Using Past Speaker Behavior to Better Predict Turn Transitions

Tomer Meshorer Peter A Heeman

Center for Spoken Language Understanding
Oregon Health & Science University, Portland, Oregon,USA

Interspeech, 2016

Outline

Problem

- For a natural conversation between human and machine, we want to conform to human to human turn taking system (Sacks et al, 1978)
- In Human-Human conversations conversant predict (Sacks et al, 1978) or signal (Duncan 1972) each other on coming turn transition
- Timeouts leads to poor user interaction(Arsikere et al, 2015)
    - Not effective in noisy environment
    - too little - machine barge in during intra turn pause.
    - too much - user waiting for the machine.
- Turn transition prediction based on local features improve turn taking but still do not match human performance.
    - Syntactic (Sacks et al 1978,De Ruiter et al. 2006)
    - Prosodic (Ford 1996,Stolcke 2002,Ferrer 2003)
    - Pragmatic (Ford 2001)

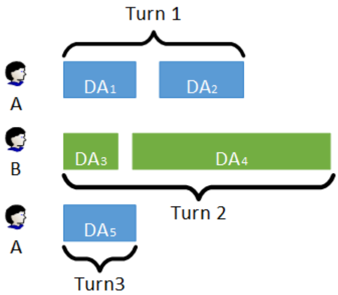
Goal of the work

Conversant's past behavior can help predict turn transitions

Past behavior represented by Summary features

Outline

Conversation



Conversation

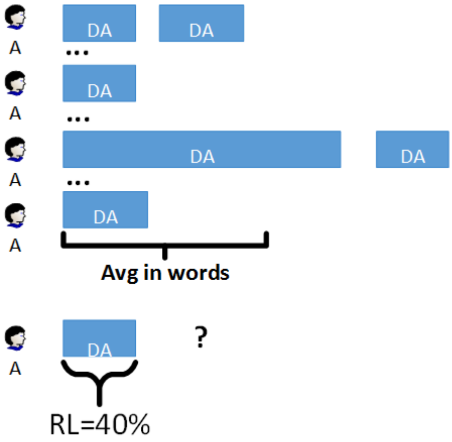
$\ldots A, DA1, A, DA2, B, DA3, B, DA4, A, DA5 \ldots$

Conversation with turn change

$\ldots DA1, 0, DA2, 1, DA3, 0, DA4, 1, DA5 \ldots$

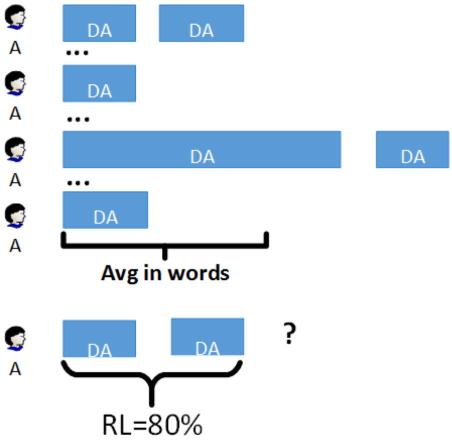
Relative Turn Length

Measure ratio of current turn length relative to average turn length.



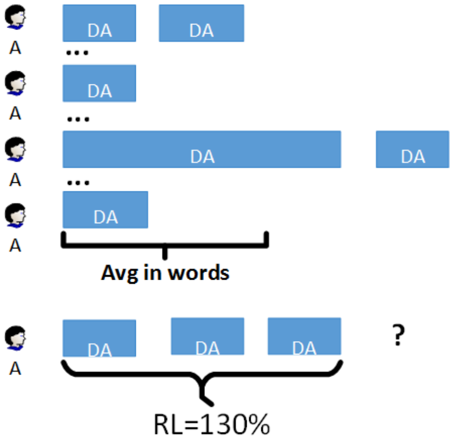
Relative Turn Length

Measure ratio of current turn length relative to average turn length.



