

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

A Manually-editing Tool of Translation Templates Based on Automatic Evaluation

¹Chun-Xiang Zhang, ²Xue-Yao Gao, ³Kun Liu, ⁴Zhi-Mao Lu, ⁵Da-Song Sun, ⁶Hong-Li Lv and ⁷Yong Liu

¹School of Software, Harbin University of Science and Technology, Harbin 150080, China

²College of Computer Science and Technology, Harbin University of Science and Technology,
Harbin 150080, China

³Material Accommodation Center of Petrochina Da Qing Petrochemical Company, Daqing 163714, China

⁴College of Information and Communication Engineering, Harbin Engineering University,
Harbin 150001, China

⁵Computer Center, Harbin University of Science and Technology, Harbin 150080, China

⁶Da Qing Chemical Research Center of Petrochina Petrochemical Research Institute,
Daqing 163714, China

⁷School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China

Abstract: Translation templates were very important translation knowledge in transfer-based MT system. They could be learned automatically from bilingual corpus and were written manually by MT knowledge engineers. But there were lots of noises in template base which would decrease the translation quality. This paper presented a manually-editing tool that provided edition functions including automatic evaluation of templates, edition of templates, insertion of templates and deletion of templates. In this tool, MT knowledge engineers could check and edit translation templates according to automatic evaluation scores. This tool aimed at helping to optimize template base in MT system which would eliminate the conflicts between translation templates and the errors in translation templates. Experimental results indicated that after MT knowledge engineers edited translation templates in this tool, translation assessment score of open test corpus improved 3.85% under 3 g bleu metric.

Key words: Translation templates, MT system, MT knowledge engineers, translation quality, translation assessment score

INTRODUCTION

Expansion of template base is very essential for a transfer-based MT system to improve its translation quality (Yamada *et al.*, 2002). At the same time, translation templates are often applied to many language processing tasks, such as question classification (Peng and Kai-Hui, 2011), information retrieval (Bedi and Chawla, 2010; Rajan and Rajagopalan, 2008) and summary (Mala and Geetha, 2008). Translation templates are often written by knowledge engineers. At the same time, they can also be acquired from bilingual sentence pairs (Zhang *et al.*, 2010). For example, error-driven learning method is also applied to get translation templates from bilingual corpus. But with new translation templates added into template base, the output quality in MT system cannot improve consistently because of redundancy and conflicts between templates. Although redundancy and conflicts can be eliminated in form, they

still exist at a deep layer, when the template base is applied and different templates work together. There are lots of noises in template base. Recent progress in natural language processing such as part-of-speech tagging, syntactic parsing, translation model, automatic translation evaluation, has made it real to develop the tool of checking translation templates semi-automatically now. In order to obtain template base with the best performance, many methods have been presented recently to evaluate translation templates. The frequency has been applied to evaluate translation templates (Menezes and Richardson, 2001). If the frequency of a translation template is low, it will occur in little language phenomenon and it should be deleted from template base. But this method only improves the efficiency of MT system rather than the translation quality. Lavoie uses log-likelihood ratio to evaluate translation templates (Lavoie *et al.*, 2001). Another current topic of machine translation is automatic evaluation of MT quality (Yasuda *et al.*, 2001). These

methods aim to replace subjective evaluation in order to speed up the development cycle of MT systems (Akiba *et al.*, 2001). However, they can be utilized not only as developers' aids but also for tuning MT systems semi-automatically and automatically (Su *et al.*, 1992). With the development of translation assessment, bleu score is presented to facilitate automatic MT evaluation (Papineni *et al.*, 2002). Nist score is also applied to automatic MT evaluation (Dodington, 2002). Automatic evaluation metric is used as feedback to evaluate translation templates and those that contribute adversely to translation quality will be deleted (Imamura *et al.*, 2003). An improved evaluation method is proposed to score translation templates (Chun-Xiang *et al.*, 2009).

In this study, a manually-editing tool is developed in order to facilitate MT knowledge engineers to check and edit translation templates. In this tool, the template's contribution to the translation of training corpus is computed under 2 g bleu metric. After ten MT knowledge engineers spend 100 h to edit translation templates in this tool, the translation quality of open test corpus is improved.

BLEU EVALUATION SCORE

The bleu score measures the similarity between MT results and reference translation results made by human beings. This similarity is measured by N-gram precision scores. Several kinds of N-grams can be used in bleu evaluation score. One gram precision score indicates the adequacy of word translation and longer N-gram precision scores indicate the fluency of sentence translation. The bleu score is calculated from the product of N-gram precision scores. So this measure combines adequacy and fluency.

