Modeling the Quality of Dialogical Explanations

Milad Alshomary, Felix Lange, Meisam Booshehri, Meghdut Sengupta, Philipp Cimiano, Henning Wachsmuth

Leibniz Universität Hannover

Motivation

- Explanations play a significant role in our daily lives → Often as dialogus
- Learning how Humans construct Explanations is insightful to Explainable AI
- No work studied the success of daily-life explanation dialogues

Contributions

- A corpus of daily-life explanations
- A comparison with expert dialogues
- A study of the automatic assessment of dialogue success

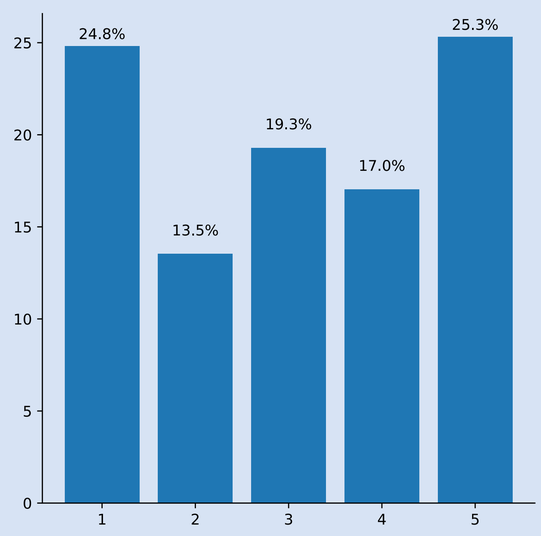
Corpus Constrution

- Threads collected from CMV Subreddit
  - 399 dialogues spanning between 2019 and 2021
  - 3457 turns, with average of 8.7 turns per dialogue
- Annotated by three annotators following the work of *Wachsmuth and Alshomary 2022*

Turn-labels distribution in our corpus

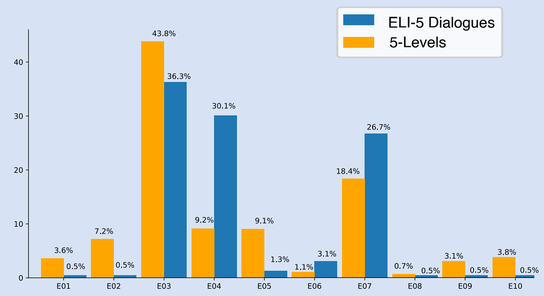
Turn Label		Train #	Train %	Test #	Test %
(t_1)	Main topic	1411	51.7	336	46.1
(t_2)	Subtopic	517	19	94	12.9
(t_3)	Related topic	346	12.7	130	17.8
(t_4)	No/Other topic	454	16.6	169	23.2
(e_1)	Test understanding	12	0.4	5	0.7
(e_2)	Test prior knowledge	13	0.5	4	0.5
(e_3)	Provide explanation	1012	37.1	244	33.5
(e_4)	Request explanation	823	30.2	217	29.8
(e_5)	Signal understanding	36	1.3	9	1.2
(e_6)	Signal non-underst.	85	3.1	23	3.2
(e_7)	Provide feedback	711	26.1	213	29.2
(e_8)	Provide assessment	14	0.5	4	0.5
(e_9)	Provide extra. Inf.	13	0.5	3	0.4
(e_{10})	Other	9	0.3	7	1
(d_1)	Check question	113	4.1	32	4.4
(d_2)	What/How question	349	12.8	83	11.4
(d_3)	Other question	462	16.9	118	16.2
(d_4)	Confirming answer	87	3.2	29	4
(d_5)	Disconfirming answer	105	3.8	21	2.9
(d_6)	Other answer	252	9.2	70	9.6
(d_7)	Agreeing statement	192	7	79	10.8
(d_8)	Disagreeing statement	364	13.3	86	11.8
(d_9)	Informing statement	733	26.9	184	25.2
(d_{10})	Other	71	2.6	27	3.7

Distribution of dialogue quality

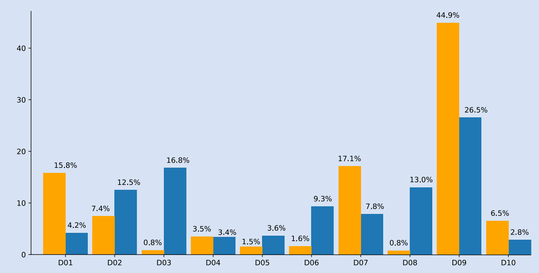


Analysis

- Comparison of explanation moves and dialogue acts between daily-life dialogues (our corpus) and expert dialogues



