Decremental Learning for Domain Adaptation in Neural Machine Translation

Kai Jin
School of Foreign Languages
Qilu University of Technology
250353, Jinan, Shandong, China
jk@qlu.edu.cn

Wuying Liu(✉)
Shandong Key Laboratory of Language Resources Development and Application
Ludong University
264025, Yantai, Shandong, China
wyliu@ldu.edu.cn

Abstract

Domain adaptation has become a research hotspot in neural machine translation (NMT). Based on the quantitative analysis results of two adaptabilities, we propose a two-phase decremental learning framework for scenes involving large-scale common bilingual parallel sentences and small-scale monolingual domain texts. In the domain filtering phase, we filter common sentence pairs with domain texts and train two domain NMT models with these domain sentence pairs. In the quality filtering phase, we use the trained domain NMT models to translate the domain sentence pairs and evaluate the translation quality to delete low-quality domain sentence pairs to get high-quality ones. With these high-quality sentence pairs, we train optimized bidirectional domain NMT models adapted to the specific domain. The experimental results of English-Chinese bidirectional NMT in the legal domain show that when the number of training steps decreases from 1,714,250 to 446,100, the BLEU value of the English-Chinese NMT model increases from 45.41 to 47.88, and that of the Chinese-English NMT model increases from 32.03 to 34.12, which demonstrates that the decremental learning is effective in achieving state-of-the-art performance with greatly reduced training space-time costs.

Keywords: Domain Adaptation, Domain Filtering, Quality Filtering, Bidirectional Legal NMT

1 Introduction

Recently, resource-rich general neural machine translation (NMT) has been well studied, resulting in myriad brilliant algorithms, data resources, and practical tools (Tan et al., 2020). With the explosive growth of language data, resource-rich machine translation (MT) research is paying more attention to transfer learning methods. Domain adaptation is a method of transfer learning featuring the same tasks and different domains. In our case, both tasks are MT tasks but the source domain and the target domain are different – the training set is a wide-domain English-Chinese corpus while the test set is a legal domain English-Chinese corpus. Different domains in domain adaptation can be embodied in that there is inconsistent data distribution between the source and target domains, or that there are a large number of labeled out-domain samples while a minimal number of or no labeled in-domain samples.

There are two main approaches to the study of domain adaptation NMT, which are evolved and developed from the study of domain adaptation statistical MT (Chu and Wang, 2018). One is model centric approach, and the other is data centric approach. There is an overlap between the two approaches since the former one may also use monolingual or parallel sentence data.

The model centric approach focuses on improving algorithms of neural networks. (1) Intervening in the architecture (Domhan and Hieber, 2017) (Britz et al., 2017) (Kobus et al., 2017). (2) Intervening in the training (Chen et al., 2017) (Wang et al., 2018) (Varga, 2017) (Dakwale and Monz, 2017) (Dou et al., 2019) (Chu et al., 2017) (Barone et al., 2017). (3) Intervening in the decoding (Adams et al., 2022) (Freitag and Al-Onaizan, 2016) (Khayrallah et al., 2017).

The data centric approach is more suitable for engineering applications in a shorter time. (1) By utilizing in-domain monolingual data (Currey et al., 2017) (Zhang and Zong, 2016) (Cheng et al., 2016). (2) By utilizing out-domain high-quality parallel data (Wang et al., 2017) (Wees et al., 2017). (3) Utilizing parallel data of unknown quality (Saunders, 2022) (Hu et al., 2019).

The research of domain adaptation MT aims for better performance of in-domain MT models by using information-rich out-domain samples. When there is an inconsistent distribution between the data of the training set and the test set, the trained model resulting from machine learning often overfits the source domain, thereby reducing the generalizability in the target domain. An ideal of domain adaptation MT should produce with higher efficiency a well-performing MT model corresponding to the domain. The data in the above data centric approaches are very close to real data environments. How to use this kind of parallel data of unknown quality to achieve NMT suitable for specific domains is a more specific and practical research issue. For this problem, we made quantitative analysis on two adaptabilities and proposed a novel decremental learning idea.

2 Examination of Adaptabilities

To calculate the adaptability between source domain and target domain, we first calculate the union V={v_1, v_2, …, v_n} of the source domain data vocabulary Vs and the target domain data vocabulary Vt. Then, basing on the n-dimensional vector base V, we calculate the word frequency vector S=<s_1, s_2, ..., s_n> of the source domain data and the word frequency vector T=<t_1, t_2, ..., t_n> of the target domain data. Finally, we measure the difference between the source domain and the target domain by Kullback-Leibler Divergence (KLD) and Maximum Mean Discrepancy (MMD). The KLD is widely used in adaptation machine learning tasks as a loss function (Nguyen et al., 2022). The MMD is mainly used to measure the distance between two different but related distributions (Wang et al., 2020).

