# Summery

## Dialog Act Distribution

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| act | test1 | test2 | test3 | test4 | train1a | train2 | train3 |
| ack | 1868 | 1829 | 453 | 0 | 2177 | 1522 | 189 |
| affirm | 2119 | 2096 | 1922 | 418 | 2500 | 1680 | 1275 |
| are-you-there | 9 | 0 | 0 | 0 | 32 | 17 | 0 |
| bebrief | 55 | 0 | 0 | 0 | 66 | 51 | 0 |
| bye | 30 | 35 | 105 | 61 | 64 | 28 | 115 |
| canthelp.cant\_find\_stop | 54 | 54 | 0 | 77 | 13 | 11 | 0 |
| canthelp.from\_equals\_to | 19 | 22 | 0 | 90 | 35 | 23 | 0 |
| canthelp.no\_buses\_at\_time | 49 | 42 | 0 | 81 | 45 | 53 | 0 |
| canthelp.no\_connection | 82 | 69 | 529 | 141 | 107 | 72 | 180 |
| canthelp.nonextbus | 0 | 0 | 8 | 0 | 0 | 0 | 0 |
| canthelp.route\_doesnt\_run | 79 | 0 | 0 | 31 | 0 | 3 | 0 |
| canthelp.system\_error | 1 | 2 | 0 | 0 | 4 | 4 | 0 |
| canthelp.uncovered\_route | 2 | 0 | 846 | 40 | 176 | 89 | 262 |
| canthelp.uncovered\_stop | 69 | 0 | 123 | 115 | 150 | 134 | 83 |
| didnthear | 0 | 0 | 0 | 60 | 0 | 0 | 0 |
| example | 1478 | 1748 | 10370 | 1003 | 3149 | 2120 | 6812 |
| expl-conf | 5134 | 5281 | 2972 | 801 | 7006 | 4049 | 1892 |
| goback | 0 | 0 | 422 | 0 | 0 | 0 | 313 |
| hello | 791 | 830 | 1020 | 466 | 1147 | 707 | 688 |
| hold-on | 490 | 463 | 0 | 0 | 883 | 650 | 0 |
| impl-conf | 13 | 0 | 4650 | 914 | 1554 | 1007 | 3149 |
| inform | 5121 | 6141 | 6240 | 2138 | 8412 | 4980 | 3604 |
| morebuses | 0 | 0 | 316 | 0 | 0 | 0 | 166 |
| negate | 1457 | 1103 | 735 | 681 | 1831 | 858 | 501 |
| nextbus | 132 | 134 | 1327 | 128 | 507 | 207 | 960 |
| null | 924 | 1038 | 1 | 467 | 1016 | 1156 | 0 |
| open-request | 0 | 0 | 0 | 0 | 1809 | 1246 | 0 |
| please-repeat | 7 | 0 | 0 | 0 | 17 | 40 | 0 |
| please-rephrase | 0 | 0 | 0 | 0 | 1 | 2 | 0 |
| prevbus | 57 | 38 | 128 | 53 | 177 | 128 | 136 |
| repeat | 11 | 18 | 344 | 49 | 27 | 17 | 344 |
| request | 1931 | 4537 | 7289 | 2642 | 3369 | 2119 | 4109 |
| restart | 376 | 197 | 381 | 72 | 523 | 346 | 218 |
| schedule | 336 | 311 | 1544 | 784 | 687 | 501 | 1243 |
| sorry | 0 | 0 | 4279 | 500 | 0 | 0 | 2616 |
| tellchoices | 0 | 0 | 35 | 0 | 0 | 0 | 20 |

Observation:

* Many acts don’t appear in other data set.
* The distribution of acts are different, such as ‘affirm’ and ‘example’

## Train on train2

Train1a is bad

* Many affirm(), but correct slot is not annotated

## Ngram Model

* Ngram: Use the top ASR

## Feature Enrich Model

|  |  |  |
| --- | --- | --- |
| Team | Features | Included |
| Henderson et al. | SLU score | Y |
|  | Rank score | Y |
|  | Affirm score | Y |
|  | Negate score | Y |
|  | Go back score | Y |
|  | Implicit score | Y |
|  | User act type | Y |
|  | Machine act type | Y |
|  | Cant help | Y |
|  | Slot confirmed | Y |
|  | Slot requested | Y |
|  | Slot informed | Y |
| Ren et al. | SLU score and rank of slot | Y |
| Metallinou et al. | rank of the current SLU result | Y |
|  | the SLU result confidence score(s) | Y |
|  | the number of possible past user negations or confirmations of the current SLU result | Y |
| My feature | Ngram | Y |
|  | Word Count | Y |
|  | Duration | Y |
|  | speaking rate | Y |
|  | General Negation | Y |

## Interesting Feature

#### Slot confirmed

If the “affirm” appears in the input, the first “expl-conf” slot-value will be confirmed

If the slot-value ONCE appeared in the Confirmed slots dictionary, it is confirmed.

#### Slot Negated

If the “negation” appears in the input, the first “expl-conf” slot-value will be confirmed.

If the slot-value ONCE appeared in the Negation slots dictionary, it is negated.

#### General Negation

If the slot appeared in the Confirmed slots dictionary but with different value, it is also a negation.