Papineni *et al.* (2002) calculates their modified precision score p_n , for each N-gram length by summing over the matches for every hypothesis sentence S in translations of test corpus C. Because the computation of bleu evaluation score is based on precision and because recall is difficult to formulate over multiple reference translations, a brevity penalty is introduced to compensate for the possibility of proposing high precision hypothesis translations which are too short. So, the brevity penalty BP is calculated in Eq. 1.

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{1-r/c} & \text{if } c \leq r \end{cases} \quad (1)$$

Here, c is the length of the corpus of hypothesis translations and r is the effective length of reference corpus.

Thus, the computation of bleu evaluation score is shown as Eq. 2 describes.

$$Bleu = BP * \exp\left(\sum_{n=1}^N w_n \log p_n\right) \quad (2)$$

Here, w_n is the uniform weight. In this paper, the translation assessment score of training corpus is computed under 2 g bleu metric because its speed is very fast. We use $N = 2$ and uniform weights $w_n = 1/N$. The bleu score can range from 0 to 1, where higher scores indicate closer matches to the reference translations and where a score of 1 is assigned to a hypothesis translation which exactly matches one of the reference translations.

EDIT TRANSLATION TEMPLATES MANUALLY BASED ON BLEU EVALUATION SCORE

In order to evaluate a translation template, Chinese-English bilingual sentence pairs are selected from target field. Chinese sentences are collected as training corpus which is denoted as $T = \{CS_1, CS_2, \dots, CS_n\}$. English sentences are collected as reference translation corpus which is denoted as $R = \{ES_1, ES_2, \dots, ES_n\}$. Training corpus is corresponded with reference translation corpus sentence by sentence. ScoreTable is used to record current evaluation score of training corpus's machine translations. ScoreTable contains translation evaluation score of every Chinese sentence in training corpus T. So, ScoreTable is corresponded with training corpus T sentence by sentence. In template base which is denoted as TB, there are lots of translation templates to be evaluated. The architecture of manually-editing tool is shown in Fig. 1.

Template optimization module is used to evaluate every translation template in TB. The evaluation score will be shown in manually-editing interface, according to which MT knowledge engineers can decide whether this translation template will improve the translation quality of training corpus T or not. When the size of training corpus T is very large, we consider that training corpus T can cover enough more language phenomenon. The process of evaluating translation templates is reasonable which will provide more information to help MT knowledge engineers finishing this edition task. In evaluation of translation templates, MT engine will be used to translate Chinese sentences in training corpus T.

The process of editing translation templates semi-automatically in this tool is shown as follows:

- In manually-editing interface, MT knowledge engineers select translation template r from template base TB for evaluation

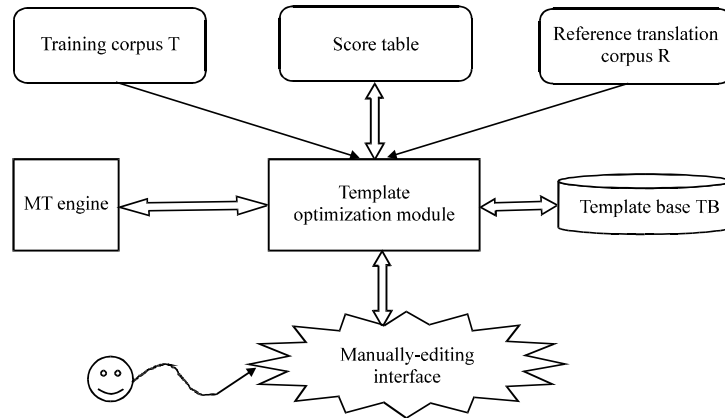


Fig. 1: Architecture for manually-editing tool

- Template optimization module removes r from template base TB and TB_r is gotten. We can know that $TB = TB_r + \{r\}$. Then template base TB_r is used to initialize MT engine
- Template optimization module uses MT Engine to translate training corpus T sentence by sentence. Then, MT results of training corpus T are gotten. Based on reference translation corpus R , template optimization module scores MT result of every Chinese sentence in training corpus T automatically under 2 g bleu metric. Based on ScoreTable, bleu score increase for MT result of every Chinese sentence in training corpus T is computed. Then, bleu score increases for MT results of all Chinese sentences in training corpus T are summed which is denoted as sum_bleu_score . The sum_bleu_score is calculated in Eq. 3. Here, $bleu_score_{new}(CS_i)$ is the translation assessment score of Chinese sentence CS_i , where MT engine uses template base TB_r to translate CS_i . The $bleu_score_{old}(CS_i)$ is the translation assessment score of Chinese sentence CS_i , where MT engine uses template base TB to translate CS_i . The $bleu_score_{old}(CS_i)$ ($I = 1, 2, \dots, n$) are recorded in ScoreTable

$$sum_bleu_score = \sum_{i=1}^n bleu_score_{new}(CS_i) - \sum_{i=1}^n bleu_score_{old}(CS_i) \quad (3)$$

