Statistical Machine Translation adding Pattern-based Machine Translation in Chinese-English Translation

Jin'ichi Murakami, Masato Tokuhisa, Satoru Ikehara

Department of Information and Knowledge Engineering Faculty of Engineering
Tottori University, Japan
murakami@ike.tottori-u.ac.jp

Abstract

We have developed a two-stage machine translation (MT) system. The first stage is a rule-based machine translation system. The second stage is a normal statistical machine translation system. For Chinese-English machine translation, first, we used a Chinese-English rule-based MT, and we obtained "ENGLISH" sentences from Chinese sentences. Second, we used a standard statistical machine translation. This means that we translated "ENGLISH" to English machine translation. We believe this method has two advantages. One is that there are fewer unknown words. The other is that it produces structured or grammatically correct sentences.

From the results of experiments, we obtained a BLEU score of 0.3151 in the BTEC-CE task using our proposed method. In contrast, we obtained a BLEU score of 0.3311 in the BTEC-CE task using a standard method (moses). This means that our proposed method was not as effective for the BTEC-CE task. Therefore, we will try to improve the performance by optimizing parameters.

1. Introduction

Machine translation (MT) systems have been studied for a long time, and there are now three generations of this technology. The first generation was a rule-based translation method, and the second generation was an example-based machine translation method. Recently, the statistical machine translation method has become very popular. This method is based on statistics. Many versions of statistical machine translation models are available. An early model of statistical machine translation was based on IBM1 $\sim$ 5[1]. Recent statistical machine translation systems usually use phrase-based models.

However some problems arise with phrase-based statistical machine translation. One problem is as follows. Normally, a translation model requires a large parallel corpus. However, if we use a smaller parallel corpus, it results in many unknown words in the output translation. The second problem is that normally, an N-gram model is used as a language model. However, this model consists of local language information and does not have grammatical information.

To solve these problems, we have developed a two-stage machine translation system. The first stage is a rule-based machine translation system, and the second stage is a normal statistical machine translation system. This idea was based on paper[3],[4],[5].

In Chinese-English translation, the first stage consists of Chinese-English rule-based machine translation. In this stage, we obtained "ENGLISH" sentences from Chinese sentences. We aim to achieve "ENGLISH" sentences that contain few unknown words and that are generally grammatically correct. However, these "ENGLISH" sentences have low levels of fluency and naturalness because they were obtained using rule-based machine translation. In the second stage, we used a normal statistical machine translation system. This stage involves "ENGLISH" to English machine translation. With this stage, we aim to revise the outputs of the first stage improve the naturalness and fluency.

We used SYSTRAN V6 for the first stage. We used general statistical machine translation tools for the second stage, such as "Giza++"GIZA++, "moses" [7], and "training-phrase-model.perl" [10]. We used these data and these tools and participated in the BTEC-CE, Challenge-CE, and Challenge-EC contests at IWSLT2009.

As a result of experiments, the proposed method was not so effective for these tasks. The BLEU score was not as good compared to the standard moses. However, the score was not optimized, and our method was still very promising. Thus, we will continue to develop the method and try again in the future.

2. Concepts of our Statistical Machine Translatlion System

Our statistical machine translation consists of a two-stage translation system. The first stage is rule-based machine translation, and the second stage is statistical machine translation. We describe our system by dividing it into two processes, training and decoding. These processes are assumed to be Chinese-English translation.

2.1. Training

The training process is as follows.

1. Parallel Corpus

   We prepare a Chinese-English parallel corpus.

2. Rule-based Machine Translation

   We used a Chinese-English rule-based machine translation. Thus, we obtain ”ENGLISH” sentences from Chinese sentences. These ”ENGLISH” sentences are pairs of English sentences.

3. Make ”ENGLISH”-English phrase table

   We make an ”ENGLISH”-English phrase table using training-phrase-model.perl[10].

4. English N-gram model

   We make an N-gram model from English sentences using SRILM [6].

Fig. 1 shows the flow chart of the training process.

