## OLR experiments and results

1. **Dataset**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **dataset**  **language** |  | | | test |
| train | val | total |
| zh-cn (Mandarin) | 6.35/4498 | 1.31/900 | 7.66/5398 | 2.66 |
| ct-cn (Cantonese) | 6.42/4799 | 1.29/960 | 7.71/5759 | 2.50 |
| ru-ru (Russian) | 8.08/4492 | 1.84/898 | 9.92/5390 | 3.03 |
| ko-kr (Korea) | 5.03/4496 | 0.95/900 | 5.98/5396 | 1.94 |
| ja-jp (Japanese) | 4.89/4788 | 0.93/950 | 5.82/5738 | 2.18 |
| Id-id (Indonesian) | 6.23/4791 | 1.25/959 | 7.48/5750 | 3.20 |
| vi-vn (Vietnamese) | 7.07/4500 | 1.38/900 | 8.45/5400 | 2.98 |
| total | 44.45/32364 | 8.58/6467 | 53.03/38831 | 18.49/12939 |

(hours/files)

train: 15 speakers/language;

val: 3 speakers/language;

test: 8 speakers/language

(wav files:16k Hz, 16 bits, uncompressed PCM)

1. **Experimental results**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| System settings | | | | Performance on val | |
|  | pre-pocessing | (backend) classifier | post-processing |
| Cavg | Accuracy |
| PPRLM |  | NN+Ngram  (based on BUTPhre) | sum norm on accumulated score | 0.5 | 0.949 |
| min-max norm on accumulated score | 0.039 |
| mean/variance norm on accumulated score | 0.025 |
| on concatenated score (including acoustic score)  (minmax norm) | SVM (linear) |  | 0.021 | 0.956 |
| on concatenated score (including acoustic score)  (minmax norm) | SVM (poly) |  | 0.083 | 0.804 |
| on concatenated score (including acoustic score)  (minmax norm) | SVM (rbf) |  | 0.060 | 0.843 |
| on concatenated score (including acoustic score)  (minmax norm) | SVM (sigmoid) |  | 0.163 | 0.611 |
| on concatenated score (including acoustic score)  (mean variance norm) | SVM (linear) |  | 0.020 | 0.954 |
| on concatenated score (including acoustic score)  (mean variance norm) | SVM (poly) |  | 0.059 | 0.743 |
| on concatenated score (including acoustic score)  (mean variance norm) | SVM (rbf) |  | 0.022 | 0.955 |
| on concatenated score (including acoustic score)  (mean variance norm) | SVM (sigmoid) |  | 0.287 | 0.363 |
| on raw accumulated score | LDA |  | 0.051 | 0.895 |
| on raw concatenated score  (including acoustic score) | LDA |  | 0.039 | 0.918 |
| on concatenated score (including acoustic score)  (minmax norm) | LDA |  | 0.039 | 0.918 |
| on concatenated score (including acoustic score)  (mean variance norm) | LDA |  | 0.039 | 0.918 |
| iVector |  |  |  |  |  |
| mfcc(19dim)/UBM(128 order)/iV(80dim) | LDA |  | 0.107 | 0.767 |
| mfcc(19dim)/UBM(128 order)/iV(80dim)/LDA | SVM(linear) |  | 0.105 | 0.914 |
| mfcc(19dim)/UBM(64 order)/iV(64dim) | LDA |  | 0.126 | 0.724 |
| mfcc(19dim)/UBM(64 order)/iV(64dim) | SVM(NB) |  | 0.169 | 0.696 |
| mfcc(19dim)/UBM(64 order)/iV(80dim) | SVM(NB) |  | ---- | 0.508 |
| SDC | mfcc(7dim)+SDC(7-1-3-7) (0 context) | SVM(NB) |  | 0.393 | 0.283 |
| mfcc(7dim)+SDC(7-1-3-7)  (20 chunksize+10 shift) | SVM(NB) | Accumulate chunk leve score | 0.340 | 0.410 |
| accumulate chunk leve score;  minmax norm | 0.326 |
| accumulate chunk leve score;  mean/variance norm | 0.329 |
|  |  |  |  | |
|  |  |  |  | |

notes:

1. PPRLM system: based on BUT Phone recognizer．
2. i-vector