## 3way Classification Results

Train set: train2

Model: SVM

Feature Set: 1) Acts: Binary features for dialog Acts 2) Ngram: top ASR unigram 3) Acts + Ngram 4) Enrich, all features in the table above (42 + # of unigram).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Precision | Recall | F-Measure |
| test1 | acts | 0.804 | 0.794 | 0.798 |
|  | ngram | 0.83 | 0.83 | 0.83 |
|  | act + ngram | 0.832 | 0.841 | 0.835 |
|  | enrich | 0.829 | 0.839 | 0.832 |
| test2 | acts | 0.828 | 0.681 | 0.718 |
|  | ngram | 0.813 | 0.795 | 0.802 |
|  | act + ngram | 0.828 | 0.794 | 0.807 |
|  | enrich | 0.829 | 0.795 | 0.808 |
| test3 | acts | 0.814 | 0.588 | 0.611 |
|  | ngram | 0.748 | 0.747 | 0.734 |
|  | act + ngram | 0.664 | 0.634 | 0.645 |
|  | enrich | 0.687 | 0.629 | 0.65 |
| test4 | acts | 0.871 | 0.736 | 0.769 |
|  | ngram | 0.828 | 0.811 | 0.818 |
|  | act + ngram | 0.85 | 0.816 | 0.828 |
|  | enrich | 0.835 | 0.805 | 0.816 |

## Cost Matrix 3-way model (Based on Act features)

### Distribution of labels

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| data | test1 | test2 | test3 | test4 | train1a | train2 | train3 |
| total | 10085 | 10857 | 13264 | 4754 | 14860 | 9345 | 8639 |
| -1 | 6366 | 7164 | 8793 | 3643 | 10978 | 5571 | 5723 |
| 0 | 2014 | 2085 | 1381 | 316 | 160 | 1507 | 875 |
| 1 | 1449 | 1332 | 2431 | 661 | 3093 | 1706 | 1756 |

### Cost Matrix:

|  |  |  |  |
| --- | --- | --- | --- |
| Classified as -> | -1 | 0 | 1 |
| -1 | 0 | 1 | 1 |
| 0 | C | 0 | 1 |
| 1 | C | 1 | 0 |

C = [1, 2, 5]

### Result:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Precision | Recall | F-Measure |
| cost matrix | cm1 | 0.799 | 0.812 | 0.804 |
|  | cm2 | 0.813 | 0.796 | 0.803 |
|  | cm5 | 0.804 | 0.791 | 0.796 |
|  | cm23 | 0.813 | 0.796 | 0.803 |

## Combined Model

3-way model is just the first step

The second step is to the label predicted by the 3-way model and then do dialogue state tracking.

Given the predicted rank, here is how to the SLU:

* If -1, None
* If 0, get the first “expl-conf” slot-value in the output
* If 1, take the first SLU

### Result

The 3-way model is based on Ngram

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | joint | | | | all | | | |
|  |  | test1 | test2 | test3 | test4 | test1 | test2 | test3 | test4 |
| schedule1 | baseline | 0.1455 | 0.1394 | 0.4732 | 0.1775 | 0.7748 | 0.7928 | 0.9178 | 0.8233 |
|  | nohistory | 0.164 | 0.1508 | 0.4572 | 0.1845 | 0.7878 | 0.7921 | 0.9156 | 0.827 |
|  | majoritybaseline | 0.2122 | 0.1723 | 0.0815 | 0.2406 | 0.8162 | 0.7971 | 0.772 | 0.8434 |
|  | 3-way | 0.4759 | 0.5308 | 0.1823 | 0.3231 | 0.9162 | 0.9205 | 0.8708 | 0.8804 |
|  | topline | 0.8724 | 0.9289 | 0.8991 | 0.8597 | 0.985 | 0.9919 | 0.9887 | 0.9835 |
| schedule2 | baseline | 0.1467 | 0.1399 | 0.4475 | 0.1572 | 0.602 | 0.4905 | 0.6202 | 0.5841 |
|  | nohistory | 0.1655 | 0.1511 | 0.4281 | 0.1629 | 0.6371 | 0.4967 | 0.6065 | 0.5906 |
|  | majoritybaseline | 0.2143 | 0.1729 | 0.1084 | 0.2485 | 0.7056 | 0.5267 | 0.259 | 0.6585 |
|  | 3-way | 0.4748 | 0.5287 | 0.2065 | 0.3216 | 0.8236 | 0.7619 | 0.5034 | 0.7482 |
|  | topline | 0.8721 | 0.9282 | 0.8844 | 0.8562 | 0.9707 | 0.9743 | 0.9131 | 0.9605 |
| schedule3 | baseline | 0.1043 | 0.1409 | 0.4923 | 0.1162 | 0.5982 | 0.4869 | 0.7033 | 0.6396 |
|  | nohistory | 0.1176 | 0.145 | 0.4675 | 0.1294 | 0.6108 | 0.479 | 0.6898 | 0.6432 |
|  | majoritybaseline | 0.1376 | 0.168 | 0.0616 | 0.1162 | 0.6255 | 0.479 | 0.12 | 0.6234 |
|  | 3-way | 0.5006 | 0.6274 | 0.1498 | 0.2127 | 0.885 | 0.8581 | 0.5219 | 0.733 |
|  | topline | 0.8679 | 0.9512 | 0.9178 | 0.8421 | 0.9777 | 0.9881 | 0.9688 | 0.9653 |

Conclusion:

* 3-way is much better the baseline on test1, test2, test4, especially on “joint”
* 3-way is bad on test3 (The performance of 3-way on test3 is bad)