Conversational Analytics

Sept 2022

Areas of Focus

- Basic time-series analyses: Traffic over defined periods of time
  - % volume of traffic by Intent / keywords / channel / etc.
  - Average confidences of predicted intents and actions
- Topic Modeling
  - Overall corpus
  - By intent
- NLU/conversation review
  - Layered keyword filtering / text clustering / word clouds
  - Model classification diffs
  - Dropoff intents/actions (where do conversations end)
- Cross-Validation reviews (from Rasa's CLI tool)
  - Individual reports
  - Comparison reports

Time Series

By intent traffic

- ➔ Prioritize by customer use
- ➔ Gauge responses to ad campaigns
- ➔ NLU updates properly re-directing

By average intent confidences

- ➔ Review NLU updates
- ➔ Identify intents potentially needing NLU sample improvements

Ad campaign response

Same ad campaign caused customers to use unexpected phrasing which also caused a drop in confidence levels of a related intent

Topic Modeling

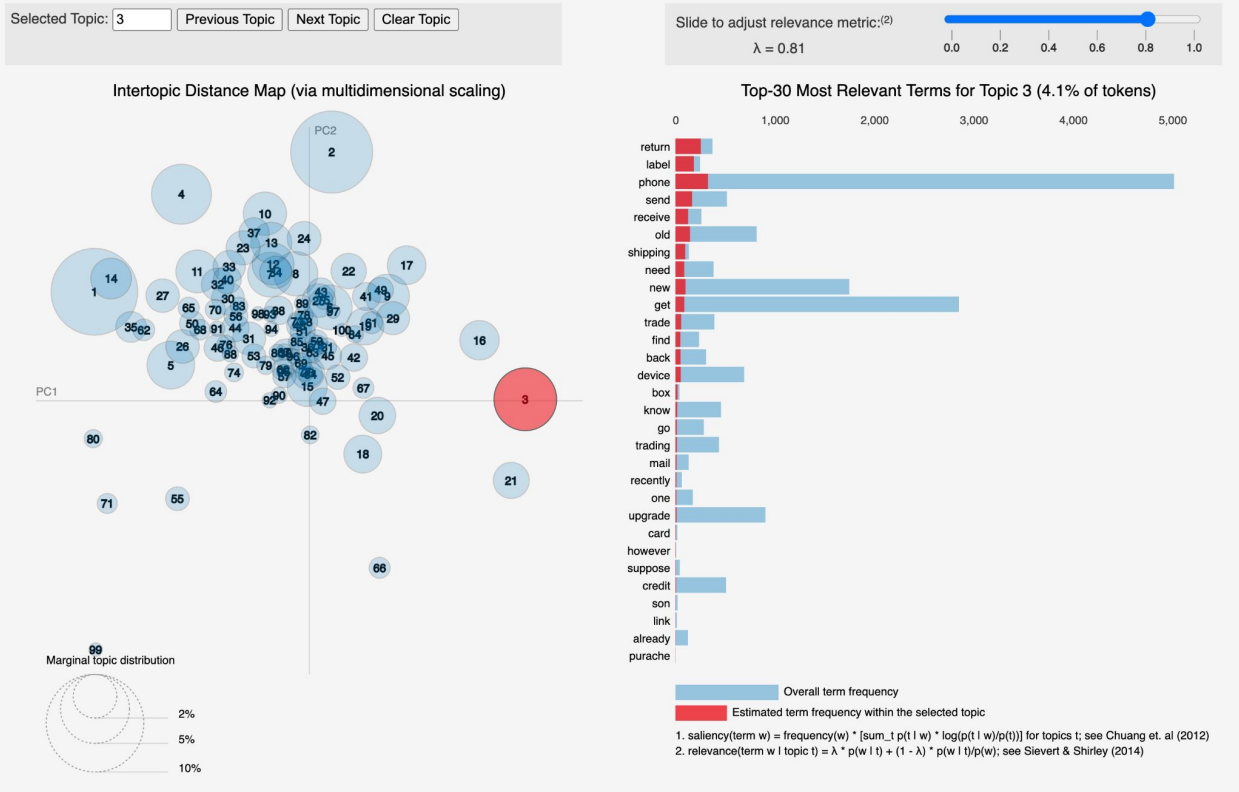
Lemmatization and other processing methods will have a major impact on this.

