一、股票价格涨跌幅＋经营范围描述：错误率

注：错误率的定义是，如果A、B两只股票不属于同一个分类，但在我们的聚类中被聚到了一起，则被认为是一次错误；反之，如果A、B本属于一类，在聚类结果中不属于同一类，则不记为错误。也就是说，我们允许比原分类更细的分类方法，但不允许相反的情况发生。如果要求两者严格相同，则称为严格错误率。

|  |  |  |
| --- | --- | --- |
|  | S50-09 | S50-10 |
| LDA | 0.2200 | 0.2600 |
| LDA-MF | 0.3400 | 0.1600 |
| PLSA | 0.3000 | 0.2800 |
| PLSA-MF | 0.2200 | 0.1200 |
| RP-KMEANS | 0.2000 | 0.2600 |
| RP-MF | 0.3000 | 0.1800 |
| KMEANS | 0.0400 | 0.0800 |
| KMEANS-MF | 0.0400 | 0.0800 |
| NNMF-KMEANS | 0.0200 | 0.0600 |
| NNMF-MF | 0 | 0 |
| MF | 0.0400 | 0.0200 |

二、Toy data

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Wunrelated | 1 | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |  |
| LDA | 0.3050 | 0.4530 | 0.2690 | 0.4560 | 0.4180 | 0.4520 | 0.3830 | 0.5060 | 0.3280 | 0.3290 | 0.4020 | 0.0748 |
| LDA-MF | 0.1010 | 0.1210 | 0.0400 | 0.0970 | 0.0660 | 0.0470 | 0.0670 | 0.0830 | 0.1380 | 0.1290 | 0.0470 | 0.0347 |
| PLSA | 0.8770 | 0.8740 | 0.8640 | 0.8560 | 0.8790 | 0.8680 | 0.8670 | 0.7460 | 0.8630 | 0.8760 | 0.8820 | 0.0384 |
| PLSA-MF | 0.5300 | 0.5170 | 0.5490 | 0.5110 | 0.5780 | 0.5830 | 0.5340 | 0.3060 | 0.5630 | 0.5400 | 0.5290 | 0.0753 |
| RP-KMEANS | 0.1240 | 0.1720 | 0.2430 | 0.2050 | 0.1420 | 0.1460 | 0.2580 | 0.3070 | 0.1850 | 0.0870 | 0.3810 | 0.0866 |
| RP-MF | 0.1770 | 0.2700 | 0.3020 | 0.3190 | 0.2200 | 0.1790 | 0.3650 | 0.3680 | 0.2230 | 0.1210 | 0.4080 | 0.0921 |
| KMEANS | 0.3890 | 0.4540 | 0.4430 | 0.3360 | 0.4590 | 0.4330 | 0.4020 | 0.3820 | 0.5110 | 0.4680 | 0.3090 | 0.0600 |
| KMEANS-MF | 0.3890 | 0.4540 | 0.4430 | 0.3360 | 0.4590 | 0.4330 | 0.4020 | 0.3820 | 0.5110 | 0.4680 | 0.3090 | 0.0600 |
| NNMF-KMEANS | 0.2490 | 0.3320 | 0.4830 | 0.5210 | 0.4850 | 0.3890 | 0.6060 | 0.3740 | 0.4250 | 0.5760 | 0.3760 | 0.1076 |
| NNMF-MF | 0 | 0 | 0.0230 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0069 |
| MF | 0.1240 | 0.1620 | 0.1890 | 0.1240 | 0.1310 | 0.1330 | 0.1610 | 0.1250 | 0.0950 | 0.1040 | 0.0870 | 0.0305 |

总得来说用NNMF初始化的我们的方法对无关词（Wunrelated）比较稳定。

三、NIPS Corpus

|  |  |
| --- | --- |
| LDA | 0.5431 |
| LDA-MF | 0.3592 |
| PLSA | 0.8218 |
| PLSA-MF | 0.8477 |
| RP-KMEANS | 0.3391 |
| RP-MF | 0.6293 |
| KMEANS | 0.5563 |
| KMEANS-MF | 0.5764 |
| NNMF-KMEANS | 0.5402 |
| NNMF-MF | 0.4328 |
| MF | 0.5569 |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0.1695 | 0.2358 | 0.9248 | 0.9772 | 0.9406 | 0.9067 | 0.8778 | 0.8778 | 0.8759 | 0.9934 | 0.9977 |
|  | 0.1705 | 0.2323 | 0.9248 | 0.9772 | 0.9352 | 0.9039 | 0.8759 | 0.8759 | 0.8655 | 0.9934 | 0.9977 |
|  | 0.1710 | 0.2283 | 0.9098 | 0.9633 | 0.9390 | 0.9082 | 0.8735 | 0.8735 | 0.8650 | 0.9929 | 0.9977 |
|  | 0.1705 | 0.2248 | 0.9098 | 0.9633 | 0.9263 | 0.9023 | 0.8761 | 0.8761 | 0.8745 | 0.9929 | 0.9977 |
|  | 0.1715 | 0.2221 | 0.9098 | 0.9633 | 0.9223 | 0.8974 | 0.8706 | 0.8706 | 0.8663 | 0.9929 | 0.9977 |
|  | 0.1730 | 0.2189 | 0.9098 | 0.9633 | 0.9193 | 0.8942 | 0.8596 | 0.8596 | 0.8619 | 0.9929 | 0.9975 |
|  | 0.1720 | 0.2134 | 0.8845 | 0.9568 | 0.9115 | 0.8850 | 0.8560 | 0.8560 | 0.8554 | 0.9929 | 0.9972 |
|  | 0.1730 | 0.2094 | 0.8845 | 0.9568 | 0.9126 | 0.8882 | 0.8602 | 0.8602 | 0.8513 | 0.9929 | 0.9962 |
|  | 0.1730 | 0.2069 | 0.8845 | 0.9568 | 0.9160 | 0.8927 | 0.8562 | 0.8562 | 0.8650 | 0.9929 | 0.9954 |
|  | 0.1740 | 0.2044 | 0.8816 | 0.9536 | 0.9132 | 0.8869 | 0.8631 | 0.8631 | 0.8646 | 0.9929 | 0.9954 |