



Figure 4: Performance comparison of different active learning approaches on imbalanced sentiment classification

Evaluation metrics

The popular geometric mean $G-mean = \sqrt{TP_{rate} \times TN_{rate}}$ is adopted, where TP_{rate} is the true positive rate (also called positive recall or sensitivity) and TN_{rate} is the true negative rate (also called negative recall or specificity) (Kubat and Matwin, 1997).

4.2 Experimental Results

For thorough comparison, various kinds of active learning approaches are implemented including:

- **Random:** randomly select the samples from the unlabeled data for manual annotation;
- **Margin-based:** iteratively select samples closest to the hyperplane provided by the SVM classifier, which is suggested by Ertekin et al. (2007a) and Ertekin et al. (2007b). One sample is selected in each iteration;
- **Uncertainty:** iteratively select samples using the uncertainty measurement according to the output of ME classifier. One sample is selected in each iteration;
- **Certainty:** iteratively select class-balanced samples using the certainty measurement according to the output of ME classifier. One positive and negative sample (the positive and negative label is provided by the ME classifier) are selected in each iteration;
- **Co-testing:** first get contention samples (i.e., unlabeled examples on which the member

classifiers predict different labels) and then select the least confidence one among the hypotheses of different member classifiers, i.e., the aggressive strategy as described Muslea et al. (2006). Specifically, the member classifiers are two subspace classifiers trained by splitting the whole feature space into two disjoint subspaces of same size;

- **Self-selecting:** first select k uncertainty samples and then randomly select a positive and negative sample from the uncertainty-sample set, which is suggested by Doyle et al. (2011). We call it self-selecting since only one classifier is involved to measure uncertainty and predict class labels.

For those approaches involving random selection of features, we run 5 times for them and report the average results. Note that the samples selected by these approaches are imbalanced. To address the problem of classification on imbalanced data, we adopt the under-sampling strategy which has been shown effective for supervised imbalanced sentiment classification (Li et al., 2011a). Our active learning approach includes two versions: the co-selecting algorithm as described in Section 3.2 and the co-selecting with selected *MA* samples automatically labeled as described in Section 3.3. For clarity, we refer the former as **co-selecting-basic** and the latter as **co-selecting-plus** in the following.