201103909
Xuanjie Liu

1. Select the mean in a cluster as the cluster centre

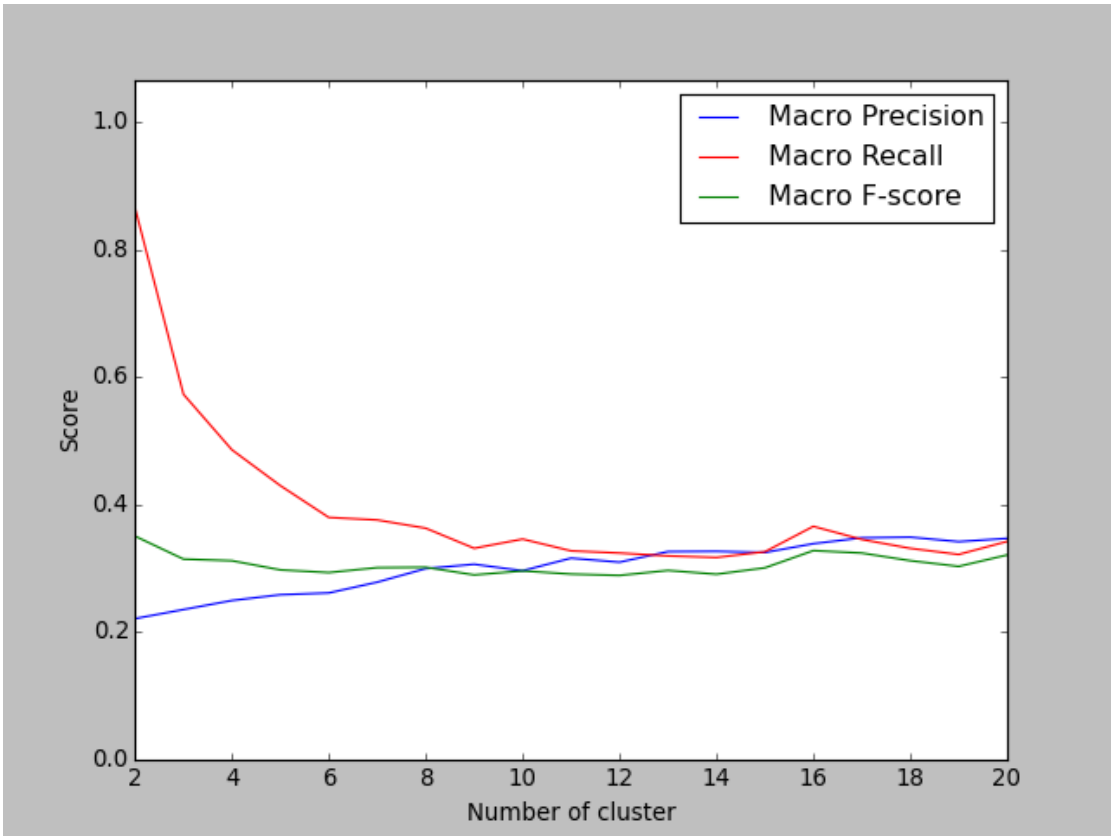

Figure 1 – 1

```
Clusters: 02  Labels: 2.0  Precision: 0.221  Recall: 0.866  F-score: 0.351  Iterations: 11.0
Clusters: 03  Labels: 3.0  Precision: 0.236  Recall: 0.573  F-score: 0.315  Iterations: 9.1
Clusters: 04  Labels: 3.9  Precision: 0.250  Recall: 0.486  F-score: 0.312  Iterations: 9.7
Clusters: 05  Labels: 4.5  Precision: 0.259  Recall: 0.430  F-score: 0.298  Iterations: 8.9
Clusters: 06  Labels: 5.3  Precision: 0.262  Recall: 0.380  F-score: 0.294  Iterations: 7.7
Clusters: 07  Labels: 5.7  Precision: 0.279  Recall: 0.376  F-score: 0.301  Iterations: 7.6
Clusters: 08  Labels: 6.3  Precision: 0.300  Recall: 0.363  F-score: 0.302  Iterations: 6.7
Clusters: 09  Labels: 6.9  Precision: 0.307  Recall: 0.332  F-score: 0.290  Iterations: 6.0
Clusters: 10  Labels: 6.5  Precision: 0.296  Recall: 0.346  F-score: 0.296  Iterations: 4.4
Clusters: 11  Labels: 7.2  Precision: 0.316  Recall: 0.328  F-score: 0.291  Iterations: 4.8
Clusters: 12  Labels: 7.1  Precision: 0.310  Recall: 0.324  F-score: 0.289  Iterations: 4.3
Clusters: 13  Labels: 7.6  Precision: 0.326  Recall: 0.320  F-score: 0.297  Iterations: 3.3
Clusters: 14  Labels: 7.6  Precision: 0.327  Recall: 0.317  F-score: 0.291  Iterations: 3.9
