Automatic Text Revision with Application to Legal Documents

Yiquan Wu
Zhejiang University
Hangzhou, Zhejiang, China
wuyiquan@zju.edu.cn

Kun Kuang
Zhejiang University
Hangzhou, Zhejiang, China
kunkuang@zju.edu.cn

Fei Wu
Zhejiang University
Hangzhou, Zhejiang, China
wufei@cs.zju.edu.cn

ABSTRACT

Most existing natural language generation (NLG) methods focus on the one-time text generation while ignoring the text revision process, which is also vital in many real applications. For example, the court's view in legal documents always needs to revise before being public. To learn the latent patterns behind the revision process, in this paper, we focus on the problem of text revision, with application to legal documents, saying the court's view revision. Firstly, we construct a dataset to support the problem with input as court's view draft and corresponding fact description, and output as the revised (whole) court's view. Then, we employ a preliminary model to automatically revise the court's view. Preliminary experimental results show the feasibility and effectiveness of the preliminary model on this task, and we will continue to put forward this work in the future.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

Datasets, Natural Language Generation, Neural Networks

1 INTRODUCTION

Owing to the prosperity of deep learning, many natural language processing (NLP) techniques have been employed in different aspects, such as text classification[12, 21], sentiment analysis[15, 18], machine translation[6, 25], etc. Natural language generation (NLG) is one of the most important branches in NLP[9]. However, existing NLG methods mainly focus on the problem of text generation, ignoring the process of text revision. Text revision is very important and necessary in many real applications. For example, news would be revised several times by journalist and editor before being reported, and a legal document (i.e., court's view) also need to be carefully revised many times by the judges before being announced in court. How to develop automatic methods for text revision is still an open problem, and is of paramount importance for both academic research and real applications.

In this paper, we focus on text revision with application to legal documents. In legal documents, court's view is an important portion of the verdict to interpret the final sentence or judgment of a legal case. Hence, a revision is always needed for court's view being announcing in court, even for the senior judge. To explore the latent patterns of revision, we focus on the real legal application of court's view revision with inputting a court's view draft and corresponding fact description, and outputting the revised court's view for publication.

To the best of our knowledge, we are the first to investigate the problem of automatic text revision. Traditional public datasets, especially for the legal datasets in NLP, can hardly be directly applied for the problem of text revision. Here, according to the public legal documents with court's view and fact description, we adopt a novel method to automatically construct a simulated court's view draft dataset, where the most factual informative sentence is masked based on the fact description. Specifically, we first use BERTScore[29] to calculate the similarity of two sentences from the fact description and the court's view. Then we mask the most factual informative sentence in the court's view, according to the similarity score. The remaining part of the court's view is the draft we construct.

With the constructed dataset, we employ a sequence-to-sequence [22] model as the preliminary model for automatic court's view revision. With extensive experiments, we validate the feasibility and effectiveness of our preliminary model. As text revision plays a central role during text generation in real applications, we will future investigate it and develop an effective method for better performance and evaluate the method combined with other generative methods on other datasets.

The main contributions of this paper can be summarized as follows:

- To the best of our knowledge, we are the first to investigate the problem of automatic text revision, which is of paramount importance for both academic research and real applications.
- We explore text revision problem with application to legal documents. To be specific, we focus on revising the court's view draft depending on the fact description.
- We adopt a novel method to automatically construct a court's view draft dataset for revision. This method can be easily applied to generate drafts in other NLG datasets.
- Preliminary experimental results show the feasibility and effectiveness of our preliminary model for the revision of court's view.