Overall corpus

- ➔ Rough outline of potential services to provide
- ➔ Identify common terminology within and between intents
- ➔ Observe shifts by news cycles, ad campaigns, etc

By intent traffic

- ➔ Identify potential new story branches or intents - depending on grouping (dis)similarities
- ➔ Discover phrasing and terminology to use in representative samples (eg. keywords)



Example topic model on the specific intent of “returns”: identifies return labels as a major request

NLU / Conversation review

NLU review/Sample Gathering

- Keyword filtering - determine % volumes of traffic with specific terminology to represent appropriately within NLU samples
  - “and”/ “or” pairings
  - Synonyms / slang
  - Word clouds/counts
- Dropoff intents/actions
  - What was the last intent in the conversation, and was it served properly
  - Was the last action taken by the bot ignored and never responded to?
  - Broken actions/story pathing
- Model Classification changes
  - Using utterances from conversations, classified by the current model, have the new model iterate through the utterances and re-classify them and compare the similarities/differences

text	intent	intent_prediction	confidence
It's not letting my choose October 15th	pa_change_dates	broken	0.924

*Fig 9: Report Intent Errors table from “*label *errors.pdf”*

intent	f1-score	precision	recall	samples	confused_with
pa_create	0.654	0.607	0.708	96	'pa_cannot_pay': 8 'pa_change_dates': 5

*Fig 10: Report Intent Accuracy summary table from “*label *figs.pdf”*

intent	confidence	count
line_add_byo	0.892	11
intent_prediction	**confidence**	**count**
affirm	0.902	28

Fig 13: tables for intent error confidence averages

intent	intent_prediction	count	confidence
out_of_scope	affirm	21	0.876

Summary of classifications

Cross-Validation Output Review

Individual Reports

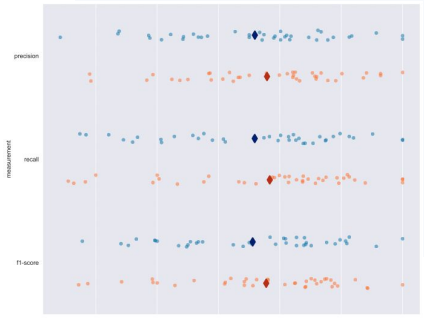
- NLU classification errors: review “missed” predictions for common terminology
- Confusion matrix of commonly mixed intents
- f1/precision/recall by intent correlated with number of samples and “commonly confused with…” from confusion matrix

Comparison Reports

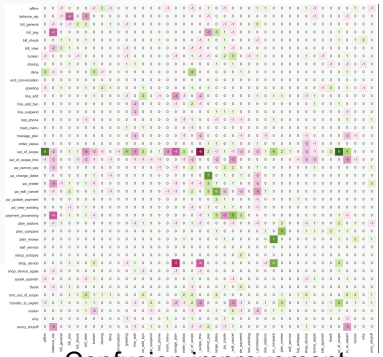
- Performance shifts
- NLU prediction differences
- Confusion Improvement Matrix

intent	intent_prediction	count	confidence
out_of_scope	affirm	21	0.876
pa_edit_cancel	pa_change_dates	18	0.897

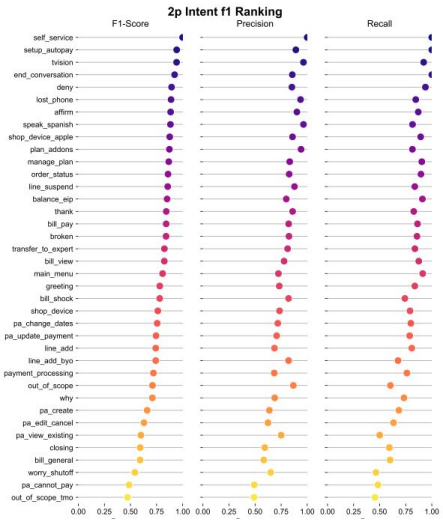
Commonly confused intents with a high average confidence during misclassification may indicate NLU samples need cleaning up



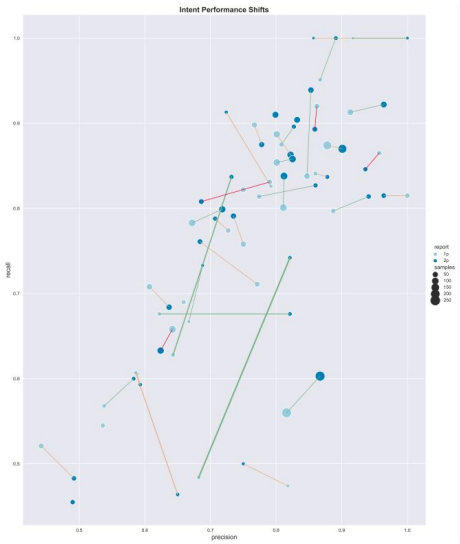
Performance mean comparison



Confusion improvement



Individual report performances



Performance shifts