Clusters: 15  Labels: 7.5  Precision: 0.325  Recall: 0.326  F-score: 0.301  Iterations: 3.4
Clusters: 16  Labels: 7.1  Precision: 0.339  Recall: 0.366  F-score: 0.328  Iterations: 3.4
Clusters: 17  Labels: 7.6  Precision: 0.348  Recall: 0.345  F-score: 0.324  Iterations: 3.3
Clusters: 18  Labels: 7.7  Precision: 0.349  Recall: 0.332  F-score: 0.312  Iterations: 3.2
Clusters: 19  Labels: 7.8  Precision: 0.342  Recall: 0.322  F-score: 0.303  Iterations: 3.1
Clusters: 20  Labels: 7.7  Precision: 0.347  Recall: 0.342  F-score: 0.321  Iterations: 2.9
```

Figure 1 – 2

As shown by figure 1 – 1, recall varies in inverse proportion to the number of cluster. It drops down from 0.86 to 0.36 as number of cluster growing form 2 to 8. Precision increases slowly, form 0.22 at 2 clusters to 0.35 at 20 clusters. The F-score levels off at 0.31.

As shown by figure 1 – 2, more and more labels (types of review) can be distinguished as the increase in number of clusters. We cannot guarantee that all the types of review can be assigned to at least one cluster even we add the number of clusters to 20. The number of iterations varies in inverse proportion to the number of cluster. Therefore, when the number of clusters is litter, it needs more iterations to reach convergence, and takes more time to train.

