NRI: Collaborative Research: RobotSLANG: Simultaneous Localization, Mapping, and Language Acquisition

Jeffrey Mark Siskind
Purdue University

Imagine that a robot is deployed in an unknown office environment where individuals work in separate areas to follow social distancing guidelines. You need to give an important package to a colleague but you're unable to go in person. You summon a robot and give it navigational instructions in plain English to your colleague's office. The robot engages you in a dialogue to clarify some ambiguity in your navigational instructions and then follows the plan it infers to deliver the package.

Key Problems:

1. Extracting a plan from spoken navigational instructions
2. Resolving uncertainty through ongoing dialogue
3. Grounding navigational instructions in the environment
4. Following the plan to the destination

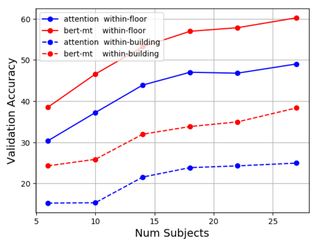
Contributions:

1. We provide a novel dataset of transcript-plan pairs for navigation in indoor environments. We apply a novel data augmentation method to train a transformer network to support multi-turn dialogue.
2. We provide a novel dataset of robot trajectories paired with navigational instructions. We use novel automatic annotation and data augmentation techniques to train a 2D CNN to produce navigational goals and statuses that correspond to the input instructions.
3. We demonstrate performance in real indoor environments with untrained volunteers that are distinct from our training set to show the generalizability of our approach.

Machine-learned Dialogue:

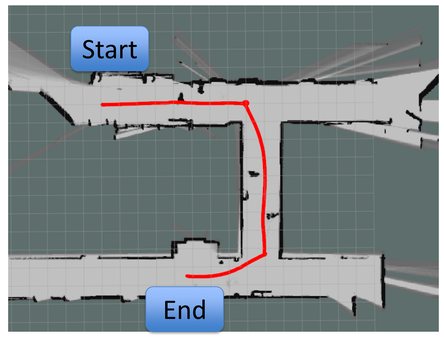
- Input:
  - Current plan
  - Utterance
- Output:
  - Updated plan
  - Follow-up question

In this plot, we measured the k-fold cross-validation accuracy across varying numbers of subjects for two networks: attention-is-all-you-need and a BERT-based transformer



Demonstration:

Robot	I'm trying to get to room 264. How do I get to my destination?
Person	First you need to turn around and hang a right.
Robot	What do I do after I turn right?
Person	Go to the end of the hallway.
Robot	What do I do after getting to the first end of the hall?
Person	Turn right again and then the door will be on your left.
Robot	Got it. Thanks!
PLAN	[turn-around int-R right end right goal-L]



Experimental Results:

- Three novel test-building environments
- Four untrained volunteers
- 39 trials consisting of spoken dialogue with the robot followed by a single attempt to follow the plan and reach the goal

	HAMP	KNOY	ME	Total
Complete Success	5	10	6	21
Dialogue Success/Navigation Failure	2	3	5	10
Dialogue Failure/Navigation Success	4	1	3	8
Complete Failure	0	0	0	0
Total	11	14	14	39

Broader impact on society:

- Can allow humans deliver items to each other without violating social distancing.

Broader impact on education:

- 3 PhDs graduated so far, 3 more PhDs plan to graduate within six months.

2021 NRI & FRR Principal Investigators' Meeting
March 10-12, 2021

Award ID#: 1522954-IIS

NRI: Collaborative Research: RobotSLANG: Simultaneous Localization, Mapping, and Language Acquisition

Jeffrey Mark Siskind
Purdue University

Imagine that a robot is deployed in an unknown office environment where individuals work in separate areas to follow social distancing guidelines. You need to give an important package to a colleague but you're unable to go in person. You summon a robot and give it navigational instructions in plain English to your colleague's office. The robot engages you in a dialogue to clarify some ambiguity in your navigational instructions and then follows the plan it infers to deliver the package.

Key Problems:

1. Extracting a plan from spoken navigational instructions
2. Resolving uncertainty through ongoing dialogue
3. Grounding navigational instructions in the environment
4. Following the plan to the destination

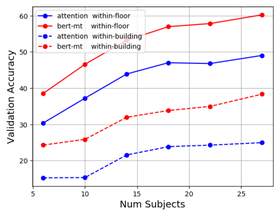
Contributions:

1. We provide a novel dataset of transcript-plan pairs for navigation in indoor environments. We apply a novel data augmentation method to train a transformer network to support multi-turn dialogue.
2. We provide a novel dataset of robot trajectories paired with navigational instructions. We use novel automatic annotation and data augmentation techniques to train a 2D CNN to produce navigational goals and statuses that correspond to the input instructions.
3. We demonstrate performance in real indoor environments with untrained volunteers that are distinct from our training set to show the generalizability of our approach.

Machine-learned Dialogue:

- Input:
  - Current plan
  - Utterance
- Output:
  - Updated plan
  - Follow-up question

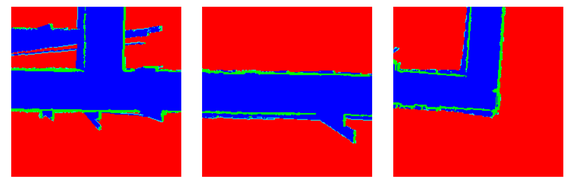
In this plot, we measured the k-fold cross-validation accuracy across varying numbers of subjects for two networks: attention-is-all-you-need and a BERT-based transformer



Machine-learned Navigation:

- Input:
  - Map
  - Instruction embedding
- Predict:
  - Goal location
  - Whether instruction was encountered, not yet present or was not found

In these maps, the robot is positioned at the center and facing rightward



int-L_left
goal: (0.51, 4.94)
class: *encountered*

elbow_either
goal: (4.99, 0.05)
class: *not yet present*

end_right
goal: (0.0, 0.0)
class: *not found*

Experimental Results:

- Three novel test-building environments
- Four untrained volunteers
- 39 trials consisting of spoken dialogue with the robot followed by a single attempt to follow the plan and reach the goal

	HAMP	KNOY	ME	Total
Complete Success	5	10	6	21
Dialogue Success/Navigation Failure	2	3	5	10
Dialogue Failure/Navigation Success	4	1	3	8
Complete Failure	0	0	0	0
Total	11	14	14	39

Broader impact on society:

- Can allow humans deliver items to each other without violating social distancing.

Broader impact on education:

- 3 PhDs graduated so far, 3 more PhDs plan to graduate within six months.

2021 NRI & FRR Principal Investigators' Meeting
March 10-12, 2021

Award ID#: 1522954-IIS