Intention and Face In Dialog

Adil Soubki and Owen Rambow

Stony Brook University

iACS INSTITUTE FOR ADVANCED COMPUTATIONAL SCIENCE

Face

- B&L define face as “The public self-image that every member [of a society] wants to claim for himself” which comes with two related aspects.
  - Negative Face - The basic claim to territories, personal preserves, rights to non-distraction - i.e. to freedom of action and freedom from imposition.
  - Positive Face - The positive consistent self-image or ‘personality’ (crucially including the desire that this self-image be appreciated and approved of) claimed by interactants.
- These may alternatively be thought of in terms of wants.
  - Negative Face - The want of every ‘competent adult member’ that his actions be unimpeded by others.
  - Positive Face - The want of every member that his wants be desirable to at least some others.

Face Acts (FAs)

- Brown & Levinson propose, analogous to speech acts, that some utterances (called face acts) inherently interact with the face of discourse participants.
  - E.g., compliments, insults, requests, etc…
- In this manner, utterances may raise (+) or threaten (-) the positive (Pos) or negative (Neg) face of hearers (H) or speakers (S).

Face Acts (FAs)

Face Act	Interpretation	Example Discourse Goals
HNEG-	Imposition	Requests, commands, questions, offers, promises, ...
HPOS-	Disagreement	Criticism, insults, disapproval, ...
HNEG+	Permissiveness	Granting permission, making exceptions, ...
HPOS+	Agreement	Seeking common ground, group cohesion, ...
SNEG-	Indebtedness	Thanking, accepting offers or thanks, commitments, ...
SPOS-	Apologies	Confessions, embarrassment, ...
SNEG+	Autonomy	Refusing requests, asserting freedoms, ...
SPOS+	Confidence	Self-promotion, signaling virtue, ...

Face Act Tagging

- The idea is to operationalize this theory of politeness starting with one of its core concepts - the face act
- A text classification task first introduced by Dutt, Joshi, and Rosé (2020)
  - Input - Some sequence of tokens
  - Output - One of the 8 possible face acts (or no face act - Other)

Will you make a donation to Save the Children? [HNeg-]

Albany is the capital of New York [Other]

Wang et al. (2019)

- Introduce a corpus of dyadic, persuasion-oriented conversations sourced from an online task where Amazon Mechanical Turk workers must convince their addressee to donate part of their task earnings to a charity, Save the Children.
- The conversations are carried out through a chat interface with one worker acting as the persuader (ER) and the other as the persuadee (EE).
- The participants were informed that the dialog must last at least 10 turns and that their reward is not penalized should they fail to convince their partner to donate.

Dutt et al. (2020)

- Augment conversations from Wang et al. (2019) with FA annotations
- Takes some small departures from politeness theory
  - Thanking is HPos+, Other to indicate no face act
- Why select this corpus for annotation?
  - (1) Goal oriented nature necessitates FTAs, which are normally avoided
  - (2) Both participants are on equal ground which mitigates issues of power/social distance
- Allowed for multiple face acts per utterance when annotating (as theory suggests)
  - Noted that this occurred in only 2% of cases and simplified to one label per utterance
- Annotated 10,716 turns averaging 10 words (or 51 characters) in length across 296 unique conversations
- Train models on this data to predict face acts (and if someone donates)

CMU Face Act Dataset

- Note that SNeg- never happens
- Highly imbalance label distribution

	speaker	utterance	face_act
1645	ER	Have you thought about giving to the Save the Children charity?	hneg-
1646	EE	I used to donate to them.	spos+
1647	EE	Do you like that charity?	hpos-
1648	ER	I think that they are good, they help a lot of children	spos+
1649	EE	What do they do exactly?	hneg-
1650	ER	They work for the rights of children in developing countries in safety, health, education	spos+
1651	EE	That is super important work, obviously.	hpos+
1652	EE	Do you know if they are a trustworthy organization?	hpos-
1653	ER	Yes they are very trustworthy, and they have been around for a long time	spos+
1654	EE	I can't stand charities where the top level execs make a ton of money	hpos-
1655	ER	I agree, that doesn't happen with this charity, most of the money goes to helping the children	spos+
1656	EE	that's good to hear.	hpos+
1657	EE	they sound like a good charity to donate to	hpos+
1658	ER	I would highly recommend it	spos+

Face Act	Count
Other	4,300
HPos+	2,844
SPos+	1,589
HNeg-	1,073
HPos-	334
HNeg+	305
SNeg+	259
SPos-	12
SNeg-	0

What's missing?