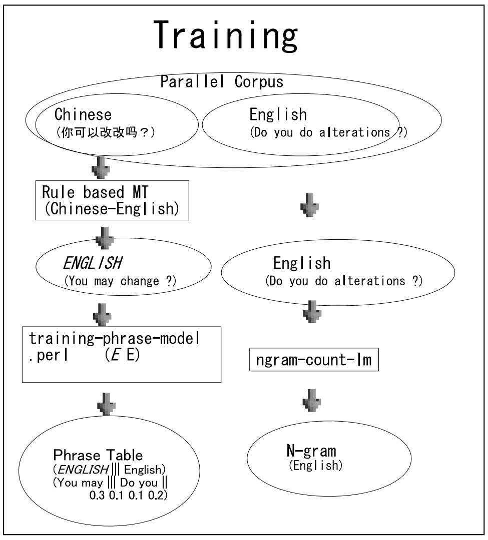

Figure 1: Flowchart of Training

2.2. Decoding

The decoding process is as follows.

1. Test Corpus

   We prepare the Chinese test sentences.

2. Rule-based Machine Translation

   We used a Chinese-English rule-based machine translation. Thus, we obtain ”ENGLISH” test sentences.

3. Statistical Machine Translation System

   Using phrase table in Section 2.1, N-gram model in Section 2.1, and moses[7], we decode the ”ENGLISH” sentences. This involves ”ENGLISH”-English translation. In this way, we obtain English sentences.

Fig.2 shows the flow chart of the decoding process.

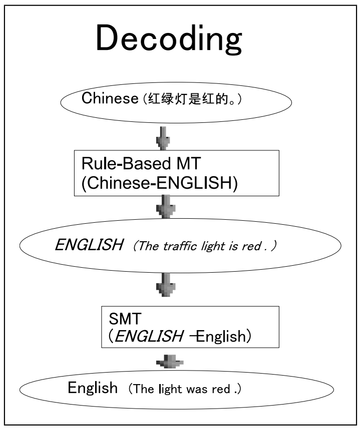

Figure 2: Flowchart of Decoding

3. Experiments with our Machine Translation

3.1. Training Data

Table 1 gives the examples of training data. We used the English punctuation procedure, which means that we changed ”,” and ”.” to ” ，” and ” ．”. Also, we did not handle English case forms.

Table 1: BTEC-CE training data

		Chinese
ZH	1	不用 担心 那个 。 我 要 它 不 需要 把 它 包 起来 。
ZH	2	可以 改改 ？
ZH	3	灯 是 的 。
ZH	4	我 想 要 靠 窗 的 子 。
ZH	5	在 那 就 在 游客 信息 的 前面 。
		English
EN	1	No worry about that . I’ll take it and you need not wrap it up .
EN	2	Do you do alterations ?
EN	3	The light was red .
EN	4	We want to have a table near the window .
EN	5	It’s over there , just in front of the tourist information .

3.2. First Stage

We used SYSTRAN V6 for the first stage. Table 2 lists examples of the first stage (SYSTRAN) output. In this page, the output is shown as ENGLISH.

Table 2: Output of First Stage

ZE	1	Does not need to worry that . I must buy its you not to need to wrap it .
ZE	2	You may change ?
ZE	3	The traffic light is red .
ZE	4	We want to open depend on the window the table .
ZE	5	In that side on in tourist information front .

3.3. "ENGLISH"-"English" Phrase Tables

For the second stage, we made an ENGLISH-English phrase table. To make this table, we used "train-phrase-model.perl[10]" in "training-release-1.3.tgz". We set parameters to default values. Table 3 lists examples of phrase tables for the second stage of our MT in the BTEC-CE task. This phrase table represents an "ENGLISH" "English" phrase table. As seen in this table, some English phrases are natural, although some of them are unnatural.

