neg} denotes the class labels of positive and negative.

In imbalanced sentiment classification, *MI* samples are much sparse yet precious for learning and thus are believed to be more valuable for manual annotation. The key in active learning for imbalanced sentiment classification is to guarantee both the quality and quantity of newly-added *MI* samples. To guarantee the selection of *MI* samples, a certainty measurement is necessary. In this study, the certainty measurement is defined as follows:

$$Cer(d) = \max_{y \in \{pos,neg\}} P(y \mid d)$$

Meanwhile, in order to balance the samples in the two classes, once an informative *MI* sample is manually annotated, an informative *MA* sample is automatically labeled. In this way, the annotated data become more balanced than a random selection strategy.

However, the two sampling strategies discussed above are apparently contradicted: while the uncertainty measurement is prone to selecting the samples whose posterior probabilities are nearest to 0.5, the certainty measurement is prone to selecting the samples whose posterior probabilities are nearest to 1. Therefore, it is essential to find a solution to balance uncertainty sampling and certainty sampling in imbalanced sentiment classification,

3.2 Co-selecting with Feature Subspace Classifiers

In sentiment classification, a document is represented as a feature vector generated from the feature set $F = \{ f_1, ..., f_m \}$. When a feature subset, i.e., $F^S = \{ f_1^S, ..., f_r^S \}$ ($r < m$), is used, the original m-dimensional feature space becomes an r-dimensional feature subspace. In this study, we call a classifier trained with a feature subspace a feature subspace classifier.

Our basic idea of balancing both the uncertainty measurement and the certainty measurement is to train two subspace classifiers to adopt them respectively. In our implementation, we randomly select two disjoint feature subspaces, each of which is used to train a subspace classifier. On one side, one subspace classifier is employed to select some certain samples; on the other side, the other classifier is employed to select the most uncertain sample from those certain samples for manual annotation. In this way, the selected samples are certain in terms of one feature subspace for selecting more possible *MI* samples. Meanwhile, the selected sample remains uncertain in terms of the other feature subspace to introduce uncertain knowledge into current learning model. We name this approach as co-selecting because it collectively selects informative samples by two separate classifiers. Figure 2 illustrates the co-selecting algorithm. In our algorithm, we strictly constrain the balance of the samples between the two classes, i.e., positive and negative. Therefore, once two samples are annotated with the same class label, they will not be added to the labeled data, as shown in step (7) in Figure 2.

```
Input:
  Labeled data $L$ with balanced samples over the two classes
  Unlabeled pool $U$
Output:
  New Labeled data $L$
Procedure:
Loop for $N$ iterations:
(1). Randomly select a feature subset $F^S$ with size $r$ (with the proportion $\theta = r/m$) from $F$
(2). Generate a feature subspace from $F^S$ and train a corresponding feature subspace classifier $C_{Cer}$ with $L$
(3). Generate another feature subspace from the complement set of $F^S$, i.e., $F - F^S$ and train a corresponding feature subspace classifier $C_{Uncer}$ with $L$
(4). Use $C_{Cer}$ to select top certain $k$ positive and $k$ negative samples, denoted as a sample set $CER_1$
(5). Use $C_{Uncer}$ to select the most uncertain positive sample and negative sample from $CER_1$
(6). Manually annotate the two selected samples
(7). If the annotated labels of the two selected samples are different from each other:
       Add the two newly-annotated samples into $L$
```

Figure 2: The co-selecting algorithm

There are two parameters in the algorithm: the size of the feature subspace for training the first subspace classifier, i.e., θ and the number of