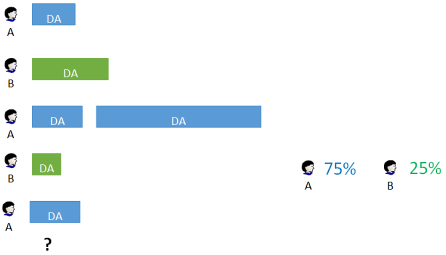
Relative Turn Length

Measure ratio of current turn length relative to average turn length.



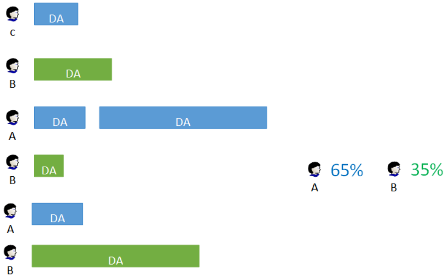
Relative Floor Control

Measure ratio that current speaker held the floor.
How dominate current speaker is.



Relative Floor Control

Measure ratio that current speaker held the floor.
How dominate current speaker is.



Outline

Preprocessing

- Removed 11 dialogue acts that were coded as 'other' in switchboard.
- Skip the first 120 seconds of the conversation.
  - Gives time for conversant to form the conversional image.
  - Reduces the dialogue acts from 50633 to 37508.
- Reduce data sparsity by collapsing 65 dialog acts into 9.

Switchboard dialog acts	Dialog act classes
sd,h,bf	statement
sv,ad,sv@	statement - opinion
aa,aar̂	agree accept
%.%-,%@	abandon
b,bh	backchannel
qy,qo,qh	question
no,ny,ng,arp	answer
+	+
o@,+@	NA

ML Classifiers

- Used random forests (N=200) to train and test the following models
  - Baseline 1: current dialog act label.
  - Baseline 2: current and previous dialog acts.
  - Summary model: just the summary features.
  - Full model: summary features and current and previous dialog acts.
- Evaluation was done using 10 fold cross validation.
- Run grid search to find the optimal hyper parameters.

Outline

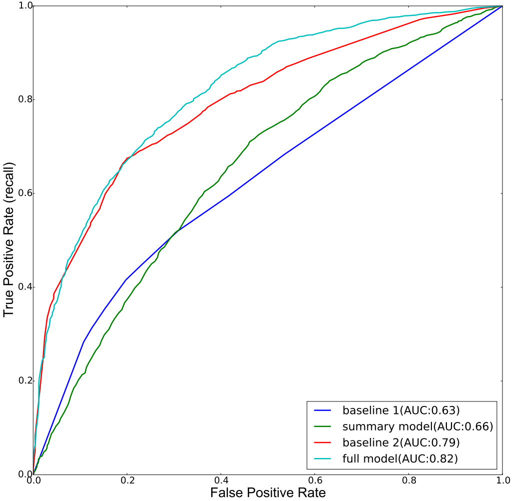
Accuracy

Model	Accuracy	AUC	hyper parameters
Baseline 1	60.26%	0.63	max_features=sqrt, max_depth=7
Baseline 2	74.43%	0.79	max_features=log2, max_depth=9
Summary	66.14%	0.65	max_features=sqrt, max_depth=5
Full	76.05%	0.82	max_features=10, max_depth=9

Table: Accuracy, Area under the curve

- The Summary model is more accurate than Baseline 1.
- The Full model is more accurate than Baseline 2.

ROC curves and AUC of the different models



Precision & Recall

Model	Precision	Recall	F1
Baseline 1	69.49%	45.52%	54.97%
Baseline 2	80.38%	68.80%	74.08%
Summary	64.55%	68.88%	66.42%
Full	76.17%	77.25%	74.87%

Table: Precision, recall and F1 results

- Summary model makes more mistakes vs Baseline 1, however it detects more turn transition.
- Summary model makes more mistakes vs Baseline 2
- Full model makes slightly more mistakes vs Baseline 2, however it detects more turn transitions.
  Overall F1 is slightly better.

Outline

Conclusion and Future Work

- Conclusion
  - Experiment proved that summary features improve turn transition prediction
- Future Work
  - Combine summary features with other local features: syntax, prosody.
  - Test simple moving average windows (5,10,20 turns)
  - Test exponential moving average.
  - Convert other local features to summary feature.