Analyzing Effects of Factored Translation Models in English to Japanese Statistical Machine Translation

Takahiro Oda Tomoyosi Akiba
Toyohashi University of Technology
1-1 Hibarigaoka, Tenpaku, Toyohashi, Aichi
441-8580, Japan
{oda, akiba}@cl.ics.tut.ac.jp

ABSTRACT

Factored translation model was proposed as an extended phrase-based statistical machine translation model. Effects of it were shown in many languages, however these were not shown in Japanese. We researched it in English to Japanese(EtoJ) and Japanese to English(JtoE) translation.

Keywords

Analyzing, Factored Translation

1. INTRODUCTION

The current state of art approach to statistical machine translation, so-called phrase-based model, is limited to the mapping of small text chunks without any explicit use of linguistic information. The factored translation model is an extension of it. Phrase-based model uses sequence of words as a phrase, while factored translation model uses sequence of factors as a phrase. A factor is different representation of word, such as surface, lemma and part-of-speech etc. Effects of factored translation models have shown in many languages, but it has not shown in Japanese. We applied factored translation model to EtoJ and JtoE translation, and researched effects of it by using standard factors.

It is important to select a combination of factors for training factored translation model, because performance of the model is influenced by selection of it. Many combinations are possible, while only a pieces of combinations may have good performance. We propose establishing effect of factors before training.

2. RELATED WORK

Kohen et al.[2] reported effects of factored translation models in English to German, English to Spanish and English to Czech translation. They reported that factored translation model has better performance in syntactically complex language, as using factors of syntactical information such as part-of-speech, lemma, morphological information etc. in addition to surface form.

3. FACTORED TRANSLATION MODEL

In this section, we will mention how to train and how to decode factored translation model.

3.1 Training

training step is divided into following three steps.

1. factorize training data
2. train translation model
3. train generation model

Before training translation model, we get factors of training data by factorizing it. the following example is factors that are obtained by factorizing "is connected to".

$$word \rightarrow surface|lemma|pos$$

$$\begin{aligned} is &\rightarrow is|be|VBZ \\ connected &\rightarrow connected|connect|VBN \\ to &\rightarrow to|TO|to \end{aligned}$$

The translation model generate phrase of target side language from phrase of source side. In factored model, the training of it is same as phrase-based model with exception training probability of phrases that are strings of factors. In phrase-based model, phrase is a string of words, while in factored model, it is a string of factors that are combinations of surface, lemma, and part-of-speech etc. . In figure 1, translation model generate target side surface phrase from source side phrase of factors that are combination of surface form and par-of-speech. In figure 2, translation model generate target side phrase of factors that combination of surface and part-of-speech from source side surface phrase. In figure 3, two translation models are used to generate target language. There are two ways that combine each phrase, one of these way is that phrase probabilities are calculated from "both" translation model, and the other is from "either" translation model. We experimented former way. In this case, probability of phrases are calculated on each translation model on same phrase boundary. And each probabilities of phrases are combined by log-linear model.

The generation model generate phrase of target side, from same side phrase that are consisted by other factors. This model is needed to generate surface, when translation model does not output surface (ex: output only lemma or part-of-speech). In this paper, generation model is not used, as all of our translation models output surface.

3.2 Decoding

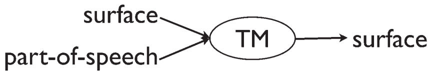

Figure 1: Translation model that receives factors

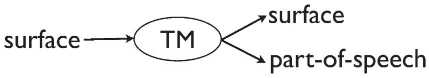

Figure 2: Translation model that outputs factors

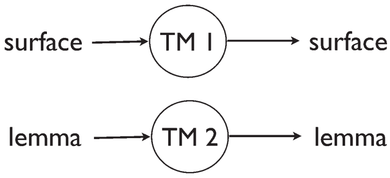

Figure 3: Multiple Translation Model

When translation model has been trained from factorized source language, such as figure 1, test data must be factorized to. On the other hand, when translation model has been trained from factorized target language, target phrases are evaluated some language models. In figure 2, target phrase are evaluated by using two language models (surface and part-of-speech).