We calculate the KLD and MMD of sentence pairs between the source domain and the target domain of language A, the KLD and MMD of sentence pairs between the source domain and the target domain of language B, and the KLD and MMD of parallel sentence pairs between the source domain and the target domain. We make a statistical analysis of three sets (LAW07, LAW08 and LAW09) of parallel sentences in both English and Chinese. LAW07 contains 21,942,400 pairs of sentences in the source domain, LAW08 contains 5,899,520 pairs of sentences in the source domain, and LAW09 contains 5,710,080 pairs of sentences in the source domain. These three sets all contain the same 50,000 pairs of sentences in the target domain. The English-Chinese adaptabilities are shown in Table 1. As the absolute values of KLD and MMD are very small, they are multiplied by 10^6 to be shown in the table. The values in Table 1 show that the source domain of LAW09 is the closest to the target domain, whether from the monolingual perspective of English and Chinese or from the English-Chinese bilingual perspective.

Source (number of sentences)	Target (number of sentences)	KLD×10^6	MMD×10^6
LAW07.train.eng (21,942,400)	LAW07.test.eng (50,000)	11.028	21.935
LAW08.train.eng (5,899,520)	LAW08.test.eng (50,000)	1.144	4.292
LAW09.train.eng (5,710,080)	LAW09.test.eng (50,000)	**1.098**	**2.384**
LAW07.train.zho (21,942,400)	LAW07.test.zho (50,000)	93.858	262.260
LAW08.train.zho (5,899,520)	LAW08.test.zho (50,000)	9.254	99.182
LAW09.train.zho (5,710,080)	LAW09.test.zho (50,000)	**8.911**	**69.141**
LAW07.train.engzho (43,884,800)	LAW07.test.engzho (100,000)	41.145	30.041
LAW08.train.engzho (11,799,040)	LAW08.test.engzho (100,000)	4.695	20.027
LAW09.train.engzho (11,420,160)	LAW09.test.engzho (100,000)	**4.297**	**19.550**

Table 1: Adaptabilities of English-Chinese Corpus

For the above three sets of corpus, we also calculate the Chinese-English adaptabilities, and the results are shown in Table 2. We compare the values in Table 1 and Table 2 and find that among the three groups of values the results of Chinese monolingual data are exactly the same, and there is no difference between the relative size relationship of English monolingual data and Chinese-English bilingual data, in spite of their different absolute values. From this result we can draw the same conclusion on adaptability of Chinese-English translation as that of English-Chinese translation. The difference mentioned above in Table 1 and Table 2 is due to the fact that all the letters in the English sentences in English-Chinese translation corpus are lowercase, and those in Chinese-English translation corpus follow the rule of capitalization in English, while, on the other hand, there is no such difference of form in Chinese.

Source (number of sentences)	Target (number of sentences)	KLD×10^6	MMD×10^6
LAW07.train.zho (21,942,400)	LAW07.test.zho (50,000)	93.858	262.260
LAW08.train.zho (5,899,520)	LAW08.test.zho (50,000)	9.254	99.182
LAW09.train.zho (5,710,080)	LAW09.test.zho (50,000)	**8.911**	**69.141**
LAW07.train.eng (21,942,400)	LAW07.test.eng (50,000)	11.502	24.319
LAW08.train.eng (5,899,520)	LAW08.test.eng (50,000)	1.510	4.768
LAW09.train.eng (5,710,080)	LAW09.test.eng (50,000)	**1.418**	**2.452**
LAW07.train.zhoeng (43,884,800)	LAW07.test.zhoeng (100,000)	49.440	25.272
LAW08.train.zhoeng (11,799,040)	LAW08.test.zhoeng (100,000)	7.295	15.736
LAW09.train.zhoeng (11,420,160)	LAW09.test.zhoeng (100,000)	**6.455**	**14.440**

Table 2: Adaptabilities of Chinese-English Corpus

Based on the values in Table 1 and Table 2, it can be drawn that the domain adaptability is not necessarily stronger when the corpus is larger. In English-Chinese bidirectional MT, LAW07 has the worst adaptability despite of its largest scale, while LAW09 has the best adaptability despite of its smallest scale.

3 Decremental Learning Framework

Inspired by the results of the statistical analysis of adaptabilities, we explored a decremental learning method which is a new idea of engineering level. Our decremental learning framework is shown in Figure 1. It mainly includes a domain filter, a quality filter and three identical NMT trainers. The preset data for implementing the framework includes common parallel sentences and domain text resources. Common parallel sentences refer to bilingual data that are large-scale, easily-available, domain-independent and errors-possibly-contained. Domain text resources can be monolingual or bilingual data, such as domain word sets, domain text documents, etc.