Explanation Move Distribution



Dialogue act Distribution

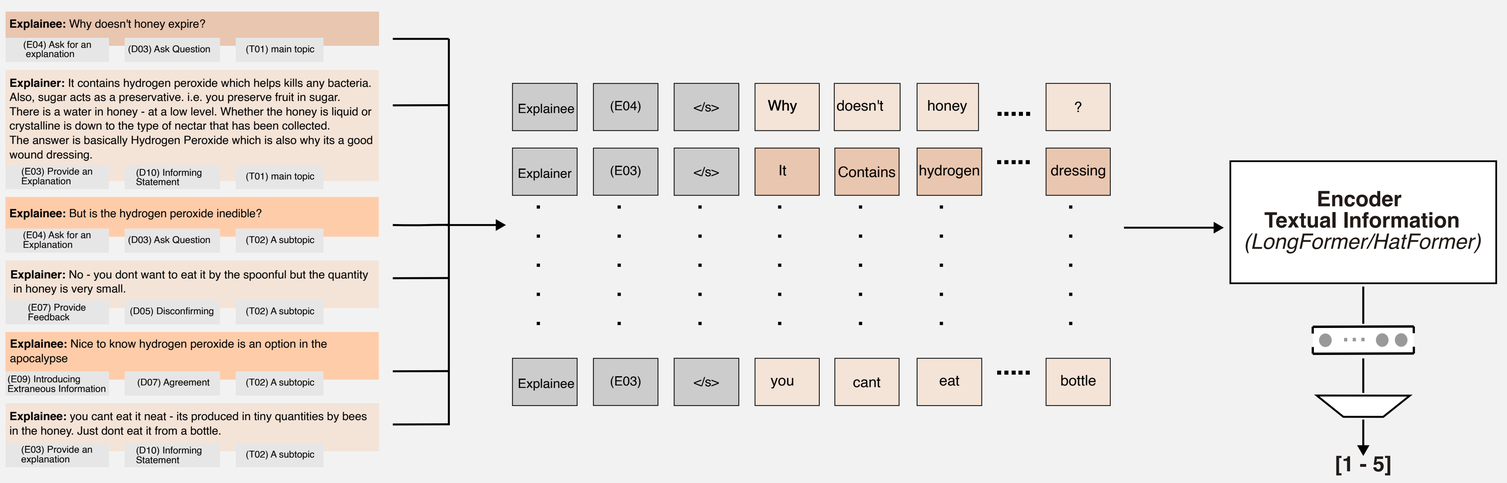
- Distribution of explanation moves and dialogue acts over the different explanation quality scores

Explanation Moves	Freq.	Score Distribution 1	2	3	4	5
(E03) Provide Exp.	1256	22%	15%	25%	17%	21%
(E04) Ask Exp.	1040	25%	15%	22%	15%	23%
(E07) Prov. Feedback	924	**40%**	11%	13%	14%	22%
(E06) Sig. Non-Under.	108	**53%**	14%	14%	12%	7%
(E05) Sig. Under.	45	20%	13%	22%	13%	**31%**
(E08) Provide Assess.	18	**72%**	0%	22%	6%	0%
(E02) Test prior know.	17	**59%**	18%	18%	0%	6%
(E01) Test Underst.	17	**53%**	24%	6%	12%	6%
(E10) Other	16	31%	19%	31%	19%	0%
(E09) Extra. Info.	16	**62%**	25%	6%	0%	6%

Dialogue Acts	Freq.	Score Distribution 1	2	3	4	5
(D09) Info. Statement	917	19%	12%	23%	18%	**28%**
(D03) Question	580	24%	13%	23%	15%	25%
(D08) Disagreement	450	**59%**	14%	18%	6%	2%
(D02) What/how Ques.	432	**30%**	17%	20%	15%	18%
(D06) Answer	322	31%	18%	11%	18%	22%
(D07) Agreement	271	17%	7%	18%	**22%**	**35%**
(D01) Check Question	145	32%	18%	21%	14%	15%
(D05) Disconfirm.	126	**39%**	16%	25%	11%	10%
(D04) Confirm.	116	28%	13%	16%	17%	25%
(D10) Other	98	37%	13%	15%	15%	19%

Table 2: The frequency of explanation moves (left) and dialogue acts (right) in our dataset broken into each of the explanation quality levels [1-5]. Highlighted in bold values that distinguish the presence of these moves in high quality dialogues compared to low quality ones.

Modeling Quality of Explanation Dialogues



- **Approach**
  - Transformer-based model with turn-labels encoded as tokens into the texts of each turn
  - **Baselines:** Average, LongFormer, and HatFormer
  - **Evaluation:** RMSE and MAE
  - **Two settings:** Groung-truth turn-labels and Predicted turn-labels

- **Results**
  - Encoding turn-labels results in lower error
  - In predicted tunr-labels settings, best reduction error comes from encoding all turn-labels into HatFormer

Approach	Ground Truth RMSE	Ground Truth MAE	Predictions RMSE	Predictions MAE
Average Baseline	1.60	1.42	1.60	1.42
HatFormer	1.42	1.17	1.42	1.17
w/ Dialogue Act	**1.29**	*1.05	1.31	1.09
w/ Expl. Move	1.41	1.21	1.43	1.22
w/ Topic	1.41	1.20	1.41	1.20
w/ ALL	1.30	**1.05**	**1.28**	**1.05**
LongFormer	1.34	1.13	1.34	1.13
w/ Dialogue Act	**1.31**	**1.05**	**1.32**	**1.06**
w/ Expl. Move	**1.31**	**1.05**	**1.32**	1.09
w/ Topic	1.35	1.15	1.34	1.14
w/ ALL	1.32	1.08	1.34	1.10