4. EXPERIMENTS

In this section, we will obtain experiments that compare factored translation model with phrase-based model. We experimented in following two types of model,

- translation model that uses factorized source language (such as figure 1)
- translation model that uses factorized target language (such as figure 2)

in this paper, all combinations of factors include surface.

4.1 Settings

In experiments, we used following tools.

Table 1: tools for training and decoding

	tool
train and decode	Moses[3]
language model	SRILM[4]
factorize English	tree-tagger[5]
tokenize and lowercase English	wmt09[6]
factorize and tokenize Japanese	mecab[7]

Parameter of phrase alignment heuristic is grow-diag-final-and, and parameter of reordering model is msd-bidirectional-fe. All Language models are 5-gram. Part-of-speech in English is represented by Penn-Treebank tagset and in Japanese is represented by mecab basic tagset. But surface and lemma language models are smoothed by Knerser-Ney smoothing, while part-of-speech language model are are smoothed by Witten-Bell smoothing.

And dataset were used in table 2. For tuning, we used head 100 sentences of development set.

Table 2: data set

	data
training	2966044 sentences
tuning	100 sentence
test(EtoJ)	1119 sentences
test(JtoE)	1251 sentences

4.2 Result

Results of EtoJ translation are shown in table 3, we experimented with four factored translation models and phrase-based models, that is row that both factors are surface. Results of JtoE translation are shown in table 4, we experimented with four same models in EtoJ translation. In this experiments, no effects of factored translation models, as phrase-based model had best performance.

When we use multiple translation model, it is a lot of time to tuning parameter. We have experimented only surface & lemma model (such as figure 3). JtoE translation BLEU is 18.88.

Table 3: EtoJ translation's BLEU

English factors	Japanese factors	BLEU
surface	surface	28.50
surface&lemma	surface	27.82
surface&pos	surface	23.24
surface	surface&lemma	26.25
surface	surface&pos	15.96

5. ANALYSIS

Generating surface from combination of factors by translation model means that generate target sentence from richer source information, generating factors from surface means that richer target information from source sentence. We researched which is better, making English richer or Japanese. BLEU scores present it. In table 3, translation models that generate surface from factors have better performance than that generate factors, while in table 4, translation models that generate factors have better performance than that generate surface from factors. It means that making English is better.

In table 5 and table 6, number of target phrase per source phrase and number of phrases that are included in translation model are presented. In both directions of translation, translation models that use combination of factors that include part-of-speech have bad performance. Because these translation models have lower phrase table size, it is caused by low phrase table size. In table 7, number of word's differences in training data are presented. If combination of factors that are generated by translation model includes part-of-speech factor, evaluated by part-of-speech language

Table 5: Number of phrase on EtoJ translation

English factors	Japanese factors	$\frac{sourcephrase}{targetphrase}$	number of phrase
surface	surface	2.31	111003227
surface,lemma	surface	2.31	111178605
surface,pos	surface	2.31	8688120
surface	surface,lemma	2.32	111079910
surface	surface,pos	5.81	7277406

Table 6: Number of phrases on JtoE translation

Japanese factors	English factors	$\frac{sourcephrase}{targetphrase}$	number of phrases
surface	surface	2.40	111127158
surface,lemma	surface	2.40	111204399
surface,pos	surface	2.34	8625831
surface	surface,lemma	2.40	111301434
surface	surface,pos	2.36	8691818

Table 4: JtoE translation's BLEU

Japanese factors	English factors	BLEU
surface	surface	26.27
surface&lemma	surface	21.47
surface&pos	surface	16.45
surface	surface&lemma	22.66
surface	surface&pos	22.45

Table 7: Number of word's types

factor	japanese	English
surface	265265	258820
pos	14	46
lemma	260274	223893
surface & pos	268582	316811
surface & lemma	266510	260694

model. As number of part-of-speech factor's differences are very low, perplexities of part-of-speech language models are high. But it is not cause of low BLEU score, as absolute weight of the part-of-speech language model are low.