- Performance on the task is lower than one might expect
  - 0.69 accuracy and 0.60 macro-F1
- Brown & Levinson situate politeness in the broader context of rational interaction (e.g., Grice, 1975)
  - This seems disconnected from the sequence classification task

Our Goals

1. Improve task performance to make the model a more viable tool for computational social science
2. Use experiments to explore implications of politeness theory

Sequence-to-Sequence Classification

- Flan-T5-base with 2 turns prior and target turn as input, and the label for the last turn as output
  - Tried a bunch of different ideas here and this worked best.
- No dev set in original splits so I train micro F1 for early stopping (usually 15-20 epochs)
- Use edit-distance and train frequency to correct any malformed labels generated.
  - Zhang et al. (2021)

[Input]
ER: Are you interested in donating?
EE: Possibly, I'm not sure.
EE: I don't even know what the charity is.

[Output]
sneg+

Evaluation

- Like prior work, we use micro and macro F1 as the primary evaluation metrics
- Note that rarity of SPos+ contributes to fluctuations in results for macro
- Use same 5 fold cross validation splits as Dutt et al. (2020)
    - Report average values.

Face Only System (FOS) Results

- 3 point boost to Macro F1
- 4 point boost to Micro F1
- Correlated (r = 0.77) with count so struggles on minority classes
- Nothing too exciting.

	F1	F1	Prec.	Recall	Count
Macro	0.60	0.63	0.63	0.63	-
Micro	0.69	0.73	0.73	0.73	-
OTHER	-	0.75	0.76	0.73	4,300
HPOS+	-	0.75	0.72	0.77	2,844
SPOS+	-	0.74	0.74	0.75	1,589
HNEG-	-	0.74	0.71	0.76	1,073
HPOS-	-	0.55	0.61	0.51	334
HNEG+	-	0.44	0.47	0.41	305
SNEG+	-	0.57	0.61	0.53	259
SPOS-	-	0.47	0.39	0.58	12
Dutt et al. (2020)			Fos		

Face Acts and Intention

- Some observations regarding intention and face.
  - There is a sense in which FAs are ambiguous without a notion of intent.
    - E.g., Admitting a mistake does not threaten speaker face if it's intent is understood to show a virtue (such as humbleness)
  - One does not risk face without a goal in mind.
- Maybe incorporating a notion of intent will improve FA model performance.
- How can we do this though?

Dialog Acts (DAs)

- Speech Act Theory was adapted to NLP in the form of dialog act tagging.
  - DAs can be thought of as a proxy for intention.
  - A well-established task
  - Many existing datasets
- In dialog act tagging, utterances are assigned labels from an existing inventory based on the speech act(s) they perform
  - The development of these inventories was challenging but is now mature.
  - Core & Allen (1998) created the most common framework, DAMSL.
- We use Switchboard and MRDA which are both transcribed from spoken dialog
  - MRDA we use a coarse tag set.
  - SWDA we use a detailed one.

Shriberg et al. (2004)
MRDA

BackChannel
Disruption
FloorGrabber
Question
Statement

Stolcke et al. (2000)
SWDA

Acknowledge (Backchannel)
Action-directive
Affirmative non-yes answers
Agree/Accept
Appreciation
Backchannel in question form
Conventional-closing
Conventional-opening
Hedge
Hold before answer/agreement
No answers
Non-verbal
Other
Other answers
Response Acknowledgement
Segment
Statement-non-opinion
Statement-opinion
Wh-Question
Yes answers
Yes-No-Question

Integrating Dialog Acts

Method 1 (Text Augmented)

- Train DA system of He et al. (2021) on MRDA and SWDA and then use it on the FA data.
- Add the annotations directly to the input.

```
[Input]
  ER: Are you interested in donating?
  (Yes-No-Question)
  EE: Possibly, I'm not sure. (Hedge)
  EE: I don't even know what the charity is.
  (Statement-non-opinion)

[Output]
  sneg+
```

Method 2 (Multitask Learning)

- "Traditional" MTL
- Prefix each task (since using T5)
  - dialog act: {3 turns of dialog}
  - face act: {3 turns of dialog}

```
[Input]
     face act:
     ER: Are you interested in donating?
     EE: Possibly, I'm not sure.
     EE: I don't even know what the
     charity is
[Output]
     sneg+

[Input]
     dialog act:
     ER: Are you interested in donating?
     EE: Possibly, I'm not sure.
     EE: I don't even know what the
     charity is
[Output]
     Statement-non-opinion
```

Results

- The DAs don't help?
- May be improving aspects of minority classes
  - MRDA (coarser) might help with precision (⅗)
  - SWDA (detailed) might help with recall (⅗)
- Friedman Rank Sum Test finds differences significant
- Definitely helped with HNeg+
- Something is going on
  - Not totally clear what

Model	F1
FOS	0.73
TA-SWDA	0.70
TA-MRDA	0.70
MTL-SWDA	0.70
MTL-MRDA	0.71
Dutt et al.	0.69

	HNeg+		
	F1‡	Prec.†	Recall‡
Fos	0.44	0.47	0.41
Ta-Swda	0.49	0.44	0.56
Ta-Mrda	0.51	0.47	0.55
Mtl-Swda	0.41	0.46	0.37
Mtl-Mrda	0.43	0.50	0.39

Error Analysis

- We didn’t get the result we were expecting
- More analysis can be found in the paper
- Goals
  - What is FOS doing?
  - Is there signal in the DA data?
  - If there is, does it look like the DA-augmented FTA systems are using it?

