policy, local net, general net gradually, and denote them as HER-3 w/o policy, HER-3 w/o policy & local net and HER-3 w/o policy & rough reading individually. The results are reported in Table 3 and it proves the effectiveness of each proposed module. Firstly, HER constructed with an automatic termination mechanism is more flexible and reliable in extracting various numbers of sentences varying different documents. Secondly, HER use ϵ-greedy to select sentences in order to raise the exploration chances on discovering important but easily ignored information. Thirdly, general cognition from rough reading process is useful in extractive summarizarion.

4.3 A Closer Look

To verify whether our proposed HER can simulate human beings' reading cognitive process, and whether such simulation are inherently helpful on extractive summarization, we conduct extensive evaluations and try to answer the following three questions.

(1) Can CNN-based network extract local features of different paragraphs?

In Figure 3, we report the distribution of selected sentences' positions on our proposed model HER, BANDITSUM and HER w/o Local Net. We test each model at 10k, 50k, 100k training steps. It shows that all the three models can focus on different parts of the context to form summary at first and BANDITSUM performs the best after training 10k steps. However, with training steps growing, BANDITSUM and HER w/o Local begin to prefer earlier sentences. HER, on the other hand, can focus on various paragraphs and extract information from different parts of the texts with constant training. The contextual bandit (CB) based frameworks seems to be able to attend on various parts of the contexts to some degree in the beginning.

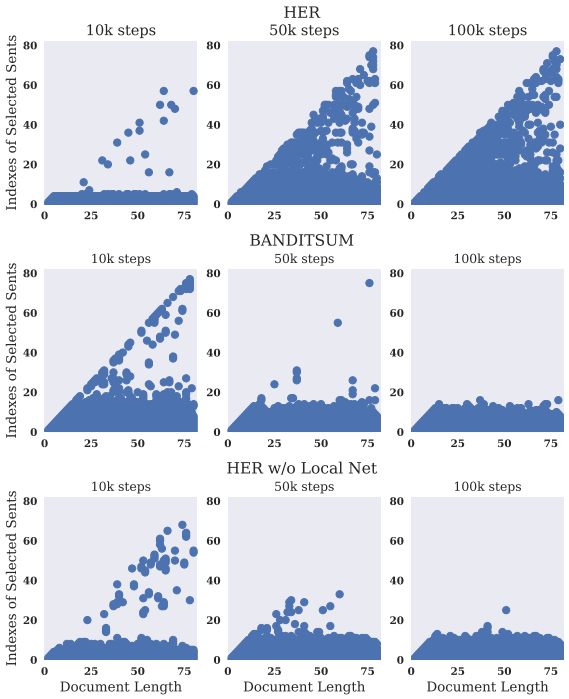

Figure 3: The statistics of model HER, BANDITSUM (Dong et al., 2018), HER w/o Local Net on the selected sentences' indexes varying different document lengths. This is reported on the documents the lengths of which are all less than 80 on the test split.

However, with constant training, both BANDITSUM and HER w/o Local start to focus on earlier sentences due to the nature that sentences similar to the main idea usually lie on the head of the text. As our proposed HER is equipped with a variant CNN to extract local features, our model can focus on gist of paragraphs rather than only the first several sentences. It also encourages the exploration on extracting information from various positions more flexibly.

(2) Can the proposed bandit policy discover low-score but easily ignored information?

To answer this question, we demonstrate a detailed case on sentence selection in Figure 4. We observe that although the fourth sentence has a high affinity, it should not be included in the summary since its meaning is close to the third sentence which has already been extracted. Instead, the 13th sentence is supposed to be selected though it has low affinity. Since our HER adopts the ϵ-greedy policy, it can explore such sentence and extract it out correctly.

(3) Can HER extract varied but proper numbers of sentences?

We answer this question by drawing the fre-

Index	Sentence	Affinity	HER	HER w/o policy
2	Two spotted leopards, two Macaque monkeys and a brown bear will be returned to Marian Thompson…	0.873	yes	yes
3	He set off a wide scare in October when he released 50 potentially dangerous animals from his farm before shooting himself.	0.872	yes	yes
4	Of the 50 animals Thompson released, 48 were killed by law enforcement, while two primates were killed by the other animals, zoo officials said.	0.767	**no**	**yes**
13	State officials have no legal power to inspect the cages before the animals are returned…	0.297	**yes**	**no**

Figure 4: A case on sentence selection of HER and HER w/o policy. The article is from CNN dataset. The highlighted indices indicate the corresponding sentences which should be extracted as summary.

