*i*CPE: A Hybrid Data Selection Model for SMT Domain Adaptation

Longyue Wang, Derek F. Wong, Lidia S. Chao, Yi Lu and Junwen Xing

Natural Language Processing & Portuguese-Chinese Machine Translation Laboratory
Department of Computer and Information Science
University of Macau, Macau S.A.R., China

{mb15505, derekfw, lidiasc, mb25435, mb15470}@umac.mo

Abstract. Data selection is a significant technique to enhance the data-driven models especially for large-scale natural language processing (NLP). Recent research on statistical machine translation (SMT) domain adaptation focuses on the usage of various individual data selection models. In this paper, we proposed a hybrid data selection model named *i*CPE, which combines three state-of-the-art similarity metrics: **C**osine *tf-idf*, **P**erplexity and **E**dit distance at both corpus level and model level. We conduct the experiments on Hong Kong Law Chinese-English corpus and the results show that this simple and effective hybrid model performs better over the baseline system trained on entire data as well as the best rival method. This consistently boosting performance of the proposed approach has a profound implication for mining very large corpora in a computationally-limited environment.

Keywords: Data Selection; Statistical Machine Translation; Domain Adaptation; Hybrid Model; Similarity Metrics.

1 Introduction

The performance of SMT [1] system depends heavily upon the quantity of training data as well as the domain-specificity of the test data with respect to the training data. A well-known challenge is that the data-driven system is not guaranteed to perform optimally if the data for training and testing are not identically distributed. Thus, domain adaptation techniques are employed to improve the translation quality for a text in a particular domain using some mixture of in-domain and out-of-domain data.

Researchers discussed the domain adaptation problems for SMT in various perspectives such as mining unknown words from comparable corpora [2], weighted phrase extraction [3], corpus weighting [4] and mixing multiple models [5, 6, 7], etc. Actually, data selection is one of the corpus weighting methods[1] [8]. One of the dominant approaches is to select data suitable for the target domain from a large general-domain corpus (general corpus). There is an underlying assumption that the general

[1] They are data selection, data weighting and translation model adaptation.

adfa, p. 1, 2011.

corpus is broad enough to cover a certain amount of sentences that fall in target domain[2]. A domain-adapted machine translation system could then be trained on these subcopora instead of the entire general corpus. These supplementary data selection approaches play an important role in i) improving the quality of word alignment, ii) preventing irrelevant phrase pairs, and iii) optimizing the re-ordering of output sentences.

Until now, three state-of-the-art selection criteria have been discussed in different perspectives. The first is cosine *tf-idf* (term frequency-inverse document frequency) similarity. Hildebrand et al. [9] applied this technique to construct TM and LM adaptation and they show it is possible to adapt TMs for SMT by selecting similar sentences from general corpus. Furthermore, Lü et al. [10] proposed re-sampling and re-weighting methods for online and offline TM optimization, which are closer to a real-life SMT system. However, they obtained a slight improvement still using a large subset of total data. The second one is perplexity-based approaches, which is used to score text segments according to an in-domain LM. Recently, Moore and Lewis [11] derived the difference of the cross-entropy from a simple variant of Bayes rule. However, this is a preliminary study which did not yet show an improvement for MT task. It was further developed by Axelrod et al. [12] for SMT domain adaptation. The experimental results show that the fast and simple technique allows to discard over 99% of the general corpus resulted in an increase of 1.8 BLEU points. However, the improvement is not stable due to the selection threshold, which is hard to be estimated to ensure optimal translation quality. The third model is not explicitly used for SMT, but is still applicable to our scenario. Edit distance (ED) is a widely used similarity measure for example-based MT (EBMT), known as Levenshtein distance (LD) [13]. Koehn and Senellart [14] applied this method for convergence of translation memory (TM) and SMT. Then Leveling et al. [15] investigated different approaches (e.g., LD and standard IR) to find similar sentences for EBMT. Therefore, we consider edit distance as a new similarity metric for this domain adaptation task.