- If sum_bleu_score is positive, it denotes that when r is removed from template base TB , MT quality of training corpus T will increase. Then, a prompt message that how many sentences are translated correctly and how many sentences are translated

incorrectly will be shown in manually-editing interface. At the same time, the increase of translation assessment score of training corpus T will be also shown in this prompt message. If MT knowledge engineers agree to remove r from template base, TB is set to TB_r and ScoreTable is updated with bleu score of MT result of every Chinese sentence in training corpus T . If MT knowledge engineers think that there are some errors in translation template r and r can cover some language phenomenon, they will correct condition part and translation part of r according to their language knowledge. New version of r is got which is denoted as r' . Template r should be deleted from template base and r' will be added into template base. TB is set to $TB_r + \{r'\}$

- If sum_bleu_score is negative, it denotes that when r is removed from template base TB , MT quality of training corpus will decrease. Then a prompt message that how many sentences are translated correctly and how many sentences are translated incorrectly will be shown in manually-editing interface. At the same time, the decrease of translation assessment score of training corpus T will be also shown in this prompt message. If MT knowledge engineers think that there are some errors in translation template r and r can cover some language phenomenon, they will correct condition part and translation part of r according to their language knowledge. New version of r is got which is denoted as r' . Template r should be deleted from template base and r' will be added into template base. TB is set to $TB_r + \{r'\}$. Otherwise, TB is set to TB_r and ScoreTable is updated with bleu score of MT result of every Chinese sentence in training corpus T

In this manually-editing tool, we allow MT knowledge engineers to decide whether *r* should be removed from template base TB or not. Although *sum_blue_score* is positive for a translation template, MT knowledge engineers think that this translation template can cover some language phenomenon and it will be kept in template base. After they correct condition part and translation part of *r* according to their language knowledge, *r* will be kept in template base TB. Although *sum_blue_score* is negative for a translation template, MT knowledge engineers think that syntax of this translation template is incorrect. So, *r* should be removed from template base. The manually-editing tool will provide more choices to select and correct translation templates for improving the quality of MT output.

EXPERIMENT

From Fig. 1, we can see that different manually-editing tools of translation templates for different transfer-based MT systems will be gotten, if we change the MT engine. Here, we use MTS2005 as MT engine in this study. MTS2005 is a transfer-based MT system which is developed by MOE-MS Key Laboratory of Natural Language Processing and Speech in Harbin Institute of Technology. There are 5480 translation templates in its template base TB. Sixty thousands of Chinese-English bilingual sentence pairs are selected from traveling fields. Chinese sentences are collected as training corpus T and English sentences are collected as reference translation corpus R. Training corpus T is corresponded with reference translation corpus R sentence by sentence.

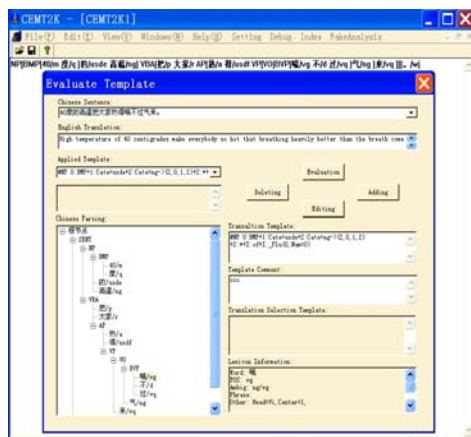


Fig. 2: Interface of manually-editing tool

Visual C++6.0 is used to develop the manually-editing tool of translation templates and its interface. The interface of this manually-editing tool is shown in Fig. 2.

In its interface, we can find Chinese sentence, parsing tree of Chinese sentence, its English translation and the template which is used in the process of translation. At the same time, we can get the lexicon information of the selected Chinese word in Chinese parsing tree.

For example, in the case of the following Chinese fake-parsing results, the process of editing translation templates is shown as follows:

- **Chinese fake-parsing results:** NP[BMP [40/m 度/q] μ的/ usde 高温/ng] VBA[把/p 大家/r AP[热/a 得/usdf VP [VO[BVP[喘/vg 不/d 过/vq]气/ng]来/vq]]]°/wj
- **Chinese sentence:** 4 0 度的高温把大家热得喘不过气来。
- **English translation:** High temperature of 40°C make everybody so hot that breathing heavily better than the breath come
- **Applied template:** # N P 0:BMP+1:Cate=usde+2:Cate=ng->(2,0,1,2)+2:*+I: of+I:_Plu(0,Num=0)

There are evaluation button, adding button, deleting button and editing button in this interface. MT knowledge engineers can click the evaluation button to score the applied template *r*. If *r* is deleted from template base, MT knowledge engineers should click the deleting button. Then *r* will be removed from template base. If *r* is kept in template base, MT knowledge engineers should click the adding button. If *r* is corrected, MT knowledge engineers should edit *r* and click the editing button. The old version of *r* will be removed from template base and the new version of *r* will be added into template base.

