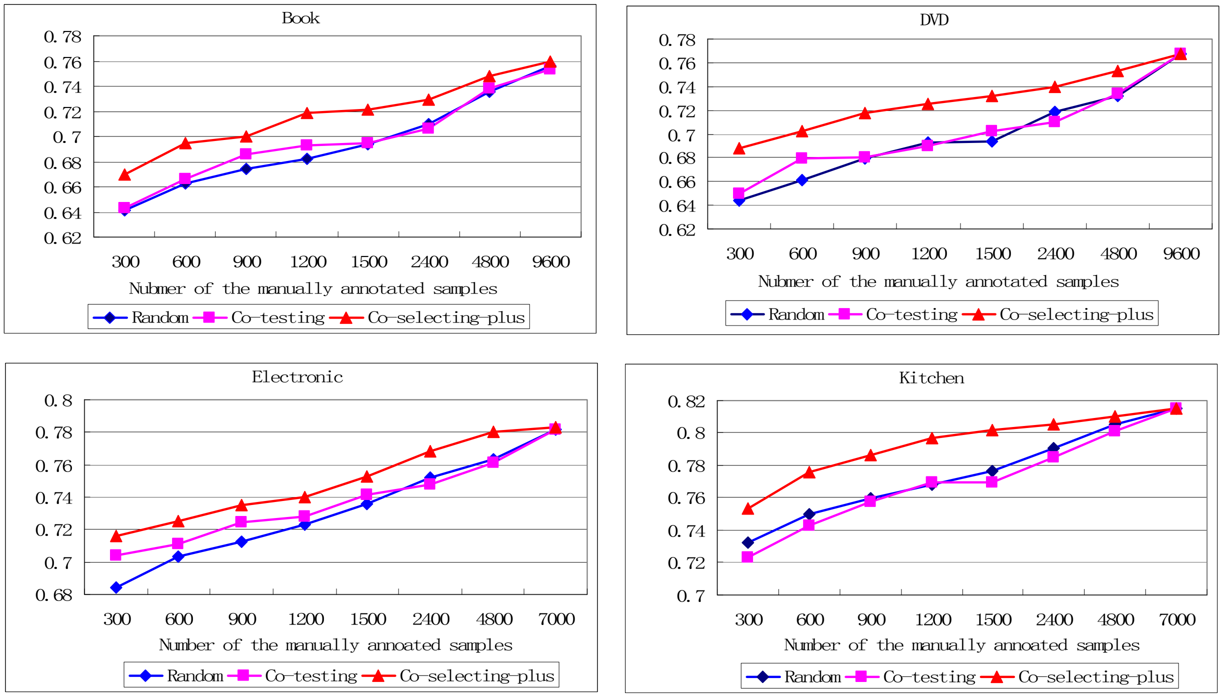

Figure 5: Performance comparison of three active learning approaches: random selection, co-testing and co-selecting-plus, by varying the number of the selected samples for manually annotation

Comparison with other active learning approaches

Figure 4 compares different active learning approaches to imbalanced sentiment classification when 600 unlabeled samples are selected for annotation. Specifically, the parameters θ and k is set to be 1/16 and 50 respectively. Figure 4 justifies that it is challenging to perform active learning in imbalanced sentiment classification: the approaches of **margin-based**, **uncertainty-based** and **self-selecting** perform no better than random selection while **co-testing** only outperforms random selection in two domains: DVD and Electronic with only a small improvement (about 1%). In comparison, our approaches, both **co-selecting-basic** and **co-selecting-plus** significantly outperform the random selection approach on all the four domains. It also shows that **co-selecting-plus** is preferable over **co-selecting-basic**. This verifies the effectiveness of automatically labeling those selected *MA* samples in imbalanced sentiment classification.

Specifically, we notice that only using the certainty measurement (i.e., **certainty**) performs worst, which reflects that only considering sample balance factor in imbalanced sentiment classification is not helpful.

Figure 5 compares our approach to other active learning approaches by varying the number of the selected samples for manually annotation. For clarity, we only include random selection and **co-testing** in comparison and do not show the performances of the other active learning approaches due to their similar behavior to random selection. From this figure, we can see that **co-testing** is effective on Book and Electronic when less than 1500 samples are selected for manual annotation but it fails to outperform random selection in the other two domains. In contract, our **co-selecting-plus** approach is apparently more advantageous and significantly outperforms random selection across all domains (p-value<0.05) when less than 4800 samples are selected for manual annotation.

Sensitiveness of the parameters θ

The size of the feature subspace is an important parameter in our approach. Figure 6 shows the performance of **co-selecting-plus** with varying sizes of the feature subspaces for the first subspace