LegalAI'20@SIGIR'20, July 30, 2020, Virtual Event, China

ACM ISBN TBA…$15.00
https://doi.org/TBA

FACT DESCRIPTION	经审理查明，二被告系夫妻关系。被告甲因做工程需要，于2009年4月至2010年3月共计向原告借款400万元。原告通过银行汇给被告乙262万元，其它款项现金支付，被告甲于2010年3月10日，出具借条一份，载明"今借到丙人民币肆佰万元正，还款日期2010年12月10日"。该款到期后，经原告多次被告催要未果，故原告诉至本院，提出前列诉讼请求。上述事实，有原告陈述及原告提供借条、汇款凭证在案为证。 After the hearing, it is found that the two defendants are husband and wife. Defendant A borrowed 4 million yuan from the plaintiff from April 2009 to March 2010 due to the need of the project. The plaintiff remitted 2.62 million yuan to Defendant B through the bank, and other money were paid in cash. Defendant A issued an debt acknowledgement on March 10, 2010, stating that "the loan was made to four million yuan and the repayment date was December 10, 2010". After the expiration, the plaintiff has repeatedly urged the defendant but with no success, therefore the plaintiff appealed to the court and filed the preceding lawsuit. The above facts are evidenced by the plaintiff's statement and the receipt provided by the plaintiff.
COURT'S VIEW DRAFT	本院认为，原告与被告甲之间的借贷关系不违反我国法律、行政法规的强制规定，应认定合法有效。被告未履行还款义务，原告要求其归还借款的诉讼请求成立，应依法予以支持。关于利息，双方在借款时约定还款期限，但未约定利息，视为不计息，其利息从借款期满后次日起，按中国人民银行同期贷款利率计算。被告甲、乙经本院合法传唤既未到庭参加诉讼，亦未向本院提供相关证据，应承担举证不能的法律后果。 The court holds that the loan relationship between the plaintiff and the defendant A does not violate the mandatory provisions of Chinese laws and administrative regulations, and should be recognized as legitimate and valid. The defendant fails to perform the repayment obligation, so the plaintiff's claim for repayment of the loan is established and shall be supported according to law. As for the interest, when both parties agree on the repayment period, no interest is agreed, so it shall be deemed as non-interest-bearing, and the interest shall be calculated according to the loan interest rate of the People's Bank of China for the corresponding period from the next day after the expiration of the loan. A and B are summoned by the court but neither attend the proceedings nor provide relevant evidence to the court, they shall bear the legal consequences of failing to provide evidence.
COURT'S VIEW	本院认为，原告与被告甲之间的借贷关系不违反我国法律、行政法规的强制规定，应认定合法有效。被告未履行还款义务，原告要求其归还借款的诉讼请求成立，应依法予以支持。关于利息，双方在借款时约定还款期限，但未约定利息，视为不计息，其利息从借款期满后次日起，按中国人民银行同期贷款利率计算。被告甲与被告乙系夫妻关系，在夫妻关系存续期间甲所欠的债务，应当按夫妻共同债务处理。被告甲、乙经本院合法传唤既未到庭参加诉讼，亦未向本院提供相关证据，应承担举证不能的法律后果。 The court holds that the loan relationship between the plaintiff and the defendant A does not violate the mandatory provisions of Chinese laws and administrative regulations, and should be recognized as legitimate and valid. The defendant fails to perform the repayment obligation, so the plaintiff's claim for repayment of the loan is established and shall be supported according to law. As for the interest, when both parties agree on the repayment period, no interest is agreed, so it shall be deemed as non-interest-bearing, and the interest shall be calculated according to the loan interest rate of the People's Bank of China for the corresponding period from the next day after the expiration of the loan. Defendant A and Defendant B are husband and wife, the debts owed by A during the existence of matrimonial relationship shall be treated as joint debts of husband and wife. A and B are summoned by the court but neither attend the proceedings nor provide relevant evidence to the court, they shall bear the legal consequences of failing to provide evidence.

Figure 1: An example of fact description, simulated court's view draft and revised court's view. The red part is the most factual informative sentence.

2 RELATED WORK

2.1 Legal Artificial Intelligence

Legal Artificial Intelligence (LegalAI) focuses on applying the technology of artificial intelligence, especially natural language processing, aiming at helping lawyers and lower court judges. In recent years, LegalAI has drawn increasing attention rapidly from both law and AI fields. While many researchers from law usually try to solve tasks through rule-based and symbol-based methods, AI researchers concentrate more on data-driven and embedding methods[32].

Since most of the legal documents appear in textual form, many NLP technologies have been brought into the legal field to improve the efficiency of legal work. Charge prediction is a common task of judgment prediction, considered as a classification problem [3, 10, 11, 13, 30]. Besides, there are also works on law articles recommendation [5], legal questions classification [28], controversy focus mining [8] and relevant case retrieval [4].

