CHSL – Training

Input:  
- represent the training set which contains *m* instances

I – base linear classifier

C – clustering algorithm

K – number of clusters

– parameter space for the base classifier

– threshold for specificity

1. –
2. // dataset with only the minor samples
3. M empty array
4. sub\_clusters C(K,)
5. FOR sub\_cluster IN sub\_clusters DO
6. D[i] sub\_cluster
7. M[i] I.train(D[i]) // train with given parameters
8. END FOR
9. A build ensemble from M
10. A // grid search on the parameters with 10 fold cross validation
11. A.train(S,) // train with the best parameters

CHSL – Prediction

Input:

- represent the test set which contains *m* instances to be labeled

Output:

– represent the prediction for the given test set

1. preds empty array
2. FOR x IN S DO:
3. tmp\_result 1
4. FOR m IN A:
5. tmp\_pred m.predict(x)
6. IF tmp\_pred=0 THEN
7. tmp\_result 0
8. STOP FOR
9. preds.add(tmp\_result)
10. Return preds