Opinions in Interactions : New Annotations of the SEMAINE Database

Valentin Barriere[1], Slim Essid[2], Chloé Clavel[2]

[1]European Commission's Joint Research Center - Ispra, [2]Télécom Paris

Overview

- We present the process to collect new annotations of opinion over the multimodal corpus SEMAINE database composed of dyadic interactions [1]
  - Using interactional context by seeing and annotation the whole conversation
  - Using multimodal context by reading the text and listening to the audio recording at the same time
- We propose a baseline for the detection of opinions in interactions, reaching a F1 of 0.72.

SEMAINE Database

- 79 sessions, composed of 5,627 speech turns and 74k words, corresponding to 6h20 of dyadic interactions
- Between an user and an operator playing the role of a virtual agent
- Emotionally colored conversations annotated in a continuous way in *Valence*, *Arousal*, *Power* and *Expectancy*

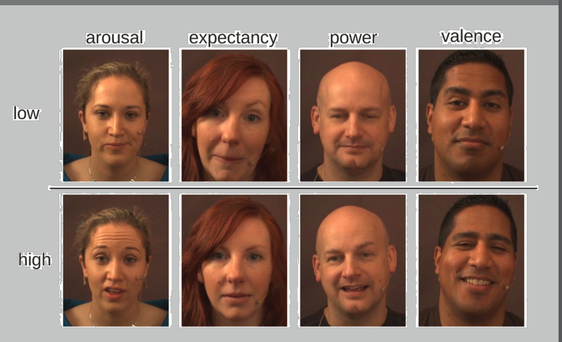


Figure 1:Examples of *Arousal*, *Expectancy*, *Power* et *Valence* from [2]

Annotation Platform: integrate interactional/multimodal context

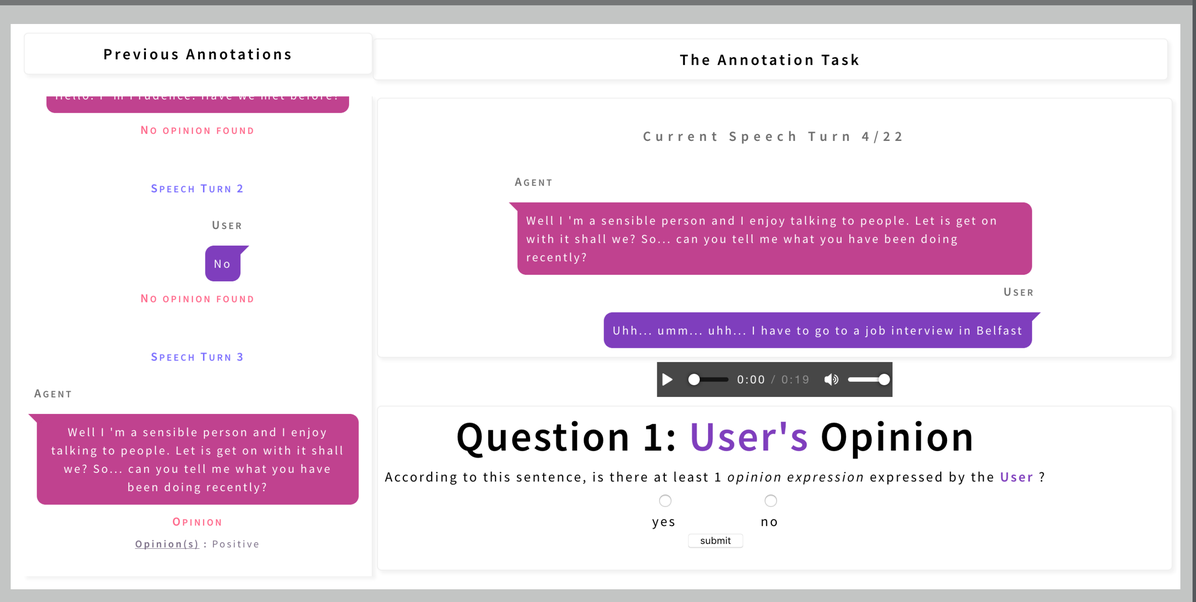


Figure 2:Screenshot of the online annotation platform

- Homemade php annotation platform
- Dialogic context: Conversational history and a contextual Adjacency Pair (pair of speech turns)
- Multimodal context: Audio and Text were aligned using [3] in order to use both the modalities to annotate

Annotation Platform: minimize the cognitive load

- Preliminary phase to annotate a dummy conversation
- Incrementally complex questions in order to reduce the cognitive load
- Special attention on mixed opinions cases to increase inter-annotator agreement opinions