MT knowledge engineers use 7000 Chinese sentences with fake-parsing results to check translation templates in this manually-editing tool. Ten MT knowledge engineers spend 100 h to finish this task. After MT knowledge engineers check and edit these translation templates semi-automatically in this tool, the size of template base becomes 4015 which is denoted as TB_{edition}.

We also select another 2100 Chinese-English bilingual sentence pairs from traveling field for open test purpose. Chinese sentences are collected as open test corpus and English sentences are collected as reference translations of open test corpus. Two groups of experiments are conducted. In experiment 1, TB is used to translate open test corpus in MTS2005. In Experiment 2, TB_{edition} is used

Table 1: The performance for two groups of experiments

Experiment	bleu3
1	0.1143
2	0.1187

to translate open test corpus in MTS2005. The translations are evaluated under 3-gram bleu metric and experimental results are described in Table 1.

As shown in Table 1, we can find that the performance of TB_{edition} is better than that of TB. The translation assessment score of open test corpus improves 3.85% under 3-gram bleu metric.

CONCLUSIONS

In this study, bleu evaluation metric is applied to develop a manually-editing tool. This tool can help MT knowledge engineers to optimize template base semi-automatically. After MT knowledge engineers use this tool to check and edit translation templates manually, the translation quality of MT system is improved.

ACKNOWLEDGMENTS

This study is supported by Science and Technology Research Funds of Education Department in Helongjiang Province under Grant No. 11541045.

REFERENCES

Akiba, Y., K. Imamura and E. Sumita, 2001. Using multiple edit distances to automatically rank machine translation output. Proceedings of the Machine Translation in the Information Age, Sept. 18-22, Spain, pp: 15-20.

Bedi, P. and S. Chawla, 2010. Agent based information retrieval system using information scent. J. Artif. Intell., 3: 220-238.

Chun-Xiang, Z., L. Sheng, Y. Mu-Yung, Z. Tie-Jun and S. Xiao-Sheng, 2009. A method of rule-base optimization based on evaluation. J. Harbin Inst. Technol., 16: 708-712.

Doddington, G., 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. Proceedings of the 2nd International Conference on Human Language Technology Research, March 24-27, San Diego, California, pp: 138-145.

Imamura, K., E. Sumita and Y. Matsumoto, 2003. Feedback cleaning of machine translation rules using automatic evaluation. Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, July 7-12, Sapporo, Japan, pp: 447-454.

Lavoie, B., M. White and T. Korelsky, 2001. Inducing lexico-structural transfer rules from parsed Bi-texts. Proc. Workshop Data-driven Methods Mach. Trans., 14: 1-8.

Mala, T. and T.V. Geetha, 2008. Story summary visualizer using L systems. J. Artificial Intell., 1: 53-60.

Menezes, A. and S.D. Richardson, 2001. A best-first alignment algorithm for automatic extraction of transfer mappings from bilingual corpora. Proc. Workshop Data-driven Methods Mach. Trans., 14: 1-8.

Papineni, K., S. Roukos, T. Ward and W.J. Zhu, 2002. BLEU: A method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, July 07-12, Philadelphia, Pennsylvania, pp: 311-318.

Peng, L. and Z. Kai-Hui, 2011. Question classification based on rough set attributes and value reduction. Inform. Technol. J., 10: 1061-1065.

Rajan, M.S. and S.P. Rajagopalan, 2008. Effective information retrieval using supervised learning approach. Inform. Technol. J., 7: 231-233.

Su, K.Y., M.W. Wu and J.S. Chang, 1992. A new quantitative quality measure for machine translation systems. Proc. 14th Conf. Comput. Linguistics, 2: 433-439.

Yamada, S., K. Imamura and K. Yamamoto, 2002. Corpus-assisted expansion of manual MT knowledge. Proceedings of the 9th International Conference on Theoretical and Methodological Issues in Machine Translation, (TMI'02), Keihanna, Japan, pp: 199-208.

Yasuda, K., F. Sugaya, T. Takezawa, S. Yamamoto and M. Yanagida, 2001. An automatic evaluation method of translation quality using translation answer candidates queried from a parallel corpus. Proc. Mach. Trans. Summit, 8: 373-378.

Zhang, C.X., Y.H. Liang, P. Li, Z.M. Lu and Y. Liu, 2010. Extracting translation equivalences automatically based on tree-string. Inform. Technol. J., 9: 371-375.