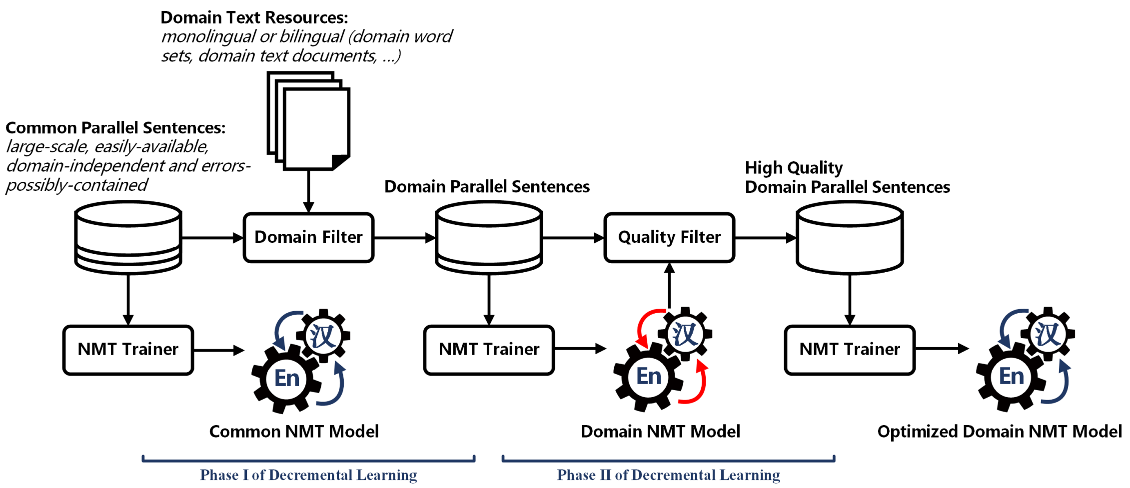

Figure 1: Decremental Learning Framework

This decremental learning framework is a meta structure independent of filtering algorithms, MT algorithms, and source languages and target languages. Here we take legal domain English-Chinese bidirectional MT as an example to implement the framework. Phase I classifies the sentence pairs in the common parallel sentences as legal domain and nonlegal domain. This is done by the domain filter with domain text resources. The result of this phase is parallel sentences. Then we train NMT models on an NMT trainer with common parallel sentences and domain parallel sentences, and respectively get an English-Chinese bidirectional NMT model and a legal-domain English-Chinese bidirectional NMT model. In Phase II we first translate every sentence pair of the domain parallel sentences with the legal-domain English-Chinese bidirectional NMT model, calculate the similarity between the original sentences and the sentences translated by the NMT model with Levenshtein string distance function in the quality filter, and delete the less similar sentence pairs based on the preset threshold. Hence we get high-quality domain parallel sentences and an optimized English-Chinese bidirectional NMT model in legal domain after training the model again. Three bidirectional English-Chinese NMT models (altogether six models) are trained in the whole framework, among which the common English-Chinese and Chinese-English models are only used for experimental comparison.

4 Domain Adaptation NMT Algorithm

To meet the actual domain adaptation NMT requirements, we designed a domain adaptation NMT algorithm based on our decremental learning framework. The algorithm is shown in Figure 2.

```
1. // Domain Adaptation NMT Algorithm
2. Input:   <String, String>[] train; // common parallel sentences of original training set
3.          <String, String>[] dev; // domain parallel sentences of development set
4.          <String, String>[] test; // domain parallel sentences of test set
5.          String[] dtr; // domain text resources
6. Output:  Model odnmt; // optimized domain NMT model
7. Model cnmt; // common NMT model only for comparison
8. cnmt.st ← NMTTrainer.train(train, dev, test);
9. cnmt.ts ← NMTTrainer.train(train, dev, test);
10. <String, String>[] train ← DomainFilter.filter(train, dtr); // domain parallel sentences
11. Model dnmt; // domain NMT model
12. dnmt.st ← NMTTrainer.train(train, dev, test);
13. dnmt.ts ← NMTTrainer.train(train, dev, test);
14. <String, String>[] mtout;
15. For Integer i ← 1 To train.size Do
16.         String TSen ← dnmt.st.translate(train[i].SSen);
17.         String SSen ← dnmt.ts.translate(train[i].TSen);
18.         mtout[i] ← <SSen, TSen>;
19. End For
20. <String, String>[] train ← QualityFilter.filter(train, mtout); // high quality domain parallel sentences
21. odnmt.st ← NMTTrainer.train(train, dev, test);
22. odnmt.ts ← NMTTrainer.train(train, dev, test);
23. Return odnmt.
```

Figure 2: Domain Adaptation NMT Algorithm

In Figure 2, **input** is the initial training set (train), development set (dev), test set (test) and domain text resource (dtr), and **output** is the opitimized domain NMT model (odnmt). The NMT training function (NMTTrainer.***train***) in lines 8, 9, 12, 13, 21 and 22 is implemented by the coder-decoder based on the attention mechanism. The input parameters are training set (train), development set (dev) and test set (test), and the output is an NMT model. The most critical two-phase filters are implemented by mature algorithms. The domain filter function (DomainFilter.***filter***) in line 10 is implemented by the text classification (SFITC) algorithm based on the String-Frequency Index (Liu et al., 2014). In Figure 2, the quality filter function (QualityFilter.***filter***) in line 20 is implemented with the algorithm of single-engine-based ensemble MT filtering (Liu and Wang, 2022).