Error Categories

- Lots of golds errors
- Far more multilabel utterances than previously reported
- Confusion matrix shows Other to be the main issue.

Conclusion: Need to improve the annotation.

Error	Count (Pct)
Both Happening (Same Part)	29 (16%)
Both Happening (Diff. Part)	17 (9%)
Gold Error (Correct)	33 (18%)
Gold Error (Incorrect)	8 (4%)
True for Previous	18 (10%)
Predicted Other	55 (30%)
No Idea	20 (11%)
Total	180

Do DAs Actually Have Signal?

- There are some nice correlations.

	SPos+	HPos+	SPos-	HPos-	SNeg+	HNeg+	HNeg-	Other
BackChannel	-0.02	0.01	-0.00	-0.01	-0.01	-0.01	-0.01	0.02
Disruption	0.03	0.01	-0.00	-0.00	0.02	-0.01	-0.04	-0.00
FloorGrabber	-0.01	-0.02	-0.00	-0.01	-0.01	-0.01	-0.01	0.04
Question	-0.19	-0.24	-0.02	0.02	-0.07	-0.05	0.49	0.09
Statement	0.14	0.20	0.02	-0.02	0.04	0.05	-0.38	-0.08
Acknowledge (Backchannel)	-0.03	0.00	-0.00	-0.02	-0.01	-0.00	-0.03	0.05
Action-directive	-0.01	-0.01	-0.00	0.00	-0.01	-0.01	0.04	-0.00
Affirmative non-yes answers	0.00	0.03	-0.00	-0.01	-0.01	-0.01	-0.01	-0.02
Agree/Accept	-0.02	0.15	-0.00	-0.02	-0.02	-0.02	-0.04	-0.07
Appreciation	-0.12	0.29	0.01	-0.05	-0.05	-0.05	-0.10	-0.06
Backchannel in question form	-0.01	-0.01	-0.00	-0.01	-0.01	-0.01	-0.00	0.03
Conventional-closing	-0.09	0.06	0.01	-0.04	-0.03	-0.03	-0.07	0.09
Conventional-opening	-0.01	-0.02	-0.00	-0.01	-0.00	-0.00	-0.01	0.04
Hedge	-0.01	-0.01	-0.00	-0.00	0.02	-0.00	-0.01	0.01
Hold before answer/agreement	-0.02	-0.01	-0.00	-0.01	-0.01	-0.01	-0.02	0.04
No answers	-0.02	0.00	-0.00	0.03	0.01	-0.01	-0.02	0.01
Non-verbal	-0.06	0.00	0.03	-0.02	0.02	0.00	-0.03	0.05
Other	-0.03	-0.05	-0.00	-0.01	-0.01	-0.01	-0.03	0.10
Other answers	-0.00	-0.01	-0.00	-0.00	-0.00	-0.00	-0.00	0.01
Response Acknowledgement	-0.01	-0.00	-0.00	-0.01	-0.01	-0.01	-0.01	0.02
Segment	0.07	0.01	-0.01	-0.00	0.02	0.02	-0.05	-0.04
Statement-non-opinion	0.30	-0.10	0.01	0.01	0.12	0.08	-0.21	-0.07
Statement-opinion	-0.03	0.13	-0.01	0.05	-0.03	0.02	-0.10	-0.05
Wh-Question	-0.14	-0.18	-0.01	0.02	-0.05	-0.05	0.35	0.07
Yes answers	0.01	0.01	-0.00	-0.01	-0.00	-0.01	-0.03	0.01
Yes-No-Question	-0.12	-0.16	-0.01	0.01	-0.05	-0.03	0.31	0.06

Do the DA Systems Use it?

- For HNeg+... it sure looks like it.
- These are instances where HNeg+ was in the predictions or gold label for the TA-MRDA system.
  - Best system for HNeg+
- HNeg+ are largely statements.
- When the DA system changes something to a TP, it is almost always a statement.
- When the DA system changes something to a TN, it is over-proportionally a question

Dialog Act	All	HNeg+	FN to TP	FP to TN
Statement	80%	93%	97%	78%
Question	20%	7%	3%	22%
Number		305	64	50

Why Doesn’t this Work Better?

(1) Dialog acts were not providing very much orthogonal information in training (i.e. FOS already learned to distinguish these)

(2) These methods of integration were not effective for this task.

(3) The data is not so good.

Thanks (SNeg-)