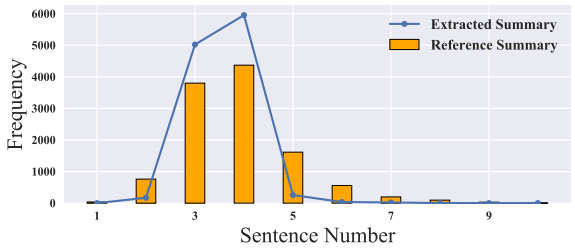

Figure 5: The statistics on extracted sentence number of our model. Frequency is the number of documents.

quency distribution of extracted sentence numbers by our model on the test set of combined CNN/DailyMail, and Figure 5 exhibits the results. We observe that the frequency distribution of extracted sentence number is basically similar to that of the gold-standard summary. Compared with BANDITSUM which extracts fixed number of sentences, our model shows more flexibility and extensibility on extractive summarization.

4.4 Human Evaluation

Lastly, we conduct a qualitative evaluation. Following Wu and Hu (2018), we randomly sample 50 documents from the test set on the combined CNN/DailyMail dataset and ask three volunteers to evaluate the summaries extracted by HER w/o Dec_2, HER w/o Local Net, BANDITSUM and HER, respectively. HER w/o Dec_2 only uses Eq. (3) to compute sentence affinities without inhibiting redundant sentences. For each document-summary pair, they are asked to rank the output of each system on three aspects, namely overall quality, coverage and non-redundancy. Notice that the best one will be marked rank 1 and so on, and two system would be ranked the same if their extracted summaries are identical. We report the average results in Table 4 and it shows that our HER is leading than BANDITSUM on overall quality and coverage. Additionally, HER w/o Dec_2 performs the worst on non-redundancy as it does not specialize these unselected sentences which are similar to already extracted ones. Furthermore, HER w/o Local Net takes on bad performance on coverage because the local features can focus on paragraphical messages and help to refine thorough information.

Model	Overall	Coverage	Non-Redundancy
HER w/o Dec_2	2.88	2.81	2.81
HER w/o L	3.02	2.96	2.75
BANDITSUM	2.06	2.07	**1.91**
HER	**1.81**	**1.75**	1.97

Table 4: Average rank of human evaluation in terms of overall performance, coverage, and non-redundancy. L is short for local Net. Lower score is better.

5 Related Work

Extractive Text Summarization Researchers have developed many statistical methods for automatic extractive summarization. Traditional methods learn to score each sentence dependently (Erkan and Radev, 2004; Mihalcea and Tarau, 2004; Wong et al., 2008). Recently neural network based extractive methods (Cheng and Lapata, 2016; Nallapati et al., 2017; Feng et al., 2018; Shi et al., 2018) usually consider extractive summarization as sequence labeling tasks and aim to minimize the cross-entropy objective function. Narayan et al. (2017) utilizes side information to help sentence classifier while Yasunaga et al. (2017) computes the salience of each sentence for selection with graph convolutional networks. In addition, reinforcement learning based methods (Wu and Hu, 2018; Narayan et al., 2018; Yao et al., 2018) have been proposed to directly optimize the evaluation metric ROUGE by combining cross-entropy loss with rewards from policy gradient reinforcement learning. Dong et al. (2018) considered extractive summarization as a contextual bandit and it performs well especially when good summary sentences appear late in the source document. Recently, Nallapati et al. (2017); Chen and Bansal (2018); Hsu et al. (2018) propose unified models and combine the advantages of both extractive and abstractive methods.

