論文 / 著書情報
Article / Book Information

論題(和文)	
Title(English)	Development of a Speech Recognition System Using Machine Translated Data
著者(和文)	Arnar Thor Jensson, 岩野 公司, 古井 貞熙
Authors(English)	Arnar Thor Jensson, Koji Iwano, Sadaoki Furui
出典(和文)	日本音響学会 2007年秋季講演論文集, Vol. , No. 2-8-7, pp. 123-124
Citation(English)	, Vol. , No. 2-8-7, pp. 123-124
発行日 / Pub. date	2007, 9

Development of a Speech Recognition System Using Machine Translated Data *

◎ Arnar Thor Jensson, Koji Iwano, Sadaoki Furui
(Tokyo Institute of Technology)

1 Introduction

Statistical language modeling is well known to be very important in large vocabulary speech recognition but creating a robust language model (LM) typically requires a large amount of training text. Therefore it is difficult to create a statistical LM for resource deficient languages. However, using text translated from other languages may possibly improve the resource deficient LM either using sentence-by-sentence (SBS) translation or word-by-word (WBW) translation. WBW translation only requires a dictionary whereas SBS machine translation (MT) needs either a large sentence-aligned parallel corpus to train the MT system, or extensive rule based system, which both are expensive to obtain. The WBW approach is expected to be successful only for closely related languages. Methods have been proposed in the literature to improve statistical language modeling in a resource-deficient language using cross-lingual information retrieval [1]. Another method proposes using latent semantic analysis for cross-lingual modeling which does not require a sentence-aligned corpus [2] but searches for similar types of texts in two languages. LM adaptation with target task machine-translated text is addressed in [3] but without speech recognition experiments. In [4], we proposed a method to improve the LM built on a task-dependent corpus using MT which is similar to [3]. This paper extends our machine translation experiments. The dictionary used to translate WBW is now created automatically by an MT system. This paper also introduces SBS machine translated texts from English to Icelandic.

2 Adaptation Method

Our method involves adapting a task dependent LM that is created from a sparse amount of text and a large translated text (*TRT*), where *TRT* denotes the machine translation of the rich corpus (*RT*), preferably in the same domain area as the task. This involves two steps shown graphically in Figure 1. First of all a sparse text in the target language is split into two, a training text corpus (*ST*) and a development text corpus (*SD*). A language model LM1 is created from *ST*, and LM2 from *TRT*. The *TRT* can either be obtained from SBS or WBW translation. The *SD* set is used to optimize the weight used in Step 2. Step 2 involves interpolating LM1 and LM2 linearly. The final perplexity or word error rate (WER) value is calculated using an evaluation text set or speech evaluation set (*Eval*) which is disjoint from all other data sets.

3 Experimental Work

3.1 Experimental Data

The weather information domain was chosen for the Icelandic experiments and translation from English (*rich*) to Icelandic (*sparse*) using WBW and SBS. For the experiments the Jupiter corpus [5] was used. It consists of 67116 unique sentences gathered from actual users' utterances. A set of 2460 sentences were manually translated from English to Icelandic and split into 1500 *ST* sentences, 300 *SD* sentences and 660 *Eval* sentences. 63116 sentences were used as *RT*.

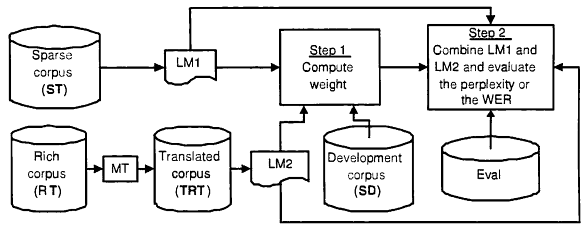

Fig. 1 Data diagram

A unique word list was made out of the Jupiter corpus and machine translated using [6] in order to create a dictionary. This MT is a rule based system. The dictionary was then used to translate *RT* into TRT_{WBW}. Another translation TRT_{SBS} was created by SBS machine translation using [6]. Names of places were identified and then replaced randomly with Icelandic place names for both TRT_{WBW} and TRT_{SBS}, since the task is in the weather information domain.

A 1-gram and 2-gram translation evaluation using BLEU [7] was performed on 20 sentences created from both the SBS and the WBW machine translators, using two human references. The 1-gram and 2-gram BLEU evaluation was 0.47 and 0.23 for WBW MT respectively. The 1-gram and 2-gram BLEU evaluation was 0.61 and 0.43 for SBS MT respectively.

A phonetically balanced (PB) Icelandic text corpus, the Jensson PB corpus [4], was used to create an acoustic training corpus. The training corpus consists of 3.8 hours of speech from 13 male and 7 famale speakers. An evaluation corpus was recorded using sentences from the *Eval* set. 2 hours of read speech was recorded from 10 male and 10 famale speakers. None of the speakers in the evaluation speech corpus are in the acoustic training corpus.

3.2 Experimental Setup

In total five different experiments were performed. The experimental setup can be viewed in Table 1. The *ST* set size varied from 100 to 1500 sentences for all the experiments. In the following text ST^n corresponds to a subset of the *ST* set where n is the number of sentences used. All LMs were built using 3-grams with Kneser-Ney smoothing. The WER experiments were performed three times with different, randomly chosen sentences, creating each *ST* and *SD* set, in order to increase the reliability of the results. An average WER was calculated over the three experiments. This increases reliability when comparing different experiments especially when the *ST* set is very sparse.