5 Experiment

To verify the effectiveness and efficiency of decremental learning, we conducted an English-Chinese bidirectional NMT experiment in the legal domain.

5.1 Implement

We first implement the domain adaptation NMT algorithm according to the decremental learning framework. The parameters of the coder-decoder integrated in the algorithm mainly include the number of neurons (*num_units* = 512), the number of coder-decoder layers (*num_encoder_layers* = *num_decoder_layers* = 4), the number of training rounds (*epoch* = 10), the batch size (*batch_size* = 128) and the beam search width (*beam_width* = 10). The other parameters remain default values. Then we run the algorithm to obtain three English-Chinese NMT models and three Chinese-English NMT models. Finally, an interactive interface is added to form a web server as is show in Figure 3.

In terms of experimental data preparation, we first construct a parallel sentence bank in legal domain (PSB.ld) containing 100,000 pairs of English-Chinese parallel sentences and a bilingual word set in legal domain (BWS.ld) containing 76,792 items. According to the principle of simple random sampling, PSB.ld is divided into a development set (50,000 sentence pairs) and a test set (50,000 sentence pairs). BWS.ld is mainly used for the domain filter. Then we collect 21,942,400 pairs of English-Chinese parallel sentences (LAW07 corpus) from the Internet to form an initial training set.

Figure 3: English and Chinese Bidirectional Legal Machine Translation Platform

In the course of the experiment, LAW07 successively produces domain parallel sentences (LAW08 corpus) and high-quality domain parallel sentences (LAW09 corpus) after decremental learning. We also reprocessed the data. For Chinese sentences, we separate each character with spaces. In the English-Chinese NMT experiment, we convert all English letters in lowercase and separate each word with spaces; while in the Chinese-English NMT experiment, we only separate each word with spaces without converting them.

5.2 English-Chinese NMT Experiment

The results of the English-Chinese NMT experiment are shown in Table 3. After phase I decremental learning, the 21,942,400 pairs of English-Chinese parallel sentences in the initial training set LAW07 corpus of unknown quality are reduced to 5,899,520 pairs in the domain training set LAW08 corpus. After phase II decremental learning, 5,710,080 pairs of English-Chinese parallel sentences of the high-quality domain training set LAW09 corpus are obtained. The BLEU values of the NMT models respectively trained with the three corpus are BLEU (LAW07) = 45.41, BLEU (LAW08) = 47.13, and BLEU (LAW09) = 47.88. It should be mentioned that the BLEU in this paper refers to the classic MT evaluation metric BLEU4. It can be seen that although the two-phase decremental learning reduces the scale of the training corpus, the MT performance does not reduce but improves. This also shows that decremental learning can effectively enhance the domain adaptability of the training corpus.

	number of sentence pairs of training set	**number of sentence pairs of development set**	**number of sentence pairs of test set**	**number of training steps**	**BLEU**
LAW07	21,942,400	50,000	50,000	1,714,250	45.41
LAW08	5,899,520	50,000	50,000	460,900	47.13
LAW09	5,710,080	50,000	50,000	446,100	**47.88**

Table 3: Results of English-Chinese NMT Model

The learning curves of the three NMT models during the training process in the English-Chinese NMT experiment is shown in Figure 4, with the abscissa being the training steps and the ordinate being BLEU values. The values on the ordinate of the end points of the three learning curves show that LAW07 model has the most training steps (1,714,250), and the number of the training steps of LAW08 model (460,900) is close to that of LAW09 model (446,100). On the other hand, the values on the abscissa of the end points of the three learning curves show that the performance of the LAW09 model is the best, which has the fewest training steps. This demonstrates that the result of decremental learning is a more efficient NMT model.