Human Reading-inspired Strategy in NLP Recently, several pioneer researches began to study how to adapt human reading cognition process, usually including pre-reading, reading and post-reading (Avery and Graves, 1997; Saricoban,

2002; Toprak and Almacıoğlu, 2009; Pressley and Afflerbach, 2012), into various NLP-related applications. For example, Li et al. (2018) solved document-based question answering and by simulating human being's reading strategy. Luo et al. (2018, 2019) utilized the prior knowledge of human reading to solve sub-tasks in sentiment analysis. Song et al. (2017, 2018) enhanced word embeddings in a similar way. Yang et al. (2019) applied it for abstractive summarization, Zheng et al. (2019) simulated human behavior for reading comprehension, and Lei et al. (2019) utilized human-like semantic cognition for aspect-level sentiment classification. In this paper, we attempt to perform extractive summarization under the inspiration of human reading recognition.

6 Conclusion

Inspired by the reading cognition of human beings, we propose HER, a two-stage method, to mimic how people extract summaries. The whole learning process is formulated as a contextual bandit trained with policy gradient reinforcement learning. In rough reading, two neural networks are taken to encode coarse-grained information. In careful reading, repeatedly reading are conducted to select fine-grained sentences as summary. Experiments on two real-world datasets demonstrate that our proposed model can significantly outperform the state-of-the-art extractive methods on summary quality, coverage and non-redundancy.

Acknowledgements

This work is partially supported by the National Key Research and Development Program of China under Grant No. 2017YFB1002104, the National Natural Science Foundation of China under Grant No. 91846113, U1811461, 61602438, 61573335, CCF-Tencent RhinoBird Young Faculty Open Research Fund No. RAGR20180111. This work is also funded in part by Ant Financial through the Ant Financial Science Funds for Security Research. Xiang Ao is also supported by Youth Innovation Promotion Association CAS. This work was also partially supported by SIAT Innovation Program for Excellent Young Researchers (Grant No. Y8G027). Min Yang was sponsored by CCF-Tencent Open Research Fund.

References

Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. 2002. The nonstochastic multi-armed bandit problem. *SIAM journal on computing*.

Patricia G Avery and Michael F Graves. 1997. Scaffolding young learners' reading of social studies texts. *Social Studies and the Young Learner*, 9(4):10–14.

Yen-Chun Chen and Mohit Bansal. 2018. Fast abstractive summarization with reinforce-selected sentence rewriting. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 675–686.

Jianpeng Cheng and Mirella Lapata. 2016. Neural summarization by extracting sentences and words. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*.

Yue Dong, Yikang Shen, Eric Crawford, Herke van Hoof, and Jackie Chi Kit Cheung. 2018. Banditsum: Extractive summarization as a contextual bandit. In *EMNLP*.

Günes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, 22:457–479.

Chong Feng, Fei Cai, Honghui Chen, and Maarten de Rijke. 2018. Attentive encoder-based extractive text summarization. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM.

Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In *Advances in neural information processing systems*.

Wan-Ting Hsu, Chieh-Kai Lin, Ming-Ying Lee, Kerui Min, Jing Tang, and Min Sun. 2018. A unified model for extractive and abstractive summarization using inconsistency loss. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *EMNLP*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization.

John Langford and Tong Zhang. 2007. The epoch-greedy algorithm for contextual multi-armed bandits. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, pages 817–824. Citeseer.

John Langford and Tong Zhang. 2008. The epoch-greedy algorithm for multi-armed bandits with side information. In *Advances in neural information processing systems*, pages 817–824.

Zeyang Lei, Yujiu Yang, Min Yang, Wei Zhao, Jun Guo, and Yi Liu. 2019. A human-like semantic cognition network for aspect-level sentiment classification. In *AAAI*.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670. ACM.

Weikang Li, Wei Li, and Yunfang Wu. 2018. A unified model for document-based question answering based on human-like reading strategy. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. *Text Summarization Branches Out*.

Miaofeng Liu, Yan Song, Hongbin Zou, and Tong Zhang. 2019. Reinforced training data selection for domain adaptation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1957–1968, Florence, Italy.

Ling Luo, Xiang Ao, Feiyang Pan, Jin Wang, Tong Zhao, Ningzi Yu, and Qing He. 2018. Beyond polarity: Interpretable financial sentiment analysis with hierarchical query-driven attention. In *IJCAI*, pages 4244–4250.