In order to promote the LegalAI, some new LegalAI datasets have been proposed, such as CAIL2018[26], Cail2019-scm[27], Cjrc[7] and Jec-qa[31]. However, these datasets require extensive human intervention and are difficult to extend.

2.2 Natural Language Generation

Our task aims at revising the court's view draft based on the fact description and court's view draft, which can be viewed as a Natural Language Generation (NLG) task. NLG is a sub-field of NLP that studies methods of automatically transforming data to a human-readable text. In practice, there are two major types of NLG applications: template-based NLG and advanced NLG.

Template-based NLG uses templates that get in advance and insert data into the templates. Such methods heavily rely on hard-coded rules, which makes them less flexible. Since template-based NLG methods have a limited number of templates and require strict data formats, they can not be easily applied to different tasks.

In advanced NLG, thanks to the recent success of sequence-to-sequence models [22], in which recurrent neural networks (RNNs) reading and generating text simultaneously, text generation task has been made more feasible. Bahdanau et al. [2] firstly applied the attention mechanism into the NLG task, where the decoder focuses on different parts of the input at each decode step. Based on the Pointer-Nerwords[23], See et al. [20] proposed a Pointer-Generator Networks (PGN), which can effectively solve the Out-Of-Vocabulary (OOV) problem.

NLG has been widely studied and applied to many tasks, such as machine translation[6, 24], text summarization [19] and question answering [1, 14].

Although the previous works on NLG can produce fluent sentences, they rarely pay attention to the revision process, which is important in the real applications.

3 PROBLEM FORMULATION

In this work, we focus on the problem of the court's view revision, where the input is the court's view draft and the fact description, and the output is the revised court's view. We formulate our problem

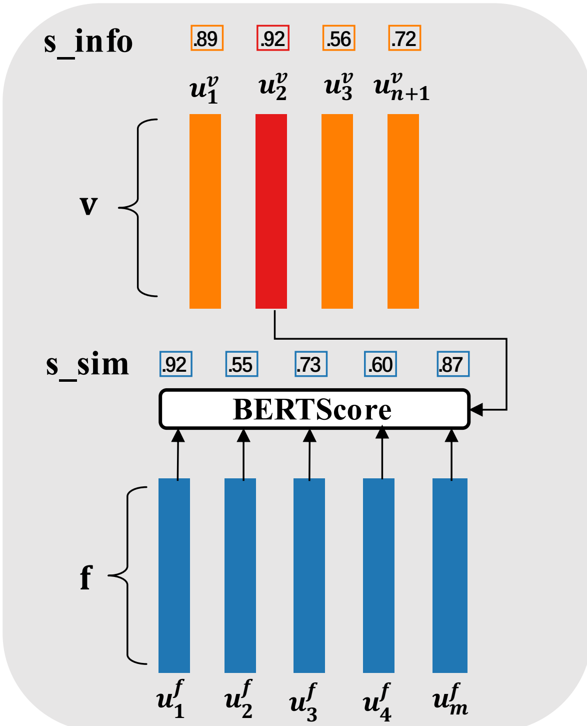

Figure 2: An illustrations of dataset construction. The red sentence will be masked because it gains the highest factual informative score.

with the definition of the fact description, the court's view draft, and the revised (whole) court's view, as shown in Fig. 1.

Fact description (F) consists of several descriptive sentences, which describe the identified facts (relevant events that have happened) in a case, as Fig. 1 shows. Here, we denote the fact description in a case as $\mathbf{f} = \{u_t^f\}_{t=1}^m$, where m is the number of sentences.

Court's view Draft (D) contains several rational sentences that summarized from fact description to interpret the judgments, but is incomplete. Here, we denote the court's view draft as $\mathbf{d} = \{u_t^d\}_{t=1}^n$, where n is the number of sentences.

Revised Court's view (V) is defined as the draft with a generated factual informative sentence in this paper. Here, we denote the revised court's view as $\mathbf{v} = \{u_t^v\}_{t=1}^{n+1}$. Specifically, we denote u_p^t as the generated sentence, where p is the position of the sentence.

Then, the problem of court's view revision can be denoted as follow:

Problem 1 (Court's View Revision). *Given the fact description $\mathbf{f} = \{w_t^f\}_{t=1}^n$ and the court's view draft $\mathbf{d} = \{u_t^d\}_{t=1}^n$, our task is to generate the revised court's view $\mathbf{v} = \{u_t^v\}_{t=1}^{n+1}$.*

4 DATASET CONSTRUCTION

In this section, we first introduce the BERTScore method. Then, we describe how to construct the draft dataset.