The analysis shows that each individual retrieval model has its own advantages and disadvantages, which result in their performance either unclear or unstable. Instead of exploring any single individual model, this paper provides a novel method to obtain a robust and effective data selection model for domain adaptation. We propose a hybrid model by performing linear interpolation on the three presented similarity metrics. We design it for both TM adaptation and LM adaptation at two levels: i) *corpus level* where joining the sub-corpora obtained via different individual model; and ii) *model level* where interpolating multiple TMs or LMs together. To compare the proposed model with the presented individual models, we conduct comparative experiments on a large Chinese-English general corpus to adapt to in-domain sentences on Hong Kong law. Using BLEU [16] as an evaluation metric, results indicate that the proposed approach can achieve consistent and significant improvement over baseline systems as well as any signal individual model.

This paper is organized as follows. We firstly review the related work in Section 2. The proposed and other related similarity models are described in Section 3. Section 4

[2] It is also defined as *pseudo in-domain subcorpus* by Axelrod et al. [12].

details the configurations of experiments. Finally, we compare and discuss the results in Section 5 followed by the conclusions to end the paper.

2 Background

In this section, we revisit three state-of-the-art data selection models: cosine *tf-idf*, perplexity and edit distance.

2.1 Cosine *tf-idf*

Cosine *tf-idf* similarity metric comes from the realm of information retrieval (IR). It is a simple but effective co-occurrence (e.g., word overlap) based matching, which is calculated by

$$w_{ij} = tf_{ij} \times \log(idf_j) \tag{1}$$

in which each document D_i is represented as a vector (w_{i1}, w_{i2},…, w_{in}), and n is the size of the vocabulary. tf_{ij} is term frequency (TF) of the j-th word in the vocabulary in the document D_i, and idf_j is the inverse document frequency (IDF) of the j-th word calculated. The similarity between two texts is then defined as the cosine of the angle between two vectors [17, 18]. It is good at retrieving similar (genres) sentences as well as reducing the number of out-of-vocabulary (OOV) words. However, only considering individual keyword may result in weakness in filtering irrelevant data (noises).

In practice, we only use the sentences in source language for indexing and query generating. Each sentence in general corpus is indexed as one document by Apache Lucene[3]. And each sentence without stop words from the reference set is used as one separate query. Besides, we make use of duplicated sentences which is similar with [9]. All retrieved sentences with corresponding target parts are ranked according to their similarity scores. Supposed that M is the size of query set and N is the number of sentences retrieved from general corpus according to each query. Thus, the size of the new sub-corpus is $Size_{Cos\text{-}IR} = M \times N$.

2.2 Perplexity Based

Perplexity can be found in the field of language modeling. As similarity metrics, it employs an n-gram language model, which considers not only the distribution of terms but also the collocation. Perplexity PP and cross-entropy $H(x)$ are monotonically related and $H(x)$ is often applied as a cosmetic substitute of PP [11]. Until now, there have been three perplexity-based variants explored for SMT domain adaptation. Among them, a metric that sums cross-entropy difference over both sides shows the best performance for this topic [12]. Cross-entropy difference is helpful to select the

[3] Available at http://lucene.apache.org.

sentences that are more similar to in-domain corpus but different from others in general corpus. Besides, considering the bilingual resources are useful in balancing the OOV and noises. However, its performance is very sensitive to quality and quantity of the model trained on a reference set. This bilingual cross-entropy difference can be simply represented as follows:

$$[H_{I-src}(x) - H_{G-src}(x)] + [H_{I-tgt}(x) - H_{G-tgt}(x)] \quad (2)$$

where $H_I(x)$ and $H_O(x)$ are the cross-entropy of string x according to a language model LM_I and LM_O which are respectively trained by in-domain data set I and general-domain data set G. *src* and *tgt* are the source and target side of training data.

The candidates with lower scores have higher relation to in-domain set. The size of the new subset $Size_{PP}$ should be equal to $Size_{Cos\text{-}IR}$. Besides, we perform SRILM toolkit[4] [19] to conduct 5-gram LMs with interpolated modified Kneser-Ney discounting [20].

2.3 Edit Distance Based

Edit distance based metric is much stricter than the former two methods, because words overlap, order and position are all involved in similarity calculation. This seems to be able to find the most ideal sentences. Given a sentence s_G from general corpus and a sentence s_R from the reference set, the edit distance for these two sequences is defined as the minimum number of edits, i.e. symbol insertions, deletions and substitutions, needed to transform s_G into s_R. Based on Levenshtein distance or edit distance, there are several different implementations. We used the normalized Levenshtein similarity score (fuzzy matching score, FMS):

$$FMS = 1 - \frac{LD(s_G, s_R)}{Max(|s_G|, |s_R|)} \quad (3)$$

which has been presented by Koehn and Senellart [14] and Leveling et al. [15]. In our system, we employed a word-based Levenshtein edit distance function. If there is a sentence of which score exceed a threshold, we would further penalize these sentences according to space and punctuations edit differences. We implemented the algorithm with map reduce technique to overcome the time-consuming problem [21].