Ling Luo, Xiang Ao, Yan Song, Jinyao Li, Xiaopeng Yang, Qing He, and Dong Yu. 2019. Unsupervised neural aspect extraction with sememes. In *IJCAI*.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*.

Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *AAAI*.

Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Caglar Gulcehre, and Bing Xiang. 2016. Abstractive text summarization using sequence-to-sequence rnns and beyond. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*.

Shashi Narayan, Shay B Cohen, and Mirella Lapata. 2018. Ranking sentences for extractive summarization with reinforcement learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Shashi Narayan, Nikos Papasarantopoulos, Shay B Cohen, and Mirella Lapata. 2017. Neural extractive summarization with side information. *arXiv preprint arXiv:1704.04530*.

Ani Nenkova, Kathleen McKeown, et al. 2011. Automatic summarization. *Foundations and Trends® in Information Retrieval*, 5(2–3):103–233.

Feiyang Pan, Qingpeng Cai, Pingzhong Tang, Fuzhen Zhuang, and Qing He. 2019a. Policy gradients for contextual recommendations. In *The World Wide Web Conference*, pages 1421–1431. ACM.

Feiyang Pan, Qingpeng Cai, An-Xiang Zeng, Chun-Xiang Pan, Qing Da, Hualin He, Qing He, and Pingzhong Tang. 2019b. Policy optimization with model-based explorations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4675–4682.

Romain Paulus, Caiming Xiong, and Richard Socher. 2017. A deep reinforced model for abstractive summarization. *arXiv preprint arXiv:1705.04304*.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *EMNLP*.

Michael Pressley and Peter Afflerbach. 2012. *Verbal protocols of reading: The nature of constructively responsive reading*. Routledge.

Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. 2015. Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732*.

Steven J Rennie, Etienne Marcheret, Youssef Mroueh, Jerret Ross, and Vaibhava Goel. 2017. Self-critical sequence training for image captioning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Arif Saricoban. 2002. Reading strategies of successful readers through the three phase approach. *The Reading Matrix*, 2(3).

Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083.

Jiaxin Shi, Chen Liang, Lei Hou, Juanzi Li, Zhiyuan Liu, and Hanwang Zhang. 2018. Deepchannel: Salience estimation by contrastive learning for extractive document summarization. *arXiv preprint arXiv:1811.02394*.

Yan Song, Prescott Klassen, Fei Xia, and Chunyu Kit. 2012. Entropy-based training data selection for domain adaptation. In *Proceedings of COLING 2012*, pages 1191–1200, Mumbai, India.

Yan Song, Chia-Jung Lee, and Fei Xia. 2017. Learning word representations with regularization from prior knowledge. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 143–152, Vancouver, Canada.

Yan Song, Shuming Shi, and Jing Li. 2018. Joint learning embeddings for chinese words and their components via ladder structured networks. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 4375–4381.

Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. 2000. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*.

Elif Leyla Toprak and Gamze Almacıoğlu. 2009. Three reading phases and their applications in the teaching of english as a foreign language in reading classes with young learners. *Journal of language and Linguistic Studies*, 5(1):20–36.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256.

Kam-Fai Wong, Mingli Wu, and Wenjie Li. 2008. Extractive summarization using supervised and semi-supervised learning. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 985–992. Association for Computational Linguistics.

Yuxiang Wu and Baotian Hu. 2018. Learning to extract coherent summary via deep reinforcement learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Min Yang, Qiang Qu, Wenting Tu, Ying Shen, Zhou Zhao, and Xiaojun Chen. 2019. Exploring human-like reading strategy for abstractive text summarization. In *AAAI*.

Kaichun Yao, Libo Zhang, Tiejian Luo, and Yanjun Wu. 2018. Deep reinforcement learning for extractive document summarization. *Neurocomputing*.

Michihiro Yasunaga, Rui Zhang, Kshitijh Meelu, Ayush Pareek, Krishnan Srinivasan, and Dragomir Radev. 2017. Graph-based neural multi-document summarization. In *Proceedings of the 21st Conference on Computational Natural Language Learning*.

Yukun Zheng, Jiaxin Mao, Yiqun Liu, Zixin Ye, Min Zhang, and Shaoping Ma. 2019. Human behavior inspired machine reading comprehension. In *SIGIR*.