On both Japanese and English, we calculated importance of part-of-speech language model that is absolute weight of the part-on-speech language model per absolute weight the surface language model, such defined by 1. As it is higher on Japanese, Japanese part-of-speech is more important for performance.

$$importance\_of\_factor\_LM = \frac{absolute\_weight\_factorLM}{absolute\_weight\_surfaceLM} \quad (1)$$

Between translation model that outputs Japanese lemma and that outputs English lemma, latter lemma is more important for performance, as the latter model has higher absolute weight of the lemma language model.

In table 8, translation result of JtoE. Factor is translation model that generate surface and lemma from surface, which is best performance in factored model. While unfactor is phrase-based model. Factor has better performace for Sentences on higher row, as these are sorted by factor BLEU - unfactor BLEU. Factor's outputs that are low BLEU score are tend to be same context as unfactor's outputs, while unfactor outputs that are low BLEU score are tend to be different context from both factor outputs and references. Factor's outputs may be better result for human, but BLEU scores don't reflect it.

6. CONCLUSION & FEATURE WORK

We researched effect of factored translation model on EtoJ translation and JtoE translation. While it is many times as long as decoding of phrase based translation to decode factored translation model, there are no effect both direction of translation. But making English's information richer is better way than making Japanese's information richer.

In the feature work, we will research effect of multiple translation models that are used by "either" way. And we will resaerch effect of factor except that we used in this paper.

7. REFERENCES

[1] Atsushi Fujii, Masao Utiyama, Mikio Yamamoto, Takehito Utsuro, Terumasa Ehara, Hiroshi Echizen-ya, Sayori Shimohata, "Overview of the Patent Translation Task at the NTCIR-8 Workshop", Proceedings of the 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-lingual Information Access, 2010.

[2] P.Koehn and H.Hoang, "Factored Translation Models", In Proceeding of Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 868-876, 2007.

[3] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin and E. Herbst, "Moses:Open source toolkit for statistical machine translation", In Proceeding of ACL 2007 demonstration session, pp.177-180, 2007. http://www.statmt.org/moses/

Table 8: sentences that has high line BLEU and low line BLEU

factor BLEU - unfactor BLEU		sentence
60.6	unfactor	a motor driver 80 , 81 and 82 edu ecu .
	factor	the motor driving device 80 , the ecu 81 and the edu 82 .
	ref	the motor operating device 80 includes the ecu 81 and the edu 82 .
58.7	unfactor	illator , there is the following problem .
	factor	however , the quartz oscillator patent document 1 has the following problems .
	ref	however , the crystal oscillator of the patent document 1 has the following problems .
56.8	unfactor	both ends of the shaft 44 of the first , the second bushing bush 35 and 36 .
	factor	the opposite ends of the shaft 44 , the first bushing 35 and the second bushing 36 is provided .
	ref	the first bushing 35 and the second bushing 36 are arranged on opposite ends of the rotary shaft 44 .
...		
-38.8	unfactor	the detecting an inspection recipe parameters are stored for each candidate .
	factor	the detected parameters are stored in the candidate for an inspection recipe .
	ref	the detection parameters are stored for each recipe candidate .
-42.1	unfactor	in this embodiment , the battery current detector 18-1 is provided will be described .
	factor	in this embodiment , the battery current detection unit 181 is provided will be described .
	ref	in this embodiment-1 , the case where the battery current detector 18-1 is provided is explained .
-45.2	unfactor	further , the coupling coefficient has a maximum value at a certain film thickness .
	factor	further , the coupling coefficient is at the maximum value of the film thickness .
	ref	also , the coupling coefficient shows a maximum value at a certain film thickness .

[4] A. Stolcke, "SRILM - an extensible language modeling toolkit", ICSLP, pp.901-904, 2002.

[5] H. Schmid, "Probabilistic Part-of-Speech Tagging Using Decision Trees", InProceeding of ECML-98, pp.25-36, 1998 .

[6] "EACL 2009 FOURTH WORKSHOP ON STATISTICAL MACHINE TRANSLATION", http://www.statmt.org/wmt09/

[7] Taku Kudo, "MeCab: Yet Another Part-of-Speech and Morphological Analyzer", http://mecab.sourceforge.net/