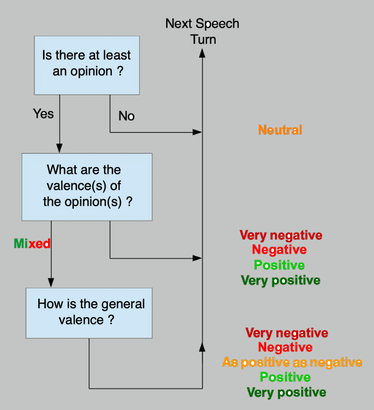
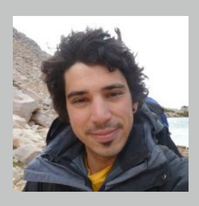


Figure 3:Annotation schema

Contact Information

- Email: valentin.barriere@ec.europa.eu
- Phone: +39 033278-3859



Annotations – Inter Annotator Agreement & Aggregation

- Given the low inter-annotator agreement using 4 classes, unreliable to use *Mixed* label as a fourth class
- Using the prominent opinion allows to reach higher IAA.

Speaker	α 3 classes μ	σ^2	min	max	med	Total
Agent	60.4	16.1	34.0	100	60.5	∅
User	54.9	15.0	13.3	85.9	56.0	∅
All	60.9	11.0	39.1	90.6	58.1	66.3

Table 1:Krippendorff's α per discussion using the Prominent label

Speaker	# Opinions by Disc. μ	σ^2	Opinions by ST (%) Neg.	Pos.	Oth.
Agent	15.22	7.71	22.84	21.14	56.02
User	17.90	8.98	18.47	33.75	47.78
All	33.12	15.14	20,66	27.42	51.92

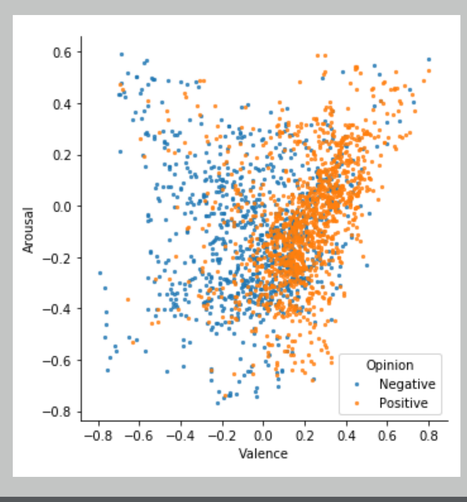
Table 2:Opinion per speech turn obtained after aggregation using majority vote

Correlations between labels

Emotion	μ (σ^2) Neg.	Pos.	Oth.	*Pearson Corr*
Valence	-3.6 (7.0)	18.7 (4.3)	11.9 (5.9)	28.9
Arousal	-11.5 (6.4)	-6.1 (5.8)	-7.9 (5.5)	7.4
Dominance	39.0 (5.2)	46.7 (3.2)	39.0 (5.2)	12.4
Surprise	33.8 (0.9)	31.4 (0.9)	32.7 (1.1)	-8.7

Table 3:Statistics between the emotion-related and the opinion annotation

- There is a correlation between Valence and Opinion.
- Positive opinions are more likely to have high valence and high arousal
- Negative opinions are more likely to have low valence and low arousal
- Positive opinions are more dominant than neutral and negative ones.



Baseline

- We provide a multimodal baseline, enhancing SEMAINE's state-of-the-art DialogueRNN model [4] using:
  - RoBERTa [5] speech turn embeddings as textual features
  - ComparE feature set [6] as audio feature

Method	Valence *MAE*	*r*	Arousal *MAE*	*r*	Opinion F1	Acc
DialogueRNN	0.171	0.37	0.164	0.60	49.56	51.31
Our model	**0.132**	**0.76**	**0.154**	**0.71**	**72.08**	**72.20**

Table 4:Baseline results on different tasks

Conclusion & Future Work

- Collected opinion annotations per speech turns on SEMAINE
- Rich in opinions: 48.08% of the speech turns
- Can be used with continuous emotional annotations of the AVEC-2012 challenge of [6]

References

[1] Gary McKeown, Michel Valstar, Roddy Cowie, Maja Pantic, and Marc Schröder.
The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent.
IEEE Transactions on Affective Computing, 3(1):5–17, 2012.

[2] Martin Wöllmer, Moritz Kaiser, Florian Eyben, Björn Schuller, and Gerhard Rigoll.
LSTM-modeling of continuous emotions in an audiovisual affect recognition framework.
Image and Vision Computing, 31(2):153–163, 2013.

[3] R M Ochshorn and M Hawkins.
Gentle forced aligner, 2017.

[4] Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria.
DialogueRNN: An Attentive RNN for Emotion Detection in Conversations.
2018.

[5] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov.
RoBERTa: A Robustly Optimized BERT Pretraining Approach.
(1), 2019.

[6] Björn Schuller, Michel Valstar, Florian Eyben, Roddy Cowie, and Maja Pantic.
AVEC 2012 - The continuous audio/visual emotion challenge.
In *ICMI'12 - Proceedings of the ACM International Conference on Multimodal Interaction*, pages 449–456, 2012.