Table 3: Examples of phrase-tables (BTEC-CE)

Extremely appropriate . |||It fits very well . ||| 1 0.0037774 1 0.000165701
Extremely appropriate |||It fits very well ||| 1 0.00394828 1 0.000167943
Extremely attractive . |||It is very beautiful . |||1 0.00468009 0.5 0.000167226
Extremely attractive . |||Very beautiful . ||| 1 0.121764 0.5 0.0529012
Extremely attractive |||It is very beautiful ||| 1 0.00489181 0.5 0.000169488
Extremely attractive |||Very beautiful ||| 1 0.127273 0.5 0.053617
want to go to eat meal . |||like to have dinner . ||| 1 4.70488e-06 0.5 0.00340606
want to go to eat meal . |||want to go to the restaurant . ||| 1 1.02487e-05 0.5 4.7193e-06
want to go to eat meal |||like to have dinner ||| 0.333333 4.91772e-06 0.5 0.00345215
want to go to eat meal |||want to go to the restaurant ||| 1 1.07123e-05 0.5 4.78316e-06
want to go to eat |||like to have ||| 0.0222222 3.18012e-05 1 0.0191019
you eaten ? |||you tried ? ||| 1 0.0705182 1 0.0519143
you eaten |||you tried ||| 1 0.0714764 1 0.0524031

3.4. 5-gram Language Model

We calculated the 5-gram model using ngram-count in the Stanford Research Institute Language Model (SRILM) toolkit [6], and used " -ukndiscount -interpolate" as the smoothing parameter.

3.5. Decoder

We used "Moses[7]" as a decoder. In a Chinese to English translation, the position of the verb is sometimes significantly changed from its original position. Thus, we set the "distortion weight (weight-d)" to "0.2" and "distortion-limit" to "-1" for standard statistical machine translation (contrastive1 in Fig8). However, our system has 2 stage machine translation and the output of first stage is "ENGLISH". In this case, the position of word did not move so widely. So, we set the "distortion-limit" to "-6" for second stage statistical machine translation for our system (primary in Fig8).

Table 4 indicates the other parameters. We did not optimize these parameters nor use a reordering model.

Table 4: Parameters of moses.ini

ttable-limit	40	0			
weight-d	0.1				
weight-l	1.0				
weight-t	0.5	0.0	0.5	0.1	0.0
weight-w	-1				
distortion-limit	(-1 or 6)				

4. Results of our Machine Translation (IWSLT 2009 Automatic Evaluation Scores)

Table 8 summarizes the results of our machine translation evaluation for the BTEC-CE, Challenge-CE, and Challenge-EC tasks.

In this table, "primary" indicates our proposed system, "contrast1" indicates the normal statistical machine translation (moses), and "contrast2" indicates the outputs of SYSTRAN. Also, ASR.1 refers to the 1-BEST task, and CRR refers to the TEXT task.

Table 5 shows examples of our statistical machine translation for the BTEC-CE task . Table 6 shows examples of moses for the BTEC-CE task. Table 7 shows examples of SYSTRAN for the BTEC-CE task.

Table 5: Outputs for BTEC-CE primary (Proposed System)

01	in	救命 ！
	out	Save a life .
02	in	把 房 打 干 。
	out	Please remove clean my room .
03	in	是 旅行 支票 存根 。
	out	Is this the purchase stub traveler's checks .
04	in	得 再 慢 一点 。
	out	Please open slow another one .
05	in	我 想 和 史密斯 先生 。
	out	I'd like to converse to Mr . Smith .
06	in	有 套餐 的 菜 ？
	out	Do you have a menu set meal ?
07	in	ł 皮革 ？
	out	What kind of leather ?
08	in	是的 但是 我 不会 英 。
	out	Yes , but I can't speak English .
09	in	。 多少 ？
	out	Thank you . How much is it ?
10	in	日本 的 十 几 青少年 很 喜 玩 子 游机 。
	out	The game a few years old young people like an electronic games mechanical well .