3 The Proposed Approach

The existing domain adaptation methods can be summarized into two broad categories: i) *corpus level* by selecting, joining, or weighting the datasets upon which the models are trained; and ii) *model level* by combining multiple models together in a weighted manner [12].

[4] Available at http://www.speech.sri.com/projects/srilm/.

For corpus level combination, we weight the sub-corpora retrieved by different methods by modifying the frequencies of the sentence in GIZA++ file [10] and then join them together. It can be formally stated as follows:

$$\begin{aligned} iCPE_{(S_i,T_i)} = & \alpha CosIR(S_x, T_x) \\ & + \beta PPBased(S_y, T_y) \\ & + \lambda EDBased(S_z, T_z) \end{aligned} \tag{4}$$

where α, β and λ are the weights for different criteria. (S_x, T_x), (S_y, T_y) and (S_z, T_z) are the sentence pairs respectively selected by cosine *tf-idf* (*CosIR*), perplexity-based (*PPBased*) and edit-distance based (*EDBased*).

For model level combination, we perform linear interpolation on the models trained with the sub-corpora retrieved by different data selection methods. The phrase translation probability $\phi(\bar{f} \mid \bar{e})$ and the lexical weight $p_w(\bar{f} \mid \bar{e}, a)$ are estimated using Eq. 5 and Eq. 6, respectively.

$$\phi(\bar{f} \mid \bar{e}) = \sum_{i=0}^{n} \alpha_i \phi_i(\bar{f} \mid \bar{e}) \tag{5}$$

$$p_w(\bar{f} \mid \bar{e}, a) = \sum_{i=0}^{n} \beta_i p_{w,i}(\bar{f} \mid \bar{e}, a) \tag{6}$$

where i = 1, 2, 3 denote phrase translation probability and lexical weight trained with the sub-corpora retrieved by *CosIR*, *PPBased* and *EDBased*. α_i and β_i are the interpolation weights.

4 Experimental Setup

4.1 Corpora

Two corpora are needed for the adaptation task. Our general-domain corpus includes more than 1 million parallel sentences comprising various genres such as newswires (LDC2005T10), sample sentences from dictionaries, law literature and other crawled sentences. The distribution of domains and sentence length of the general corpus are shown in Table 1 and Fig. 1, respectively. The in-domain corpus and test set are randomly selected that are disjoined from the LDC corpus (LDC2004T08), consisting of texts of Hong Kong law. All of them were segmented (with the same segmentation scheme)[5] [22, 23] and tokenized[6] [24]. In the preprocessing, we also removed the sentences with length more than 80. To evaluate the methods for both LM and TM, we used the target side sentences of the corpora to train all the LMs for translation. The sizes of the test set, in-domain corpus and general corpus we used are summarized in Table 2.

[5] IC-TCLAS2013 is available at http://ictclas.nlpir.org/.

[6] The scripts are available at http://www.statmt.org/europarl/.

In previous work, cosine *tf-idf* method often selected data using test set as reference set [9, 10], which limits the practical applicability of the method in a real-life SMT system. For perplexity-based approaches, an in-domain corpus which is identical to the test sentences is employed for data selection [11, 12]. To compare the different methods fairly, we propose two strategies: one is *offline strategy* where we use test set to find similar sentences in general corpus; the other one is called *online strategy* where an additional in-domain corpus is used to select useful data.

Table 1. Domain proportions in general corpus.

Statistics	**Domains**				**Total**
	News	*Novel*	*Law*[b]	*Miscellaneous*[a]	
Sentence Number (#)	279,962	304,932	48,754	504,396	1,138,044
Percentage (%)	24.60	26.79	4.28	44.33	100.00

[a] Miscellaneous part includes crawled sentences from various sources.

[b] The law part includes the articles of law in Chinese mainland, Hong Kong and Macau.