4.1 BERTScore

BERTScore[1] [29] leverages the pre-trained contextual embeddings from BERT and matches words in the two sentences by cosine similarity. It has been shown to correlate with human judgment on sentence-level and system-level evaluation. Moreover, BERTScore computes precision, recall, and F1 measure, which can be useful for evaluating different language generation tasks.

In this paper, we use the F1 score to measure the similarity.

4.2 Draft Construction

As Fig. 2 shows, for each sentence u_i^v in court's view $\mathbf{v}$, we calculate its similarity score with each sentence in fact description $\mathbf{f}$ through BERTScore. The similarity score between u_i^v and u_j^f is calculated as follow:

$$s\_sim_j^i = BERTScore(u_i^v, u_j^f) \tag{1}$$

Then we take the highest similarity score in s_sim^i as the factual informative score of u_i^v, which is produced as follow:

$$s\_info_i = \max_{1<j<m} (s\_sim_j^i) \tag{2}$$

After that, every sentence in court's view $\mathbf{v}$ gains a factual informative score. We assume the sentence at position p gains the highest factual informative score. We mask the sentence u_p^v, and the rest of the sentences make up the draft $\mathbf{d}$.

$$\mathbf{d} = \sum_{1<t<n+1 \ \& \ t!=p} u_t^v \tag{3}$$

Take Fig. 1 as an example, the red sentence is the most factual informative sentence, and should be masked.

5 PRELIMINARY MODEL

Since the task is in the process, we haven't designed the final model. In this section, we will describe the preliminary sequence-to-sequence model, which consists of an encoder and a decoder.

5.1 Encoder

We first concat the fact description $\mathbf{f}$ and darft $\mathbf{d}$ to $\mathbf{x}$, then transforms the words to embeddings. The embedding sequences are fed to the Bi-LSTM, producing a sequence of hidden states $\mathbf{h}$.

5.2 Decoder

We adopt attention mechanism in the decoder. At each decode step t, given the hidden states $\mathbf{h}$ and the decode state s_t, the attention distribution a^t is calculated as follow:

$$e_i^t = v^T tanh(W_h h_i + W_s s_t + b_{attn}) \tag{4}$$

$$a^t = softmax(e^t) \tag{5}$$

where v, W_h, W_s and b_attn are learnable parameters. Next, the attention distribution is used to produce a weighted sum of the hidden states, known as the context vector h_t^*:

$$h_t^* = \sum_i a_i^t h_i \tag{6}$$

The vocabulary distribution P_{vocab} is then calcuated as follow:

$$P_{vocab} = softmax(V'(V[s_t, h_t^*]) + b) + b') \tag{7}$$

[1]https://github.com/Tiiiger/bert_score

Table 1: Results of court's view revision

Settings	ROUGE			BLEU			BERTScore		
	r-1	r-2	r-l	b-1	b-2	b-n	p	r	f1
SSwP	**44.4**	**30.6**	**44.0**	**55.2**	**43.7**	**35.9**	**83.0**	**81.0**	**81.9**
LSwP	40.5	22.2	39.1	55.1	41.5	33.6	79.6	77.3	78.4
LSw/oP	41.7	23.1	40.1	56.0	42.6	34.6	80.0	77.6	78.7

Table 2: The hyperparameters of the model.

Name	value	Note
hidden_dim	256	dimension of hidden states
emb_dim	300	dimension of word embeddings
batch_size	32	minibatch size
max_enc_steps	600	max timesteps of encoder (max source text tokens)
max_dec_steps	100	max timesteps of decoder (max generated text tokens)
beam_size	4	beam size for beam search decoding
min_dec_steps	1	Minimum sequence length of generated text
vocab_size	10000	Size of vocabulary

where V, V', b and b' are learnable parameters.

5.3 Training

During the training, the loss for step t is the negative log-likelihood of the target word w_t^*:

$$\mathcal{L}_t = -logP(w_t^*) \tag{8}$$

Thus, the overall loss is:

$$\mathcal{L}_{gen} = \frac{1}{T}\sum_{t=0}^{T}\mathcal{L}_t \tag{9}$$

where T is the length of target sequence.