*機械翻訳データを用いた音声認識システムの開発
アーナー ジェンソン，岩野公司，古井貞熙（東工大）

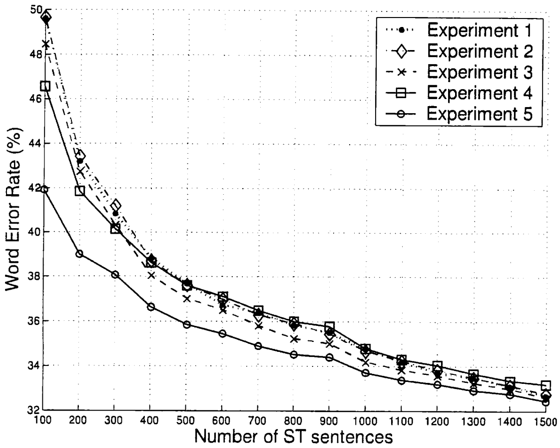

Fig. 2 Word error rate results.

3.3 Results

The WER results are shown in Figure 2. Perplexity and out-of-vocabulary (*OOV*) results are shown in Table 2 and Table 3 respectively for some *ST* values. The perplexity results for Experiment 1 to 3 should be compared together since the vocabulary, V_{ST}, is the same for those experiments. As shown in Table 2 all perplexity results get improved when a *TRT* corpus is introduced and interpolated with the corresponding *ST* set. The *OOV* rate shown in Table 3 is reduced by adding the unique words found in the *TRT* set to V_{ST} as expected.

4 Discussion

The improvement of the Icelandic LM with translated English text/data was confirmed by reduction in WER by using either WBW or SBS MT. Experiment 1 (*baseline*), should be compared with the other experiments since Experiment 1 does not assume any foreign machine translation. When the *baseline* is compared with the interpolated results using WBW MT in Experiment 4, we get a WER 49.6% reduced to 46.6% respectfully, a 6.0% relative improvement when using 100 *ST* sentences. The relative improvement reduces as more *ST* sentences are added to the system and converges to the *baseline* when 500 *ST* sentences are added to the system.

When the *baseline* is compared with the interpolated results using SBS MT in Experiment 5, we get a WER 49.6% reduced to 41.9% respectfully, a 15.5% relative improvement when 100 *ST* sentences are added to the system. The WER improvement of the SBS MT over the WBW MT is confirmed for all the *ST* sets as the BLEU evaluation results in Section 3.1 suggests. This can be seen by comparing Experiment 4 with Experiment 5 in Figure 2. The improvement is as well confirmed with perplexity results when Experiment 2 and Experiment 3 are compared in Table 2.

5 Conclusions

The results presented in this paper show that an LM can be improved considerably using either WBW or SBS translation. The WBW translation is especially important for resource deficient languages that do not have SBS machine translation tools available. Future work involves applying the WBW and SBS translation methods to a larger domain such as broadcast news.

Table 1 *Experimental Setup.*

Experiment nr.	*TRT* Corpus	Vocabulary
Experiment 1	None	V_{ST}
Experiment 2	TRT_{WBW}	V_{ST}
Experiment 3	TRT_{SBS}	V_{ST}
Experiment 4	TRT_{WBW}	$V_{ST} + V_{TRT_{WBW}}$
Experiment 5	TRT_{SBS}	$V_{ST} + V_{TRT_{SBS}}$

Table 2 *Perplexity results*

	ST^n			
Experiment nr.	ST^{100}	ST^{500}	ST^{1000}	ST^{1500}
Experiment 1	30.7	26.4	26.3	26.5
Experiment 2	29.4	26.0	26.1	26.3
Experiment 3	26.6	25.3	25.3	25.4

Table 3 *OOV results*

	ST^n			
Vocabulary	ST^{100}	ST^{500}	ST^{1000}	ST^{1500}
V_{ST^n}	14.0	6.8	5.5	4.6
$V_{ST^n} + V_{TRT_{WBW}}$	8.4	4.8	4.0	3.4
$V_{ST^n} + V_{TRT_{SBS}}$	4.4	2.6	2.5	2.2

6 Acknowledgements

We would like to thank Drs. J. Glass and T. Hazen at MIT and all the others who have worked on developing the Jupiter system. We also would like to thank Dr. Edward W. D. Whittaker for his valuable input. Special thanks to Stefan Briem for his English to Icelandic machine translation tool and allowing us to use his machine translation results. This work is supported in part by 21st Century COE Large-Scale Knowledge Resources Program.

References

[1] Khudanpur, S. and Kim, W., "Using Cross-Language Cues for Story-Specific Language Modeling", *Proc. ICSLP*, Denver, CO, vol 1, pp. 513-516, 2002.

[2] Kim, W. and Khudanpur, S., "Cross-Lingual Latent Semantic Analysis for Language Modeling", *Proc. ICASSP*, Montreal, Canada, vol 1, pp. 257-260, 2004.

[3] Nakajima, H., Yamamoto, H., Watanabe, T., "Language Model Adaptation with Additional Text Generated by Machine Translation", *Proc. COLING*, vol 2, pp. 716-722, 2002.

[4] Jensson, A., Iwano, K., Furui, S., "Development of a speech recognition system using a sparse training corpus", *Proc. LKR*, Tokyo, Japan, pp.133-136, 2007

[5] Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T. and Hetherington, L., "JUPITER: A Telephone-Based Conversational Interface for Weather Information", *IEEE Trans. on Speech and Audio Processing*, 8(1):100-112, 2000.

[6] Briem, S., "Machine translation tool for automatic translation from English to Icelandic", *http://www.simnet.is/stbr/*, Iceland, 2007.

[7] Papineni, K., Roukos, S., Ward T. and Zhu W., "BLEU: a Method for Automatic Evaluation of Machine Translation", *Proc. ACL*, Philadelphia, PA, pp. 311-318, 2002.