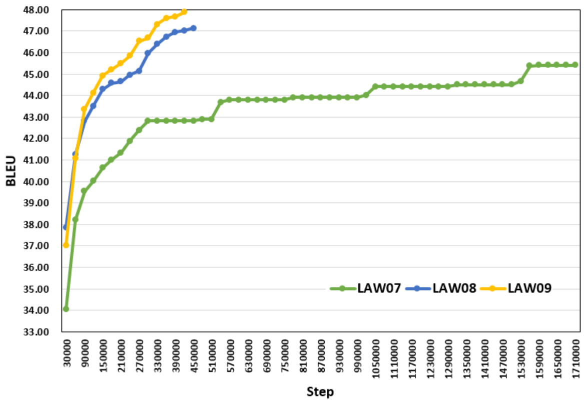

Figure 4: Learning Curves of English-Chinese NMT

5.3 Chinese-English NMT experiment

We also run the Chinese-English NMT experiment in the legal domain in reverse on the same corpus. Table 4 shows the results of the Chinese-English NMT experiment. Compared with Table 3, it is easy to know that the results of the Chinese-English NMT experiment are consistent with those of the English-Chinese NMT experiment. It is found that the translation performance of LAW09 (BLEU (LAW09) = 34.12) is better than that of LAW08 (BLEU (LAW08) = 33.81) and that of LAW07 (BLEU (LAW07) = 32.03). This demonstrates that decremental learning is also effective for Chinese-English NMT. We also find that the BLEU value of the Chinese-English translation is more than 10 points lower than that of the English-Chinese translation, both of which are trained with the same corpus. This is due to the need to maintain the uppercase and lowercase form of letters in Chinese-English NMT training, which increases data sparsity compared with all lowercase letter forms in English-Chinese NMT training.

	number of sentence pairs of training set	number of sentence pairs of development set	number of sentence pairs of test set	number of training steps	BLEU
LAW07	21,942,400	50,000	50,000	1,714,250	32.03
LAW08	5,899,520	50,000	50,000	460,900	33.81
LAW09	5,710,080	50,000	50,000	446,100	**34.12**

Table 4: Results of Chinese-English NMT Model

The learning curves of the three NMT models during the training process in the Chinese-English NMT experiment is shown in Figure 5. Compared to Figure 4, it is easy to find that the learning results of NMT are consistent in both Chinese-English and English-Chinese.

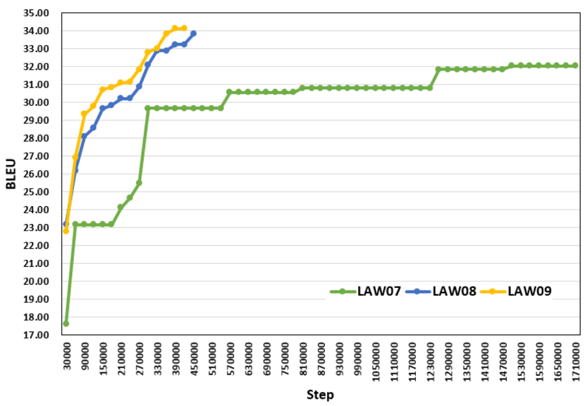

Figure 5: Learning Curves of Chinese-English NMT

5.4 Examples and Discussions

In order to show the effect of our English-Chinese NMT model more intuitively, we randomly selected three sentence pairs (Sen0, Sen1, and Sen2) for display. Table 5 lists the output translations of our three models LAW07, LAW08, and LAW09, as well as the output translations of Google, DeepL, and NiuTrans systems, where REF represents the human reference translation in the original sentence pair. A cursory glance shows that the output of all 6 models is acceptable. Careful discussion revealed that: "Pharmaceutical trading enterprise" in Sen0 is translated as "药品贸易企业" by DeepL, while the other 5 models are all translated as "药品经营企业"; "relevant judicial interpretations" in Sen1 is translated as "相关司法解释" by LAW07 and "有关司法解释" by DeepL, while other models add "的规定"; "regulating property relations” in Sen2 is translated as “规范产权关系” by LAW08, “调整财产关系” by the 3 general business models, and “规范财产关系” by our LAW07 and LAW09.

