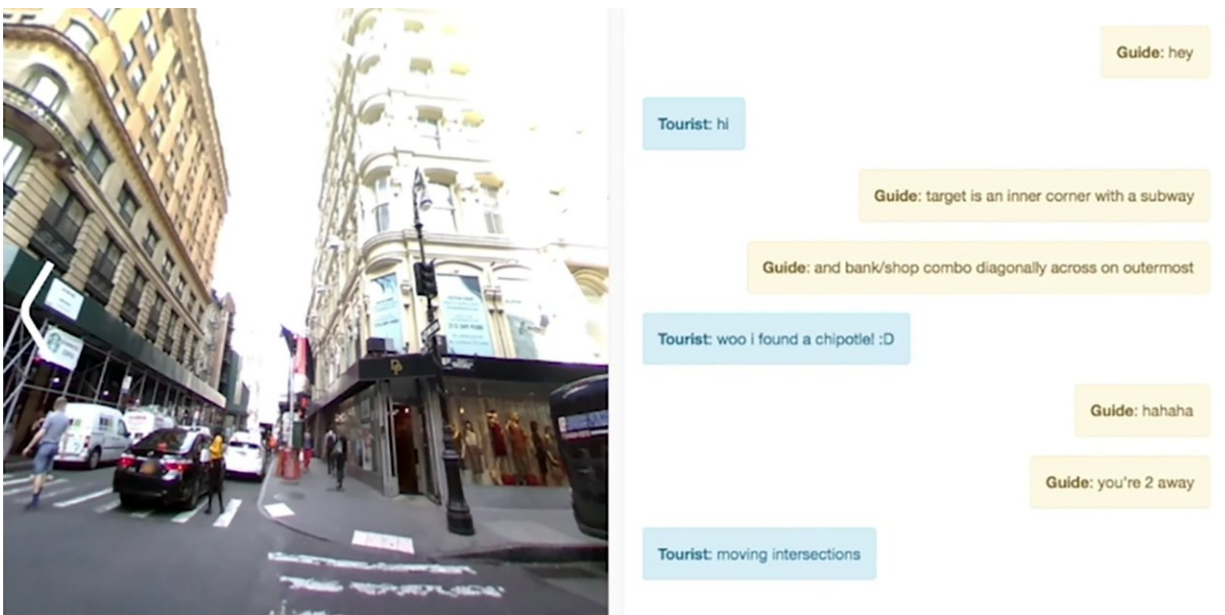


Research community can go on Facebook AI's NYC conversation tour

July 12 2018, by Nancy Owano



Credit: FB

Jason Weston, with doctorate in machine learning at University of London and Douwe Kiela, with doctorate from the University of Cambridge with thesis on grounding semantics in perceptual modalities, are research scientists at Facebook Research and have introduced the world to their formidable team's [Talk the Walk](#).

Talk the Walk is an eye-opener for scientists interested in doing more

for AI as a conversation agent. These days, they do not just gloat over voice assistants telling people when the concert starts or if it will rain. Scientists are exploring goal-directed dialogues.

How easy does that sound? Don't kid yourselves. Trying to get there is hard.

Fast Company turned to Kiela for reasons why the tourist guide effort has research weight. "This task is very important for AI research because it's very hard," Kiela says, "and because it combines all these interesting problems—three-hundred-sixty visual perception, map-based navigation, [visual](#) reasoning, and [natural language](#) communications via dialogue."

They made the point, first off, that natural language is understandable to most people "without requiring extra steps or knowledge to decipher its meaning." Toward that end, Facebook's AI research group, FAIR, are hooked on a certain strategy for AI to show human-level language understanding.

That strategy, they wrote, "is to train those systems in a more natural way, by tying language to specific environments. Just as babies first learn to name what they can see and touch, this approach—sometimes referred to as embodied AI—favors learning in the context of a system's surroundings, rather than training through large data sets of text (like Wikipedia)."

Enter Talk the Walk. They are teaching AI systems to navigate the streets of New York using language exchanges that sound natural between guide and tourist. Two bots have tasks. The tourist bot wants to navigate its way through 360-degree images of New York City neighborhoods. The guide bot is to help with a map of the neighborhood. The team used MASC (Masked Attention for Spatial Convolution) so that the guide bot could focus on the right place on the map.

They said their goal is "to achieve that high degree of synthetic performance through natural language interaction, and to challenge the community to do the same."

Information for Talk the Walk is on [GitHub](#). "Sharing this work will provide other researchers with a framework to test their own embodied AI systems, particularly with respect to dialogue."

A 360-degree camera captured 5 neighborhoods, Hell's Kitchen, East Village, Financial District, Upper East Side, and Williamsburg in Brooklyn. Daniel Terdiman in *Fast Company* said the guide bot used a standard 2D map with generic waypoints—"bank," "coffee shop," "deli"—to deliver its instructions on how to navigate.

The AI work involved is about perceiving a certain environment, navigating through it, and communicating about it. Lucas Matney in *TechCrunch* wrote that "In "Talk the Walk," the guide AI bot had all of this 2D map data and the tourist bot had all of this rich 360 visual data, but it was only through [communication](#) with each other that they were able to carry out their directives."

Tourist: Woo I found a Chipotle

Guide: Haha

Tourist: "I'm diagonal from a bank"

Guide: "Cool."

The paper discussing their work can be found on arXiv. It is titled "Talk the Walk: Navigating New York City through Grounded Dialogue," by

Harm de Vries, Kurt Shuster, Dhruv Batra, Devi Parikh, Jason Weston and Douwe Kiela.

More information: Talk the Walk: Navigating New York City through Grounded Dialogue, arXiv:1807.03367 [cs.AI]
arxiv.org/abs/1807.03367

Abstract

We introduce "Talk The Walk", the first large-scale dialogue dataset grounded in action and perception. The task involves two agents (a "guide" and a "tourist") that communicate via natural language in order to achieve a common goal: having the tourist navigate to a given target location. The task and dataset, which are described in detail, are challenging and their full solution is an open problem that we pose to the community. We (i) focus on the task of tourist localization and develop the novel Masked Attention for Spatial Convolutions (MASC) mechanism that allows for grounding tourist utterances into the guide's map, (ii) show it yields significant improvements for both emergent and natural language communication, and (iii) using this method, we establish non-trivial baselines on the full task.

© 2018 Tech Xplore

Citation: Research community can go on Facebook AI's NYC conversation tour (2018, July 12) retrieved 18 April 2024 from

<https://techxplore.com/news/2018-07-facebook-ai-nyc-conversation.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.