Learning to Speak and Act in a Fantasy Text Adventure Game

Authors:

Jack Urbanek, Angela Fan, Siddharth Karamcheti, Saachi Jain, Samuel Humeau, Emily Dinan, Tim Rocktäschel, Douwe Kiela, Arthur Szlam, Jason Weston

Conference: EMNLP-IJCNLP 2019(Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing)

Mukesh Kumar Vidam (DH73736)

12 Nov 2024

Summary of the Paper

- **Goal**: Improve AI dialogue by grounding it in a virtual world, enabling both dialogue and actions.
- **Environment**: Large multiplayer game world, **LIGHT** (Learning in Interactive Games with Humans and Text), featuring:
  - 663 locations
  - 3,462 objects
  - 1,755 characters
- **Dataset**: 11,000+ player interactions capturing both conversations and actions.
- **Key Finding**: Grounding in context improves AI's action and dialogue relevance, but performance still lags in new scenarios.

LIGHT Environment Overview

- **Game World**: Text-based fantasy setting for both human and AI agents.
- **Crowdsourced Content**: Diverse locations, characters, and objects created by the crowd.
- **Interactions**: Characters can:
  - Perform actions (e.g., “get,” “drop”)
  - Emote (e.g., “smile,” “laugh”)
  - Engage in dialogue
- **Character & Object Descriptions**: Each character has a persona and backstory; objects have specific uses (like drinking from a cup).

	Train	Valid	Test Seen	Test Unseen
Locations	589	352	499	74
Objects	2658	1412	1895	844
Characters	1369	546	820	360
Dialogues	8538	500	1000	739
Utterances	110877	6623	13272	9853
Emotes	17609	1156	2495	1301
Actions	20256	1518	3227	1880
Vocabulary Size	32182	11327	11984	9984
Utterance Length	18.3	19.2	19.4	16.2

LIGHT dataset statistics.

Methodology

- **Models Used**:
  - **Bi-ranker model**: Ranks responses/actions for faster performance.
  - **Cross-ranker model**: Provides deeper context understanding but is slower.
- **Training**: Trained on 11,000 interactions from LIGHT with both actions and dialogues grounded in context.
- **Evaluation**:
  - **Seen Environments**: Familiar settings, with trained locations/characters.
  - **Unseen Environments**: New settings, with untrained locations/characters.
  - **Performance Metrics**: Dialogue accuracy, action accuracy, coherence.
- **Human Benchmark**: Compared model performance to human results to identify gaps.

Persona 1: Servant	**Persona 2: King**
I come from the lower class. I do what I am told without question. I can not read. I have not seen my family in a long time.	I am a king of the whole empire. I give rules and pursuit them. I am brave and fearless.
Carrying: a duster, a small bucket, a rag Wearing: a shirt	Carrying: a crown, a scepter

Setting: The main foyer is massive. A grand staircase sits to the back of the foyer leading to the upstairs. At the front of the foyer stand two servants ready to help anyone who comes to visit. To the left of the room there is a doorway leading into a corridor. To the right there is a door leading to another corridor for the King's servants. At the foot of the stairs there is a bearskin rug that is staring at you almost as if still hungry. The walls are lined with portraits of the king and his family.

Servant: my humble king. What am I to do to serve you?
King: Ahhh. My loyal servant. Polish my scepter.
Action: give scepter to servant
Servant: Yes my lord. I will polish it immediately. Am I to return it to you personally?
Action: put scepter in small bucket
King: Yes. Yes. Of course. Also check the jewels in my crown. They seem loose.
Action: give crown to servant
Servant: But sire I am not qualified to do that. Would you prefer I take it to someone?
King: Oh fine then.
Emote: gesture sigh
Servant: I am sorry sir the rug startled me
Action: drop crown
King: Haha! That's bear I slain on my latest hunting trip. He's a mighty beast!
Emote: gesture laugh
Servant: and if I may ask where did you go hunting sire?
King: The great woods of course. This bear was stealing children in the kingdom. Surely you heard about it.
Servant: sire. I have not been outside of these walls in quiet some time. I have not seen my family in ages.
King: Such is the life of a servant I suppose. How's that scepter looking?
Servant: it is almost ready sire. and the crown who would you like me to take it to?
Action: get scepter from small bucket
King: Here just give it back. I'll have the queen find someone.

Figure 1: Example dialogue from the LIGHT dataset.

Key Findings and Results

- **Seen vs. Unseen Environments**:
  - Models perform better in familiar ("seen") settings.
  - Accuracy drops in new ("unseen") settings, showing generalization challenges.
- **Human Comparison**:
  - Bi-Ranker and Cross-Ranker models perform well but fall short of human-level accuracy, especially with complex dialogue and actions.
- **Contextual Grounding**:
  - Grounding improves model relevance and realism in responses.
- **Limitations**:
  - Generalization is a key challenge; models struggle with new environments.

Method	Test Seen - Dialogue R@1/20	Test Seen - Action Acc	Test Seen - Emote Acc	Test Unseen - Dialogue R@1/20	Test Unseen - Action Acc	Test Unseen - Emote Acc
Random baseline	5.0	12.2	4.5	5.0	12.1	4.5
IR baseline	23.7	20.6	7.5	21.8	20.5	8.46
FastText Classification	-	-	13.2	-	-	9.92
Starspace	53.8	17.8	11.6	27.9	16.4	9.8
Transformer MemNet	70.9	24.5	25.0	66.0	21.1	16.6
BERT-based Bi-Ranker	76.5	42.5	25.0	70.5	38.6	25.7
BERT-based Cross-Ranker	74.9	50.7	25.8	69.7	51.8	28.6
Human Performance*	**87.5 ± 2.4**	**62.0 ± 3.1**	**27.0 ± 2.5**	**91.8 ± 1.9**	**71.9 ± 3.5**	**34.4 ± 2.6**

Ranking Model Test Performance

Strengths of the Paper

- **Comprehensive Dataset**: Rich variety of characters, objects, and locations, making it valuable for studying grounded dialogue.
- **Dialogue & Action Integration**: Combines language and behavior in a fantasy setting.
- **Innovative Model Design**: Creative use of BERT-based bi-ranker and cross-ranker models for contextual accuracy.
- **Focus on Generalization**: Tests in both seen and unseen settings provide insights into real-world adaptability.

Weaknesses of the Paper

Limited Realism: Text-only format lacks visual/sensory data, limiting real-world applicability.

Performance Gap: Models fall short of human benchmarks, especially in unseen environments.

Evaluation Metrics: Reliance on in-game metrics. Broader tests could provide more insights.

Reliance on BERT: BERT has limitations in contextual understanding, exploring alternative models might improve performance.

Applications to Storytelling and Interactive Fiction

- **Enhanced Story Generation**: Dynamic, context-responsive stories through dialogue and action models.
- **Improved Player Interaction**: Characters respond naturally to players' actions, creating immersive experiences.
- **Adaptive Narratives**: Evolving stories based on player choices, making each playthrough unique.
- **Lifelike Characters**: Grounded dialogue allows characters to display realistic personalities and emotions.

Conclusion

Summary: LIGHT provides a platform for studying grounded AI interactions in a detailed virtual world.

Key Findings: Contextual grounding improves AI relevance but generalization remains challenging.

Future Directions: Focus on bridging AI-human performance gaps, incorporating multi-modal inputs, and improving adaptability.

Applications: Potential in storytelling, games, virtual assistants, and education for immersive and dynamic character interactions.

THANK YOU