MT	Sen0.eng	Sen0.zho	Sen1.eng	Sen1.zho	Sen2.eng	Sen2.zho
LAW07	Pharmaceutical trading enterprise means an enterprise exclusively or partly engaged in the trade of pharmaceuticals.	药品经营企业是指专营或者部分从事药品贸易的企业。	The people's procuratorates shall, based on the facts, nature, circumstances and degree of harm to society of a crime, provide sentencing recommendations in accordance with the Criminal Law, Criminal Procedure Law and relevant judicial interpretations.	人民检察院应当根据犯罪的事实、性质、情节和危害程度，根据刑法、刑事诉讼法和相关司法解释，提出量刑建议。	The compilation of the civil code shall be advanced more quickly, the legal system relating to real rights, contract and property rights shall be improved, laws, regulations and provisions in contravention of fairness shall be reviewed, and equal protection shall be established as a basic principle of regulating property relations.	更快推进民法汇编，完善与权利、合同、财产权有关的法律制度，审查违反公平的法律、法规、规定，确立平等保护作为规范财产关系的基本原则。
LAW08		药品经营企业是指专营或者部分从事药品经营的企业。		人民检察院应当根据犯罪社会的事实、性质、情节和程度，依照刑法、刑事诉讼法和相关司法解释的规定提出量刑建议。		加快编纂《民法典》，完善与物权、合同、产权有关的法律制度，审查违反公平的法律、法规和规定，以平等保护为规范产权关系的基本原则。
LAW09		药品经营企业是指专营或者部分从事药品交易的企业。		人民检察院应当根据犯罪社会危害的事实、性质、情节和程度，依照刑法、刑事诉讼法和有关司法解释的规定提出量刑建议。		更快推进民法汇编，完善与物权、合同、产权相关的法律制度，审查违反公平的法律法规规定，建立平等保护作为规范财产关系的基本原则。
Google		药品经营企业，是指专门从事药品经营或者兼营药品经营的企业。		人民检察院应当根据犯罪的事实、性质、情节和社会危害程度，依照刑法、刑事诉讼法和有关司法解释的规定，提出量刑建议。		加快推进民法典编纂，完善物权、合同权、财产权法律制度，清理违背公平的法律、法规和规定，把平等保护确立为调整财产关系的基本原则。
DeepL		药品贸易企业是指专门或部分从事药品贸易的企业。		人民检察院应当根据犯罪的事实、性质、情节和对社会的危害程度，依照刑法、刑事诉讼法和有关司法解释提出量刑建议。		加快推进民法典编纂工作，完善物权、合同、财产权等相关法律制度，审查有违公平原则的法律法规和规定，将平等保护作为调整财产关系的基本原则。
NiuTrans		药品经营企业是指专门或者部分从事药品经营的企业。		人民检察院应当根据犯罪的事实、性质、情节和对社会的危害程度，依照刑法、刑事诉讼法和有关司法解释的规定提出量刑建议。		加快推进民法典编纂工作，完善物权、合同、财产权法律制度，对有违公平的法律法规和规定进行审查，把平等保护确立为调整财产关系的基本原则。
REF		药品经营企业是指经营药品的专营企业或者兼营企业。		人民检察院应当根据犯罪的事实、犯罪的性质，情节和对于社会的危害程度，依照刑法、刑事诉讼法以及相关司法解释的规定提出量刑建议。		加快推进民法典编纂工作，完善物权、合同、知识产权相关法律制度，清理有违公平的法律法规条款，将平等保护作为规范财产关系的基本原则。

Table 5: English-Chinese NMT Examples

Table 6 shows the effect of the Chinese-English NMT model. Comparing the reference translations, we find that the performance of each model has decreased, which is consistent with the BLEU evaluation results. In addition to the sparse data caused by keeping English case when we train NMT models, another potential reason is that the scale of English Token space is very different from that of Chinese Token space.

MT	Sen0.zho	Sen0.eng	Sen1.zho	Sen1.eng	Sen2.zho	Sen2.eng
LAW07	药品经营企业是指经营药品的专营企业或者兼营企业。	“ Pharmaceutical business enterprises ” means specialized enterprises that engage in drug management or concurrently operated enterprises.	人民检察院应当根据犯罪的事实、犯罪的性质，情节和对于社会的危害程度，依照刑法、刑事诉讼法以及相关司法解释的规定提出量刑建议。	The people's procuratorate shall, in accordance with the facts of the crime, the nature of the crime, the circumstances of the crime and the extent of harm to the society, submit sentencing suggestions in accordance with the provisions of the Criminal Law, the Criminal Procedure Law and the relevant judicial interpretations.	加快推进民法典编纂工作，完善物权、合同、知识产权相关法律制度，清理有违公平的法律法规条款，将平等保护作为规范财产关系的基本原则。	We shall accelerate the work of the codification of civil code, improve the legal system related to property rights, contracts and intellectual property rights, clarify the provisions of laws and regulations that violate the fairness of the law, and make equal protection the basic principle of regulating property relations.
LAW08		The term “ pharmaceutical trading enterprises ” refers to specialized enterprises or joint ventures that operate pharmaceuticals.		The people's procuratorate shall, according to the facts of the crime, the nature of the crime, the circumstances and the degree of harm to the society, put forward sentencing proposals in accordance with the Criminal Law, the Criminal Procedure Law and the relevant judicial interpretations.		To accelerate the promotion of the codification of civil codes, improve the legal systems relating to property rights, contracts, and intellectual property rights, clean up the provisions of laws and regulations that are in violation of fairness, and make equal protection a basic principle for regulating property relations.
LAW09		“ Pharmaceutical trading enterprises ” means specialized enterprises or joint ventures that engage in pharmaceutical products.		With regard to the facts of the crime, the nature of the crime, the circumstances and the degree of harm to society, the people's procuratorate shall put forward sentencing suggestions in accordance with the Criminal Law, the Criminal Procedure Law and the relevant judicial interpretations.		We shall accelerate the advancement of the compilation of civil law, improve the relevant legal systems relating to property rights, contracts and intellectual property rights, clean up the provisions of laws and regulations that are in violation of the law, and use equality protection as the basic principle for regulating property relations.
Google		A pharmaceutical business enterprise refers to an enterprise that specializes in or also operates pharmaceuticals.		The People's Procuratorate shall make sentencing recommendations based on the facts of the crime, the nature and circumstances of the crime and the degree of harm to society in accordance with the provisions of the Criminal Law, the Criminal Procedure Law and relevant judicial interpretations.		Accelerate the compilation of the Civil Code, improve the legal systems related to property rights, contracts, and intellectual property rights, clean up legal and regulatory provisions that violate fairness, and take equal protection as the basic principle for regulating property relations.
DeepL		A pharmaceutical business is a franchised or part-time business that deals in pharmaceuticals.		The people's procuratorate shall, on the basis of the facts of the crime, the nature of the crime, the circumstances and the degree of harm to society, make a recommendation on sentencing in accordance with the provisions of the Criminal Law, the Criminal Procedure Law and the relevant judicial interpretations.		Accelerating the codification of the Civil Code, improving the legal systems relating to property rights, contracts and intellectual property rights, clearing up legal and regulatory provisions that run counter to fairness, and making equal protection the basic principle governing property relations.
NiuTrans		A pharmaceutical trading enterprise refers to a franchised enterprise or concurrent enterprise that deals in pharmaceuticals.		The people's procuratorate shall, according to the facts, nature, circumstances and degree of harm to society of the crime, put forward sentencing suggestions in accordance with the provisions of the Criminal Law, the Criminal Procedure Law and relevant judicial interpretations.		Accelerate the compilation of the Civil Code, improve the legal systems related to property rights, contracts and intellectual property rights, clean up laws and regulations that violate fairness, and take equal protection as the basic principle for regulating property relations.
REF		Pharmaceutical trading enterprise means an enterprise exclusively or partly engaged in the trade of pharmaceuticals.		The people's procuratorates shall, based on the facts, nature, circumstances and degree of harm to society of a crime, provide sentencing recommendations in accordance with the Criminal Law, Criminal Procedure Law and relevant judicial interpretations.		The compilation of the civil code shall be advanced more quickly, the legal system relating to real rights, contract and property rights shall be improved, laws, regulations and provisions in contravention of fairness shall be reviewed, and equal protection shall be established as a basic principle of regulating property relations.