2. Select the instance that is closest to the mean as the cluster centre

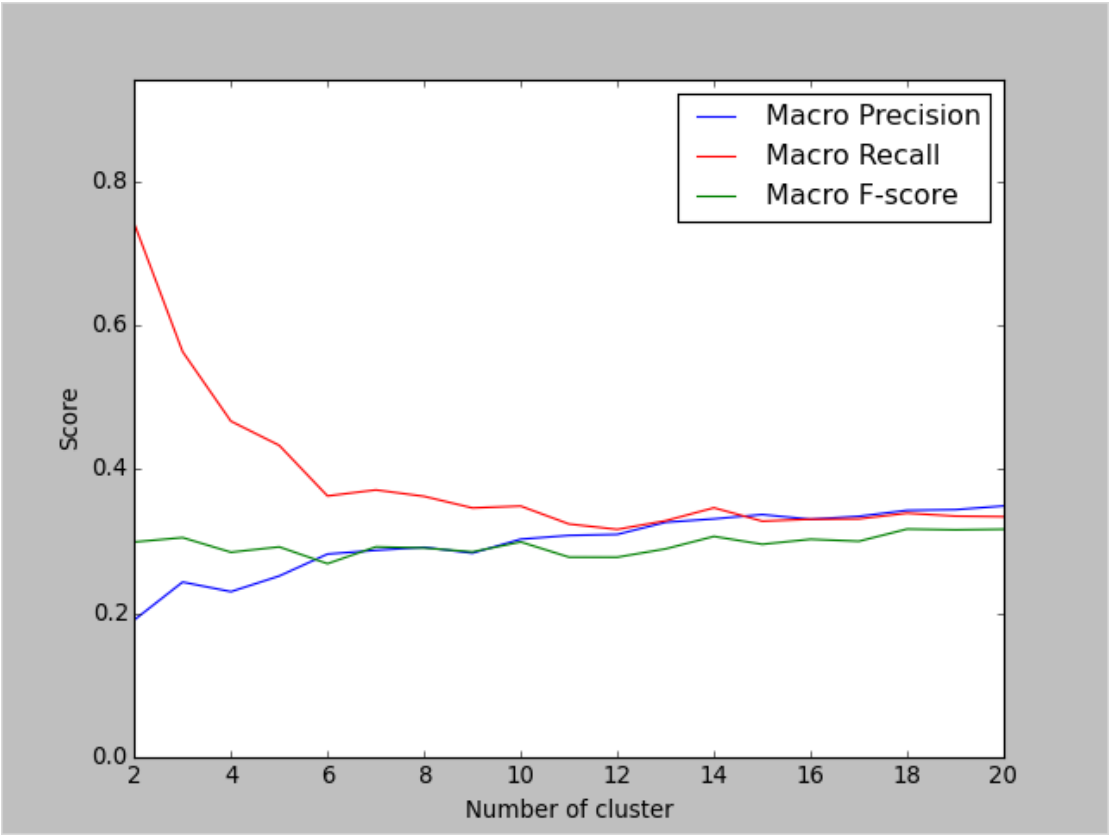

Figure 2 – 1

```
Clusters: 02  Labels: 2.0  Precision: 0.191  Recall: 0.741  F-score: 0.299  Iterations: 3.0
Clusters: 03  Labels: 3.0  Precision: 0.243  Recall: 0.563  F-score: 0.305  Iterations: 3.0
Clusters: 04  Labels: 3.7  Precision: 0.230  Recall: 0.467  F-score: 0.285  Iterations: 3.1
Clusters: 05  Labels: 4.5  Precision: 0.252  Recall: 0.433  F-score: 0.292  Iterations: 2.9
Clusters: 06  Labels: 5.3  Precision: 0.282  Recall: 0.363  F-score: 0.269  Iterations: 3.1
Clusters: 07  Labels: 5.6  Precision: 0.287  Recall: 0.371  F-score: 0.292  Iterations: 2.7
Clusters: 08  Labels: 5.8  Precision: 0.292  Recall: 0.362  F-score: 0.291  Iterations: 2.8
Clusters: 09  Labels: 6.2  Precision: 0.284  Recall: 0.346  F-score: 0.285  Iterations: 2.6
Clusters: 10  Labels: 6.5  Precision: 0.303  Recall: 0.349  F-score: 0.299  Iterations: 2.6
Clusters: 11  Labels: 6.8  Precision: 0.308  Recall: 0.324  F-score: 0.278  Iterations: 2.9
Clusters: 12  Labels: 7.1  Precision: 0.310  Recall: 0.317  F-score: 0.278  Iterations: 2.5
Clusters: 13  Labels: 7.1  Precision: 0.326  Recall: 0.329  F-score: 0.289  Iterations: 2.2
Clusters: 14  Labels: 6.9  Precision: 0.331  Recall: 0.347  F-score: 0.307  Iterations: 2.9
Clusters: 15  Labels: 7.5  Precision: 0.337  Recall: 0.328  F-score: 0.296  Iterations: 2.4
Clusters: 16  Labels: 7.5  Precision: 0.331  Recall: 0.331  F-score: 0.303  Iterations: 2.6
Clusters: 17  Labels: 7.5  Precision: 0.335  Recall: 0.331  F-score: 0.300  Iterations: 2.8
Clusters: 18  Labels: 7.6  Precision: 0.343  Recall: 0.339  F-score: 0.317  Iterations: 2.6
Clusters: 19  Labels: 7.7  Precision: 0.344  Recall: 0.335  F-score: 0.316  Iterations: 2.8
Clusters: 20  Labels: 7.8  Precision: 0.349  Recall: 0.334  F-score: 0.317  Iterations: 2.3
```

Figure 2 – 2

The trends of precision, recall, f-score and the number of distinguished types of review (labels) are very similar as those when select the mean in a cluster as the cluster centre.

The number of iterations to reach convergence, however, is much less than that in selecting mean clustering, extremely when the number of clusters is little. For example, when the number of clusters is 2, it takes 11 iterations to reach converge in selecting mean clustering. In contrast, it takes only 3 iteration to reach converge in selecting the instance that is closest to the mean of the cluster as centre. So in this way, little time is spent on the performance of the k-means clustering.