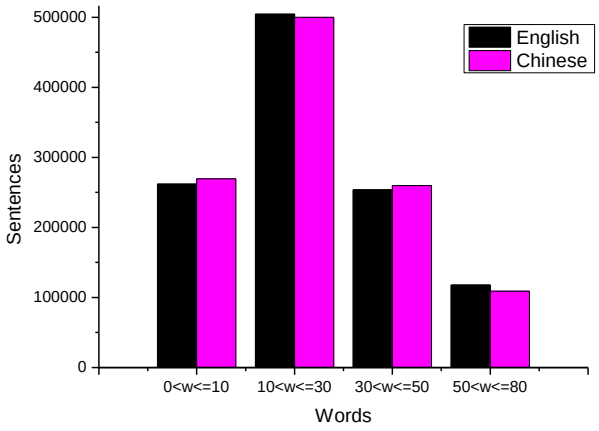

Fig. 1. Distributions of sentences (length) of general corpus.

Table 2. Statistics summary of used corpora.

Data Set	**Language**	**Sentences**	**Tokens**	**Ave. Len.**
Test Set	*English*	2,050	60,399	29.46
	Chinese		59,628	29.09
In-domain	*English*	45,621	1,330,464	29.16
	Chinese		1,321,655	28.97
Training Set	*English*	1,138,044	28,626,367	25.15
	Chinese		28,239,747	24.81

4.2 System Description

The experiments presented in this paper were carried out with the Moses toolkit [25], a state-of-the-art open-source phrase-based SMT system. The translation and the re-ordering model relied on "*grow-diag-final*" symmetrized word-to-word alignments built using GIZA++ [26] and the training script of Moses. A 5-gram language model was trained using the IRSTLM toolkit [27], exploiting improved Modified Kneser-Ney smoothing, and quantizing both, probabilities and back-off weights.

4.3 Baseline

The baseline systems were trained with the toolkits and settings as described above. The in-domain baseline (IC-Baseline) was trained on the in-domain corpus which produces a 12.26M phrase table. The general-domain baseline (IC-Baseline) was substantially larger, having a 1.57G phrase table. The BLEU scores of the baseline systems are in Table 3.

The baseline results show that a translation system trained on the general corpus outperforms a system trained on the in-domain corpus by over 2.85 BLEU points. Although the in-domain data could improve the quality of word alignment, it is not broad enough to reduce the OOV words. As described in next section, the GC-Baseline system result will be further improved by data selection methods.

Table 3. BLEU via General and In-domain corpus.

Baseline	Phrase Table Size	BLEU
GC-Baseline	1.57G	**39.15**
IC-Baseline	12.26M	36.30

5 Results and Discussion

In order to evaluate the performance of the presented models, we implemented three individual data selection models: cosine *tf-idf* (Cos-IR), bilingual cross-entropy difference (B-CED), fuzzy matching scorer (FMS) as well as the proposed model at corpus level (*i*CPE-C) and model level (*i*CPE-M). For each method, we selected the top N={80K, 160K, 320K} sentence pairs out of the 1.1M in the general corpus[7]. Table 4 contains BLEU scores of the systems trained on subsets selected via different models.

All the methods but FMS could be used to train a state-of-the-art SMT system. Cos-IR improves by at most 1.02 (offline) and 0.88 (online) BLEU points using 28.12% of the general corpus. The results approximately match with the conclusions of [9, 10]. This shows that keywords overlap plays a significant role in finding sentences in similar domains. Besides, Cos-IR has a strong robustness because the selection with online strategy still works well. However, it needs a large amount of select-

[7] Roughly 7.0%, 14.0%, 28.0% of general-domain corpus. Besides, K is short for thousand and M is short for million.

ed data (28.0%) to obtain an ideal performance. The main reason is that the sentences including same keywords still may be irrelevant. For instance, there are two sentences including the same phrase "*according to the article*", but one may be in the domain of law and other one may be from news.

Table 4. Translation results via different methods.