Table 6: Chinese-English NMT Examples

Based on the statistical results of the adaptabilities and the results of the English-Chinese bidirectional NMT experiment in the legal domain, the following findings can be drawn. (1) The BLEU value ranking of the English-Chinese bidirectional NMT model in the legal domain is fully consistent with the KLD and MMD numerical ordering of the corpora. This verifies that the KL divergence and maximum mean discrepancy can effectively measure the adaptability of the source domain corpus and the target domain corpus in adaptive NMT in a quantitative way. (2) Decremental learning is an effective domain adaptive strategy. In the process, the decremental learning of domain filtering can enhance the domain adaptability of the training corpus, and the decremental learning of quality filtering can improve the quality of domain-related translations of the training corpus. (3) The domain adaptation NMT algorithm only needs about a quarter of the original training steps and increase the BLEU values by over two points. This verifies that the decremental learning framework can efficiently optimize the performance of the NMT model. In summary, to train an NMT model for a specific domain, what matters are not the amount of the corpus but the high-quality domain-related corpus. Our decremental learning approach is a positive attempt for data engineering to convert readily available large-scale corpus of unknown quality to high-quality domain-related corpus.

6 Conclusion

This paper proposes a data centric approach focusing on the domain adaptation NMT problem, which gives full play to the large scale of parallel data of unknown quality and improves the domain translation quality through domain filtering and quality filtering. Finally, the effectiveness and efficiency of decremental learning is verified in the English-Chinese bidirectional NMT experiment in the legal domain.

Future research will mainly focus on domain knowledge modeling and NMT models based on domain knowledge intervention. We are going to construct an explicit multilingual domain knowledge graph, increase the neural computability of cross-language complex domain knowledge, and further improve the translation quality of domain adaptation NMT.

Acknowledgments

The research is supported by the Humanity and Social Science Research Project of Ministry of Education of China (No. 20YJAZH069, No. 20YJC740062), the New Liberal Arts Research and Reform Practice Project of Ministry of Education of China (No. 2021060049), the Postgraduate Education and Teaching Reform Research Project of Shandong (No. SDYJG21185), the Key Project of Undergraduate Teaching Reform Research of Shandong (No. Z2021323), the Key Project of the 14th Five-Year Plan of Educational Science of Shandong (No. 2023ZD021), and the 2024 Scientific Research and Innovation Platform Project of Xianda SISU (Language Data and AI Research, Xianda SISU).

References

Zhixing Tan, Shuo Wang, Zonghan Yang, Gang Chen, Xuancheng Huang, Maosong Sun, and Yang Liu. 2020. Neural Machine Translation: A Review of Methods, Resources, and Tools. *AI Open*, 1:5–21.

Chenhui Chu and Rui Wang. 2018. A Survey of Domain Adaptation for Neural Machine Translation. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1304–1319.

