**LING 570 – HW9**

**Q2**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Expt id | rare thres | feat thres | training accuracy | test accuracy | # of feats | # of kept feats | running time |
| 1\_1 | 1 | 1 | 0.98728 | 0.83544 | 39380 | 39380 | 2 min 59 sec |
| 1\_3 | 1 | 3 | 0.97331 | 0.83422 | 39380 | 12191 | 1 min 47 sec |
| 2\_1 | 2 | 1 | 0.98883 | 0.88322 | 44167 | 44167 | 2 min 55 sec |
| 2\_3 | 2 | 3 | 0.98670 | 0.88240 | 44167 | 14125 | 1 min 49 sec |
| 3\_3 | 3 | 3 | 0.98749 | 0.89179 | 44297 | 15279 | 1 min 54 sec |
| 3\_5 | 3 | 5 | 0.98560 | 0.89424 | 44297 | 8780 | 1 min 59 sec |
| 5\_10 | 5 | 10 | 0.98310 | 0.89873 | 44169 | 4527 | 1 min 33 sec |

The number of features increased as the “rare thres” number increased from 1 to 3. Beyond the “rare thres” number of 3, the number of features generated would start to decline as shown in experiment id 5\_10.

The number of kept features depended on the “rare thres” and “feat thres” numbers, and would be very high if a higher “rare thres” number was matched with a low “feat thres” number, as shown in experiment id 2\_1. This was because a higher “rare thres” number would have generated more features and a comparatively lower “feat thres” would have resulted in lesser number of those features generated by the “rare thres” number being discarded. This should have resulted in a training vector that contained a high number of kept features.

Experiments with high number of kept features took the longest to run as shown in experiment id 1\_1 and 2\_1. Hence in experiment id 5\_10, the running time was the fastest given that the number of kept features was also the lowest. As described above, the number of kept features would have depended largely on how the numbers for “rare thres” and “feat thres” were set.

Given a “rare thres” number, the training accuracy declined as the “feat thres” number was adjusted higher. On the other hand, given a “feat thres” number, the training accuracy increased as the “rare thres” number was adjusted higher. The training accuracy was highest with the experiment id 2\_1. Beyond experiment id 2\_1, as the “rare thres” and “feat thres” numbers were both adjusted higher, the training accuracy started to show a declining trend.

When the “rare thres” number was at 1 or 2, the test accuracy declined as the “feat thres” number was adjusted comparatively higher. That is to say, given a “rare thres” number of either 1 or 2, the test accuracy would move lower as the “feat thres” number was adjusted higher. However beyond the “rare thres” number of 2, the test accuracy started to increase consistently as both the number for “rare thres” and “feat thres” were adjusted higher.

*End of HW9 – submitted by Wee Teck Tan*

*Student ID: 0937003*

*Course Name: LING 570*