6 PRELIMINARY EXPERIMENTS

In this section, we will show some preliminary experiments with their results.

6.1 Metrics

Firstly, we introduce the evaluation metrics we adopt.

ROUGE. ROUGE is a set of metrics used in the NLP task. The metrics compare an automatically produced result against the result of a reference. We use the official ROUGE script and keep the results of ROUGE-1, ROUGE-2 and ROUGE-L. ROUGE-1 and ROUGE-2 refer to the overlap of unigram and bigram between the generated and reference documents, respectively. ROUGE-L is a Longest Common Subsequence (LCS) based statistics.

BLEU. BLEU[16] is a method of automatic text-generation evaluation that correlates highly with human evaluation. We use BLEU-1, BLEU-2 to evaluate from the perspectives of unigram, bigram. BLEU-N is an average of BLEU-1, BLEU2, BLEU-3, BLEU-4.

BERTScore. Still, we use BERTScore to evaluate the results.

6.2 Experiment Settings

We describe three experiment settings here.

6.2.1 Short Sentence With Position. In this setting, we mark the masked position with a special token in the draft, so the model knows where it should add a sentence. And the length of the masked sentence is less than 10 tokens.

6.2.2 Long Sentence With Position. In this setting, the length of the masked sentence is unlimited.

6.2.3 Long Sentence Without Position. In this setting, the task becomes more difficult. The model doesn't know the position of the masked sentence in advance.

6.3 Experiment Results

As Tab. 1 shows, SSwP achieves the best performance among the three settings, which means the short sentence is easier to generate. The performance gap between LSwP and LSw/oP is small, which means the model has little dependence on the position information.

Overall, all the scores are relatively high, which shows the feasibility of our task.

6.4 Experiment Details

We use Gensim [17] with a large-scale legal corpus to train a language model as the pre-trained model, then use it to initialize the word embeddings. The more details are shown in Fig. 2.

We use two V100 GPU to train and evaluate the model.

7 CONCLUSION AND FUTURE WORK

In this paper, we focus on the task of court's view revision, which has great particle application value. We adopt a novel method to automatically construct a draft dataset. Preliminary experimental results show the feasibility and effectiveness of the task.

In the future, we will explore in the following directions: (1) Construct a more realistic draft dataset, like changing the sentence order. (2) Develop a effective method for this task, which can make better use of the input information. (3) Verify the effectiveness of the method on other datasets. (4) Combine the revision method with other generative methods to form a two-stage method and improve the generated results.

8 ACKNOWLEDGEMENT

This work is supported by National Key R&D Program of China (2018YFC0830200;2018YFC0830206;2018AAA0101900).

REFERENCES

[1] Sugato Bagchi and Laura Wynter. 2013. Method for a natural language question-answering system to complement decision-support in a real-time command center. US Patent 8,601,030.

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[3] Ilias Chalkidis, Ion Androutsopoulos, and Nikolaos Aletras. 2019. Neural Legal Judgment Prediction in English. *arXiv preprint arXiv:1906.02059* (2019).

[4] Yen-Liang Chen, Yi-Hung Liu, and Wu-Liang Ho. 2013. A text mining approach to assist the general public in the retrieval of legal documents. *Journal of the American Society for Information Science and Technology* 64, 2 (2013), 280–290.

[5] Yuh-Shyan Chen, Shin-Wei Chiang, and Tong-Ying Juang. 2019. *A Few-Shot Transfer Learning Approach Using Text-label Embedding with Legal Attributes for Law Article Prediction.* Technical Report. EasyChair.

[6] Yong Cheng. 2019. Semi-supervised learning for neural machine translation. In *Joint Training for Neural Machine Translation.* Springer, 25–40.

[7] Xingyi Duan, Baoxin Wang, Ziyue Wang, Wentao Ma, Yiming Cui, Dayong Wu, Shijin Wang, Ting Liu, Tianxiang Huo, Zhen Hu, et al. 2019. CJRC: A Reliable Human-Annotated Benchmark DataSet for Chinese Judicial Reading Comprehension. In *China National Conference on Chinese Computational Linguistics.* Springer, 439–451.