Tobias Domhan and Felix Hieber. 2017. Using Target-Side Monolingual Data for Neural Machine Translation through Multi-Task Learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1500–1505.

Denny Britz, Quoc Le, and Reid Pryzant. 2017. Effective Domain Mixing for Neural Machine Translation. In *Proceedings of the 2nd Conference on Machine Translation*, pages 118–126.

Catherine Kobus, Josep Crego, and Jean Senellart. 2017. Domain Control for Neural Machine Translation. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 372–378.

Boxing Chen, Colin Cherry, George Foster, and Samuel Larkin. 2017. Cost Weighting for Neural Machine Translation Domain Adaptation. In *Proceedings of the First Workshop on Neural Machine Translation*, pages 40–46.

Rui Wang, Masao Utiyama, Andrew Finch, Lemao Liu, Kehai Chen, Eiichiro Sumita. 2018. Sentence Selection and Weighting for Neural Machine Translation Domain Adaptation. *IEEE/ACM*

Transactions on Audio, Speech, and Language Processing, 26(10):1727–1741.

Adam Csaba Varga. 2017. Domain Adaptation for Multilingual Neural Machine Translation. Master Thesis, Saarlandes University.

Praveen Dakwale and Christof Monz. 2017. Fine-Tuning for Neural Machine Translation with Limited Degradation across in- and out-domain Data. In *Proceedings of the 16th Machine Translation Summit*, pages 156–169.

Zi-Yi Dou, Xinyi Wang, Junjie Hu, and Graham Neubig. 2019. Domain Differential Adaptation for Neural Machine Translation. In *Proceedings of the 3rd Workshop on Neural Generation and Translation*, pages 59–69.

Chenhui Chu, Raj Dabre, and Sadao Kurohashi. 2017. An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 385–391.

Antonio Valerio Miceli Barone, Barry Haddow, Ulrich Germann, and Rico Sennrich. 2017. Regularization Techniques for Fine-Tuning in Neural Machine Translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1489–1494.

Virginia Adams, Sandeep Subramanian, Mike Chrzanowski, Oleksii Hrinchuk, and Oleksii Kuchaiev. 2022. Finding the Right Recipe for Low Resource Domain Adaptation in Neural Machine Translation. arXiv preprint arXiv:2206.01137v1.

Markus Freitag and Yaser Al-Onaizan. 2016. Fast Domain Adaptation for Neural Machine Translation. arXiv preprint arXiv:1612.06897.

Huda Khayrallah, Gaurav Kumar, Kevin Duh, Matt Post, and Philipp Koehn. 2017. Neural Lattice Search for Domain Adaptation in Machine Translation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing*, pages 20–25.

Anna Currey, Antonio Valerio Miceli Barone, and Kenneth Heafield. 2017. Copied Monolingual Data Improves Low-Resource Neural Machine Translation. In *Proceedings of the Second Conference on Machine Translation*, pages 148–156.

Jiajun Zhang and Chengqing Zong. 2016. Bridging Neural Machine Translation and Bilingual Dictionaries. arXiv preprint arXiv:1610.07272v1.

Yong Cheng, Wei Xu, Zhongjun He, Wei He, Hua Wu, Maosong Sun, and Yang Liu. 2016. Semi-Supervised Learning for Neural Machine Translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 1965–1974.

Rui Wang, Andrew Finch, Masao Utiyama, and Eiichiro Sumita. 2017. Sentence Embedding for Neural Machine Translation Domain Adaptation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 560–566.

Marlies van der Wees, Arianna Bisazza, and Christof Monz. 2017. Dynamic Data Selection for Neural Machine Translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1400–1410.

Danielle Saunders. 2022. Domain Adaptation and Multi-Domain Adaptation for Neural Machine Translation: A Survey. arXiv preprint arXiv:2104.06951v2.

Junjie Hu, Mengzhou Xia, Graham Neubig, and Jaime Carbonell. 2019. Domain Adaptation of Neural Machine Translation by Lexicon Induction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2989–3001.

A. Tuan Nguyen, Toan Tran, Yarin Gal, Philip H. S. Torr, and Atılım Güneş Baydin. 2022. KL Guided Domain Adaptation. In *Proceedings of the 10th International Conference on Learning Representations*.

Wei Wang, Haojie Li, Zhengming Ding, and Zhihui Wang. 2020. Rethink Maximum Mean Discrepancy for Domain Adaptation. arXiv preprint arXiv:2007.00689v1.

Wuying Liu, Lin Wang, Mianzhu Yi, and Nan Xie. 2014. Active Multi-Field Learning for Spam Filtering. *Computing and Informatics*, 33(6):1400–1427.

Wuying Liu and Lin Wang. 2022. Ensemble Machine Translation to Filter Low Quality Corpus. In *Proceedings of the 26th International Conference on Asian Language Processing Proceedings*, pages 500–504.