As expected, using the same domain data for training and testing achieves the best results as indicate by **bold fonts**. The results demonstrate again that relevant data is better data.

To test our online model optimization method, we divide the baseline corpus according to the origins of sub corpus. That is, the FBIS, HK_ Hansards and HK_News models are used as three sub-models and the baseline model is used as general model. The four weighting schemes described in section 3.2 are used as online weighting schemes individually. The experimental results are shown in Table 6. S_*i* indicates the system using weighting scheme *i*.

System Test data	FBIS	HK_ Hansards	HK_ News	Baseline
FBIS-part	**0.1096**	0.0687	0.0622	0.1030
HK_Hans_part	0.0726	**0.0918**	0.0846	0.0897
HK_News_part	0.0664	0.0801	**0.0936**	0.0870
MT05_part	0.1130	0.0805	0.0776	0.1116
Whole test set	0.0937	0.0799	0.0781	0.0993

Table 5. Baseline results on new test set

System Test data	S_1	S_2	S_3	S_4
FBIS-part	0.1090	0.1090	0.1089	0.1089
HK_Hans_part	0.0906	0.0903	0.0902	0.0902
HK_News_part	0.0952	0.0950	0.0933	0.0934
MT05_part	0.1119	0.1123	0.1149	0.1151
Whole test set	0.1034	0.1034	0.1038	0.1038

Table 6. Online model optimization results

Different weighting schemes don't show significant improvements from each other. However, all the four weighting schemes achieve better results than the baseline system. The improvements are shown not only on the whole test set but also on each part of the sub test set. The results justify the effectiveness of our online model optimization method.

5 Related work

Most previous research on SMT training data is focused on parallel data collection. Some work tries to acquire parallel sentences from web (Nie et al. 1999; Resnik and Smith 2003; Chen et al. 2004). Others extract parallel sentences from comparable or non-parallel corpora (Munteanu and Marcu 2005, 2006). These work aims to collect more parallel training corpora, while our work aims to make better use of existing parallel corpora.

Some research has been conducted on parallel data selection and adaptation. Eck et al. (2005) propose a method to select more informative sentences based on n-gram coverage. They use n-grams to estimate the importance of a sentence. The more previously unseen n-grams in the sentence the more important the sentence is. TF-IDF weighting scheme is also tried in their method, but didn't show improvements over n-grams. This method is independent of test data. Their goal is to decrease the amount of training data to make SMT system adaptable to small devices. Similar to our work, Hildebrand et al. (2005) also use information retrieval method for translation model adaptation. They select sentences similar to the test set from available in-of-domain and out-of-domain training data to form an adapted translation model. Different from their work, our method further use the small adapted data to optimize the distribution of the whole training data. It takes the full advantage of larger data and adapted data. In addition, we also propose an online translation model optimization method, which make it possible to select adapted translation model for each individual sentence.

Since large scale monolingual corpora are easier to obtain than parallel corpora. There has some research on language model adaptation recent years. Zhao et al. (2004) and Eck et al.(2004) introduce information retrieval method for language model adaptation. Zhang et al.(2006) and Mauser et al.(2006) use adapted language model for SMT re-ranking. Since language model is built for target language in SMT, one pass translation is usually needed to generate n-best translation candidates in language model adaptation. Translation model adaptation doesn't need a pre-translation procedure. Comparatively, it is more direct. Language model adaptation and translation model adaptation are good complement to each other. It is possible that combine these two adaptation approaches could further improve machine translation performance.

6 Conclusion and future work

This paper presents two new methods to improve statistical machine translation performance by making better use of the available parallel training corpora. The offline data selection method