Table 6: Outputs for BTEC-CE contrastive1 (Moses)

01	in	救命 ！
	out	Help .
02	in	把 房 打 干 。
	out	Please make up clean the room .
03	in	是 旅行 支票 存根 。
	out	Is this the stub purchase traveler's checks .
04	in	得 再 慢 一点 。
	out	Please drive more slowly .
05	in	我 想 和 史密斯 先生 。
	out	I'd like to talk to Mr . Smith , .
06	in	有 套餐 的 菜 ？
	out	Do you have set meals menu ?
07	in	ł 皮革 ？
	out	What kind of leather ?
08	in	是的 但是 我 不会 英 。
	out	Yes , but I can't speak English .
09	in	。 多少 ？
	out	Thank you . How much is it ?
10	in	日本 的 十 几 青少年 很 喜 玩 子 游机 。
	out	How many Japanese ten years old 青少年 electronic 游机 really like fun .

Table 7: Outputs for BTEC-CE contrastive2 (SYSTRAN)

01	in	救命 ！
	out	Save a life !
02	in	把 房 打 干 。
	out	Please clean the room cleanly .
03	in	是 旅行 支票 存根 。
	out	This is the traller's check purchase checkstub .
04	in	得 再 慢 一点 。
	out	Please open again slow one .
05	in	我 想 和 史密斯 先生 。
	out	I want to converse with Mr . Smith .
06	in	有 套餐 的 菜 ？
	out	Has the prix fixe the menu ?
07	in	ł 皮革 ？
	out	Which kind of leather ?
08	in	是的 但是 我 不会 英 。
	out	Is but I cannot speak English .
09	in	。 多少 ？
	out	Thanking . How much money ?
10	in	日本 的 十 几 青少年 很 喜 玩 子 游机 。
	out	Japan's several year old young people like playing the electronic mechanical games very much .

As seen in these results, our method was not so effective, as indicated by the BLEU score.

5. Discussion

With our system, we aim to reduce the number of unknown words and ungrammatical sentences. Thus, we analyze the outputs according to these factors.

5.1. Unknown Words

If we compare the outputs of moses and of our system, we find very few unknown words. Therefore, the proposed method is effective in reducing the number of unknown words.

5.2. Grammatical Correctness

We analyzed the outputs of our MT system. However, there were no native speakers of Chinese to check the inputs. Therefore, it was impossible to analyze these results and determine what was wrong. However, by comparing the output of moses and the output of our system, we found that our system produced more grammatically correct sentences. However, the BLEU score was not so good, so we need to find out why this occurred.

6. Conclusion

We have developed a two-stage machine translation system. The first stage is a rule-based machine translation system. The second stage is a statistical machine translation system. Our goal with this system was to obtain fewer unknown words and fewer ungrammatical sentences. However, the results we obtained in experiments were not so good.

We did not optimize parameters nor did we use a reordering model. In future experiments, we will try these techniques, which we expect will enable our system to perform better.

7. Appendix: Experiments with Parameter Tunings

We try to optimize parameters using MERT and use reordering models to improve these results.

Table9 shows the results of these experiments. As can be seen this table, proposed method was not so effective for BLEU score. However, it was effective for METEOR score. It means that proposed method was less unknown words.

8. Acknowledgements

We thank the students of Tottori University for their valuable help.

9. References

[1] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. "The machinematics of machine translation: Parameter estimation", Computational Linguistics, 19(2): pp. 263-311. (1993).

[2] Philipp Koehn, Franz J. Och, and Daniel Marcu. " Statistical phrase-based translation". In Marti Hearst and Mari Ostendorf, editors, HLT-NAACL 2003: Main Proceedings, pages 127.133, Edmonton, Alberta, Canada, May 27 -June 1. Association for Computational Linguistics. (2003).

[3] Pierre Isabelle, Cyril Goutte, and Michel Simard., "Domain Adaptation of MT systems through automatic post-editing", MT Summit XI, 102, 2007.

[4] Yushi Xu and Stephanie Seneff, "Two-Stage Translation: A Combined Linguistic and Statistical Machine Translation Framework", Proceedings of the Eighth Conference of the Association for Machine Translation (AMTA) 2008.

[5] Jason Katz-Brown and Michael Collins, " Syntactic Reordering in Preprocessing for Japanese English Translation: MIT System Description for NTCIR-7 Patent Translation Task", Proceedings of the 7th NTCIR Workshop Meeting, 2008.