Method	**Sentences**	**BLEU (Offline)**	**BLEU (Online)**
GC-Baseline	1.1*M*	**39.15**	
IC-Baseline	1.1*M*	36.30	
Cos-IR	80*K*	39.04	37.53
	160*K*	39.85	39.45
	320*K*	**40.17**	**40.03**
B-CED	80*K*	40.91	35.50
	160*K*	**41.12**	39.47
	320*K*	40.02	**40.98**
FMS	80*K*	37.42	36.22
	160*K*	37.90	36.71
	320*K*	38.15	38.00
*i*CPE-C	80*K*	42.25	39.39
	160*K*	**43.04**	**41.87**
	320*K*	42.42	40.44
*i*CPE-M	80*K*	42.93	40.57
	160*K*	43.65	41.95
	320*K*	**43.97**	**42.21**

PPBased variant B-CED works very well with the offline strategy. It achieves 41.12 (using 7.0% data) and 40.98 (using 14.0% data) BLEU with offline and online strategies. This indicates that bilingual resources are very useful to build a stable in-domain model. When using an in-domain corpus as the reference set, B-CED should enlarge the size of selected data to obtain an ideal BLEU. It has a good but unstable performance with different strategies. The main reason is that considering the word order may be helpful to filter the noise, but it depends heavily upon the in-domain LMs. Similar to the discussion in Section 1, they are so sensitive to the quality and quantity of reference sets, which was not reported by [12].

FMS fails to outperform the baseline system even it is much stricter than other criteria. When adding word position factor into similarity measuring, FMS tries to find nearly the same sentences on length, collocation and semantics. But our general corpus seems not large enough to cover a certain amount of FMS-similar sentences. With increasing the size of general or in-domain corpus, we believe FMS may work better.

We combined Cos-IR, FMS and B-CED (which is the best one among PPBased criteria) and gave equal weights (set $\alpha = \beta = \lambda = 1$ in Eq. 4 and $\alpha_i = \beta_i = 1/3$ in Eq. 5 and 6) to each component at two combination levels. At both levels, *i*CPE performs much better than other methods as well as the baseline systems. This shows a strong

ability to balance the OOV and noise problems. On the one hand, filtering too much unmatched words may not sufficiently address the data sparsity issue of the SMT model; on the other hand, adding too much of the selected data may lead to the dilution of the in-domain characteristics of the SMT model. However, it seems to succeed the advantage of each individual model when combining them together. For instance, the performance of *i*CPE does not drop sharply (like PPBased approaches) when using an in-domain corpus as reference set. This not only shows its stronger robustness for building a real-life SMT system, but also proves that combination method works better than any single individual approach.

Furthermore, *i*CPE has achieved at most 3.89 (offline) and 2.72 (online) improvements over the baseline system at corpus level combination. Besides, the result is still higher than the best individual model (B-CED) by 1.92 (offline) and 0.91 (online). The performance can be further improved by interpolating at the model level. It works better (obtained around 1 BLEU point improvement) than the corpus combination method in the same settings.

6 Conclusions

In this paper, we regard data selection as a problem of measuring similarities via different criteria. This is the first time to systematically compare the state-of-the-art data selection methods for SMT adaptation. We not only explore edit-distance based method for this task for the first time, but also present offline and online strategies for fair comparison. We further integrate the presented individual data selection model at both corpus and model levels. It achieves a good performance in terms of its robustness and effectiveness.

In order to evaluate the proposed data selection model on a large general corpus, we compare it with 3 other related methods: Cos-IR, B-CED, FMS as well as two baseline systems. We can analyze the results from three different aspects:

- *Translation Quality*. The results show a significant performance of the most methods in particular the proposed *i*CPE. It suggests better to use bilingual resources in similarity measuring.
- *Noise Filtering*. *i*CPE could discard about 93% data of the general corpus with a better translation quality. While other models perform either badly or unsteadily.
- *Robustness*. To build a real-life system, in-domain data set is preferable (online strategy). However, only *i*CPE gives a consistently boosting performance.

Acknowledgment

The authors are grateful to the Science and Technology Development Fund of Macau and the Research Committee of the University of Macau for the funding support for our research, under the reference No. 017/2009/A and MYRG076(Y1-L2)-FST13-WF. The authors also wish to thank the anonymous reviewers for many helpful comments.