[8] Xinyu Duan, Yating Zhang, Lin Yuan, Xin Zhou, Xiaozhong Liu, Tianyi Wang, Ruocheng Wang, Qiong Zhang, Changlong Sun, and Fei Wu. 2019. Legal Summarization for Multi-role Debate Dialogue via Controversy Focus Mining and Multi-task Learning. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management.* 1361–1370.

[9] Albert Gatt and Emiel Krahmer. 2018. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research* 61 (2018), 65–170.

[10] Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, and Maosong Sun. 2018. Few-shot charge prediction with discriminative legal attributes. In *Proceedings of the 27th International Conference on Computational Linguistics.* 487–498.

[11] Xin Jiang, Hai Ye, Zhunchen Luo, Wenhan Chao, and Wenjia Ma. 2018. Interpretable rationale augmented charge prediction system. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations.* 146–151.

[12] Kamran Kowsari, Kiana Jafari Meimandi, Mojtaba Heidarysafa, Sanjana Mendu, Laura Barnes, and Donald Brown. 2019. Text classification algorithms: A survey. *Information* 10, 4 (2019), 150.

[13] Wan-Chen Lin, Tsung-Ting Kuo, Tung-Jia Chang, Chueh-An Yen, Chao-Ju Chen, and Shou-de Lin. 2012. Exploiting machine learning models for chinese legal documents labeling, case classification, and sentencing prediction. *Processdings of ROCLING* (2012), 140.

[14] Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. 2018. The natural language decathlon: Multitask learning as question answering. *arXiv preprint arXiv:1806.08730* (2018).

[15] Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2019. SemEval-2016 task 4: Sentiment analysis in Twitter. *arXiv preprint arXiv:1912.01973* (2019).

[16] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics.* Association for Computational Linguistics, 311–318.

[17] Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks.* ELRA, Valletta, Malta, 45–50. http://is.muni.cz/publication/884893/en.

[18] Sara Rosenthal, Noura Farra, and Preslav Nakov. 2019. SemEval-2017 task 4: Sentiment analysis in Twitter. *arXiv preprint arXiv:1912.00741* (2019).

[19] Alexander M Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685* (2015).

[20] Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368* (2017).

[21] Roberta A Sinoara, Jose Camacho-Collados, Rafael G Rossi, Roberto Navigli, and Solange O Rezende. 2019. Knowledge-enhanced document embeddings for text classification. *Knowledge-Based Systems* 163 (2019), 955–971.

[22] I Sutskever, O Vinyals, and QV Le. 2014. Sequence to sequence learning with neural networks. *Advances in NIPS* (2014).

[23] Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In *Advances in neural information processing systems.* 2692–2700.

[24] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144* (2016).

[25] Ziyu Wu, Hongxu Hou, Ziyue Guo, Xuejiao Wang, and Shuo Sun. 2019. Mongolian-Chinese Unsupervised Neural Machine Translation with Lexical Feature. In *China National Conference on Chinese Computational Linguistics.* Springer, 334–345.

[26] Chaojun Xiao, Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Yansong Feng, Xianpei Han, Zhen Hu, Heng Wang, et al. 2018. Cail2018: A large-scale legal dataset for judgment prediction. *arXiv preprint arXiv:1807.02478* (2018).

[27] Chaojun Xiao, Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Zhiyuan Liu, Maosong Sun, Tianyang Zhang, Xianpei Han, Heng Wang, Jianfeng Xu, et al. 2019. CAIL2019-SCM: A Dataset of Similar Case Matching in Legal Domain. *arXiv preprint arXiv:1911.08962* (2019).

[28] Guangyi Xiao, Even Chow, Hao Chen, Jiqian Mo, Jingzhi Guo, and Zhiguo Gong. 2017. Chinese Questions Classification in the Law Domain. In *2017 IEEE 14th International Conference on e-Business Engineering (ICEBE).* IEEE, 214–219.

[29] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675* (2019).

[30] Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018. Legal judgment prediction via topological learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.* 3540–3549.

[31] Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. 2019. JEC-QA: A Legal-Domain Question Answering Dataset. *arXiv preprint arXiv:1911.12011* (2019).

[32] Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. 2020. How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence. *arXiv preprint arXiv:2004.12158* (2020).