[5] GIZA++, http://www.fjoch.com/GIZA++

[6] SRILM, The SRI Language Modeling Toolkit, http://www.speech.sri.com/projects/srilm

[7] Moses, moses.2007-05-29.tgz, http://www.statmt.org/moses/

[8] NIST Open Machine Translation, http://www.nist.gov/speech/tests/mt

[9] The METEOR Automatic Machine Translation Evaluation System, http://www.cs.cmu.edu/ alavie/METEOR/

[10] training-phrase-model.perl Evaluation System, http://www.cs.cmu.edu/ alavie/METEOR/

Table 8: Results

TASK	BTEC_CE									
case+punc	bleu	meteor	f1	prec	recl	wer	per	ter	gtm	nist
primary	0.3151	0.6169	0.6569	0.6465	0.6676	0.5590	0.4760	48.0710	0.6478	6.3834
contrastive1	0.3311	0.6109	0.6610	0.6758	0.6468	0.5377	0.4567	44.8140	0.6423	6.1511
contrastive2	0.1070	0.4697	0.5619	0.5671	0.5567	0.7017	0.6182	60.0070	0.4863	3.9727
TASK	CT_CE									
case+punc	bleu	meteor	f1	prec	recl	wer	per	ter	gtm	nist
primary.CRR	0.2797	0.5971	0.6306	0.6092	0.6536	0.6590	0.5099	61.3850	0.6592	5.5309
contrastive1.CRR	0.2706	0.5881	0.6189	0.5945	0.6453	0.6712	0.5113	62.4990	0.6533	5.4633
contrastive2.CRR	0.0642	0.3953	0.4928	0.5051	0.4811	0.8046	0.6823	74.9560	0.4312	3.2979
primary.ASR.1	0.2482	0.5489	0.5910	0.5773	0.6053	0.6943	0.5456	64.8360	0.6136	5.0705
contrastive1.ASR.1	0.2650	0.5610	0.6000	0.5876	0.6128	0.6647	0.5220	62.0140	0.6307	5.2804
contrastive2.ASR.1	0.0602	0.3654	0.4644	0.4822	0.4479	0.8148	0.7018	76.1960	0.4009	2.9995
TASK	CT_EC									
case+punc	bleu	meteor	f1	prec	recl	wer	per	ter	gtm	nist
primary.CRR	0.2759	0.5328	0.5500	0.5150	0.5900	0.7421	0.5382	68.6970	0.6914	5.3888
contrastive1.CRR	0.3391	0.5744	0.6204	0.6430	0.5994	0.5942	0.4356	52.3780	0.6930	6.1764
contrastive2.CRR	0.2300	0.5063	0.5596	0.5599	0.5594	0.6993	0.4987	63.2230	0.6304	5.4766
primary.ASR.1	0.2214	0.4417	0.4516	0.4100	0.5025	0.8518	0.6447	80.8210	0.6399	4.5091
contrastive1.ASR.1	0.2853	0.5134	0.5604	0.5784	0.5435	0.6609	0.4986	59.2510	0.6331	5.4212
contrastive2.ASR.1	0.1902	0.4483	0.4986	0.4948	0.5025	0.7627	0.5683	70.5120	0.5689	4.6699

primary: Proposed method contrastive1: Moses contrastive2: Systran

Table 9: Appendix: Results with Parameter Tunings

TASK	BTEC_CE									
case+punc	bleu	meteor	f1	prec	recl	wer	per	ter	gtm	nist
primary	0.3351	0.6256	0.6522	0.6301	0.6759	0.5704	0.4874	0.5048	0.6613	6.5972
contrastive1	0.3423	0.6135	0.6500	0.6463	0.6538	0.5436	0.4721	0.4674	0.6551	6.5624
contrastive2	0.107	0.4697	0.5619	0.5671	0.5567	0.7017	0.6182	60.007	0.4863	3.9727

primary: Proposed method contrastive1: Moses contrastive2: Systran