# Summery

## Baseline VS. Non-History (All Metrics)

(Highlight cells are better)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | test1 |  | test2 |  | test3 |  | test4 |  |
|  |  | baseline | nonhistory | baseline | nonhistory | baseline | nonhistory | baseline | nonhistory |
| schedue1 | accuracy | 0.7748 | 0.7878 | 0.7928 | 0.7921 | 0.9178 | 0.9156 | 0.8233 | 0.827 |
|  | avgp | 0.7953 | 0.8023 | 0.8089 | 0.8139 | 0.9005 | 0.8983 | 0.827 | 0.8281 |
|  | l2 | 0.2895 | 0.2796 | 0.2702 | 0.2632 | 0.1408 | 0.1439 | 0.2447 | 0.2431 |
|  | mrr | 0.8585 | 0.8682 | 0.8522 | 0.8609 | 0.9428 | 0.9405 | 0.898 | 0.897 |
|  | roc.v1\_ca05 | 0.6783 | 0.6828 | 0.7334 | 0.7234 | 0.8853 | 0.8751 | 0.7321 | 0.7322 |
|  | roc.v1\_ca10 | 0.7079 | 0.7197 | 0.7538 | 0.7533 | 0.9178 | 0.9156 | 0.7738 | 0.7734 |
|  | roc.v1\_ca20 | 0.7357 | 0.7747 | 0.7834 | 0.7847 | 0.9178 | 0.9156 | 0.8233 | 0.827 |
|  | roc.v1\_eer | 0.153 | 0.158 | 0.1113 | 0.1209 | 0.0862 | 0.0927 | 0.1392 | 0.1457 |
|  | roc.v2\_ca05 | 0.8598 | 0.8466 | 0.9 | 0.9047 | 0.8226 | 0.8198 | 0 | 0 |
|  | roc.v2\_ca10 | 0.8684 | 0.8552 | 0.9104 | 0.9095 | 0.8346 | 0.8291 | 0 | 0 |
|  | roc.v2\_ca20 | 0.8755 | 0.8667 | 0.9207 | 0.9132 | 0.8642 | 0.8507 | 0.8639 | 0.8611 |
| schedue2 | accuracy | 0.602 | 0.6371 | 0.4905 | 0.4967 | 0.6202 | 0.6065 | 0.5841 | 0.5906 |
|  | avgp | 0.6282 | 0.6441 | 0.518 | 0.5274 | 0.5679 | 0.5549 | 0.595 | 0.596 |
|  | l2 | 0.5258 | 0.5033 | 0.6816 | 0.6684 | 0.611 | 0.6294 | 0.5728 | 0.5714 |
|  | mrr | 0.7428 | 0.7641 | 0.6309 | 0.6468 | 0.7264 | 0.7069 | 0.7574 | 0.7529 |
|  | roc.v1\_ca05 | 0.4136 | 0.4303 | 0.2959 | 0.2958 | 0.1451 | 0.0938 | 0 | 0 |
|  | roc.v1\_ca10 | 0.4386 | 0.433 | 0.3206 | 0.3042 | 0.2692 | 0.1718 | 0.3743 | 0.3764 |
|  | roc.v1\_ca20 | 0.4774 | 0.5226 | 0.3545 | 0.365 | 0.4562 | 0.3712 | 0.4574 | 0.455 |
|  | roc.v1\_eer | 0.2764 | 0.2832 | 0.2962 | 0.3116 | 0.3756 | 0.4287 | 0.3028 | 0.3273 |
|  | roc.v2\_ca05 | 0.6598 | 0.6274 | 0.5602 | 0.5712 | 0.1132 | 0.0801 | 0 | 0 |
|  | roc.v2\_ca10 | 0.682 | 0.657 | 0.6032 | 0.5955 | 0.1801 | 0.1265 | 0 | 0 |
|  | roc.v2\_ca20 | 0.699 | 0.6796 | 0.657 | 0.6123 | 0.3408 | 0.2249 | 0.599 | 0.6012 |
| schedue3 | accuracy | 0.5982 | 0.6108 | 0.4869 | 0.479 | 0.7033 | 0.6898 | 0.6396 | 0.6432 |
|  | avgp | 0.6402 | 0.6475 | 0.5418 | 0.5523 | 0.6367 | 0.6242 | 0.6411 | 0.6404 |
|  | l2 | 0.5088 | 0.4984 | 0.648 | 0.6331 | 0.5137 | 0.5315 | 0.5076 | 0.5085 |
|  | mrr | 0.7418 | 0.7516 | 0.6379 | 0.6571 | 0.7981 | 0.7843 | 0.7905 | 0.7862 |
|  | roc.v1\_ca05 | 0.4471 | 0.4504 | 0.3334 | 0.3328 | 0.212 | 0.1364 | 0 | 0 |
|  | roc.v1\_ca10 | 0.45 | 0.4573 | 0.3513 | 0.3397 | 0.4212 | 0.2761 | 0.3974 | 0.4123 |
|  | roc.v1\_ca20 | 0.4971 | 0.529 | 0.3788 | 0.3758 | 0.6123 | 0.5672 | 0.5178 | 0.5201 |
|  | roc.v1\_eer | 0.2507 | 0.2541 | 0.2517 | 0.2021 | 0.3102 | 0.3475 | 0.3085 | 0.3184 |
|  | roc.v2\_ca05 | 0.7139 | 0.7005 | 0.6447 | 0.6747 | 0.1144 | 0.0821 | 0 | 0 |
|  | roc.v2\_ca10 | 0.7396 | 0.7291 | 0.6848 | 0.6968 | 0.1816 | 0.1315 | 0 | 0 |
|  | roc.v2\_ca20 | 0.7523 | 0.7487 | 0.7215 | 0.7093 | 0.3536 | 0.2369 | 0 | 0.5694 |