References

1. Brown P. F., V. J. D. Pietra, Pietra S. A. D., Mercer and R. L.: The mathematics of statistical machine translation: Parameter estimation. Computational linguistics vol. 19, no. 2, pp. 263–311 (1993)
2. Daumé III H. and Jagarlamudi J.: Domain adaptation for machine translation by mining unseen words. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011) (2011)
3. Mansour S. and Ney H.: A simple and effective weighted phrase extraction for machine translation adaptation. IWSLT (2012)
4. Koehn P. and Haddow B.: Towards effective use of training data in statistical machine translation. Proceedings of the Seventh Workshop on Statistical Machine Translation pp. 317–321 (2012)
5. Civera J. and Juan A.: Domain adaptation in statistical machine translation with mixture modeling. Proceedings of the Second Workshop on Statistical Machine Translation pp. 177–180 (2007)
6. Foster G. and Kuhn R.: Mixture-model adaptation for SMT. Proceedings of the Second ACL Workshop on Statistical Machine Translation pp. 128 – 136 (2007)
7. Eidelman V., Boyd-Graber J., and Resnik P.: Topic models for dynamic translation model adaptation. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers Volume 2, pp. 115–119 (2012)
8. Matsoukas S., Rosti A.V. I., and Zhang B.: Discriminative corpus weight estimation for machine translation. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2, pp. 708–717 (2009)
9. Hildebrand A. S., Eck M., Vogel S., and Waibel A.: Adaptation of the translation model for statistical machine translation based on information retrieval. Proceedings of EAMT vol. 2005, pp. 133–142 (2005)
10. Lü Y., Huang J., and Liu Q.: Improving statistical machine translation performance by training data selection and optimization. Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL) pp. 343–350 (2007)
11. Moore R. C. and Lewis W.: Intelligent selection of language model training data. Proceedings of the ACL 2010 Conference Short Papers, pp. 220–224 (2010)
12. Axelrod A., He X., and Gao J.: Domain adaptation via pseudo in-domain data selection. Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 355–362 (2011)
13. Levenshtein V. I.: Binary codes capable of correcting deletions, insertions and reversals. Soviet physics doklady, vol. 10, pp. 707 (1966)
14. Koehn P. and Senellart J.: Convergence of translation memory and statistical machine translation. Proceedings of AMTA Workshop on MT Research and the Translation Industry, pp. 21–31 (2010)
15. Leveling J. et al.: Approximate sentence retrieval for scalable and efficient example-based machine translation. COLING 2012 pp. 1571-1586 (2012)
16. Papineni K., Roukos S., Ward T., and Zhu W. J.: BLEU: a method for automatic evaluation of machine translation. Proceedings of the 40th annual meeting on association for computational linguistics pp. 311–318 (2002)
17. Wang L. Y., Wong D. F., and Chao L. S.: TQDL: Integrated models for cross-language document retrieval. International Journal of Computational Linguistics and Chinese Language Processing (IJCLCLP), 17(4): 15-31 (2012)

18. Wang L. Y., Wong D. F., and Chao L. S.: An improvement in cross-language document retrieval based on statistical models. Processing of the 24th Conference on Computational Linguistics and Speech (ROCLING 2012), 144-155 (2012)
19. Stolcke A. et al.: SRILM-an extensible language modeling toolkit. Proceedings of the international conference on spoken language processing vol. 2, pp. 901–904 (2002)
20. Chen S. F. and Goodman J.: An empirical study of smoothing techniques for language modeling. Proceedings of the 34th annual meeting on Association for Computational Linguistics, pp. 310–318 (1996)
21. Wang Longyue, Wong Derek F., Chao Lidia S., Xing J. W., Lu Y. and Trancoso Isabel: Edit Distance: A new data selection criterion for SMT domain adaptation. Proceedings of Recent Advances in Natural Language Processing. (2013)
22. Zhang H. P., Yu H. K., Xiong D. Y., and Liu Q.: HHMM-based Chinese lexical analyzer ICTCLAS. Proceedings of the second SIGHAN workshop on Chinese language processing-Volume 17, pp. 184–187 (2003)
23. Wang L. Y., Wong D. F., and Chao L. S., Xing J. W.: CRFs-based Chinese word segmentation for micro-blog with small-scale data. Proceedings of the Second CIPS-SIGHAN Joint Conference on Chinese Language, pp. 51–57, 20-21 (2012)
24. Koehn P.: Europarl: A parallel corpus for statistical machine translation. MT summit, vol. 5 (2005)
25. Koehn P. et al.: Moses: Open source toolkit for statistical machine translation. Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, pp. 177–180 (2007)
26. Och F. J. and Ney H.: A systematic comparison of various statistical alignment models. Computational linguistics, vol. 29, no. 1, pp. 19–51 (2003)
27. Federico M., Bertoldi N., and Cettolo M.: IRSTLM: an open source toolkit for handling large scale language models. Proceedings of Interspeech, pp. 1618–1621 (2008)