Observation:

* Baseline > nonhistory on Test3 on all metrics
* Nonhistory > Baseline on some of them

## Correct SLU Rank

* “-1” means the correct doesn’t appear
* ‘0’ means the correct one appears in the output (system)
* Others means the correct one appears in the SLU of input (speakers)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | test1 | test2 | test3 | test4 | train1a | train2 | train3 |
| -1 | 6366 | 7164 | 8793 | 3643 | 10978 | 5571 | 5723 |
| 0 | 2014 | 2085 | 1381 | 316 | 160 | 1507 | 875 |
| 1 | 1449 | 1332 | 2431 | 661 | 3093 | 1706 | 1756 |
| 2 | 72 | 73 | 266 | 94 | 297 | 236 | 137 |
| 3 | 34 | 28 | 121 | 22 | 80 | 93 | 54 |
| 4 | 21 | 16 | 70 | 9 | 38 | 42 | 29 |
| 5 | 19 | 17 | 45 | 4 | 17 | 26 | 17 |
| 6 | 16 | 10 | 37 | 0 | 15 | 14 | 7 |
| 7 | 7 | 12 | 24 | 3 | 10 | 12 | 6 |
| 8 | 9 | 7 | 13 | 0 | 11 | 11 | 8 |
| 9 | 5 | 15 | 8 | 1 | 12 | 7 | 4 |
| 10 | 5 | 3 | 11 | 0 | 12 | 6 | 4 |

## Top10 SLU Correct Rank

## Correct Slot Values

See Res.xlsx, sheet “slot values”

Observation:

* Many of the values appear only once (Ngram feature might not be good)
* The distributions of values among different sets are different

## New Topline

If there is a correct one among the topK, used the correct one, or else, use the None

### “ALL” Slot

### “Joint” Slot

## Basic Discriminate Model

Model: SVM

3-way Classification:

* None/Output/TopOne

Features:

* Dialog Acts

Data:

Train: train1a

Test: Test1, Test2, Test3, Test4

3-way Classification Results:

Always predicts as “None” -> Majority Classifier