The Conversational Intelligence Challenge

NIPS 2017 Live Competition

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

Schedule

01:30 PM - 01:45 PM Introduction by the competition chairs, presentation of the challenge, its design, obtained results.

01:45 PM - 01:50 PM Awarding prize.

01:50 PM - 02:00 PM Short presentation by winning team.

02:00 PM - 02:30 PM Inspirational talk by Alexander Rudnicky

02:30 PM - 03:00 PM Panel discussion moderated by Julian Serban. The participants are Alexander Rudnicky (CMU), Dilek Hakkani-Tur (Google), Jianfeng Gao (MS Research) and Joelle Pineau (McGill U and FAIR).

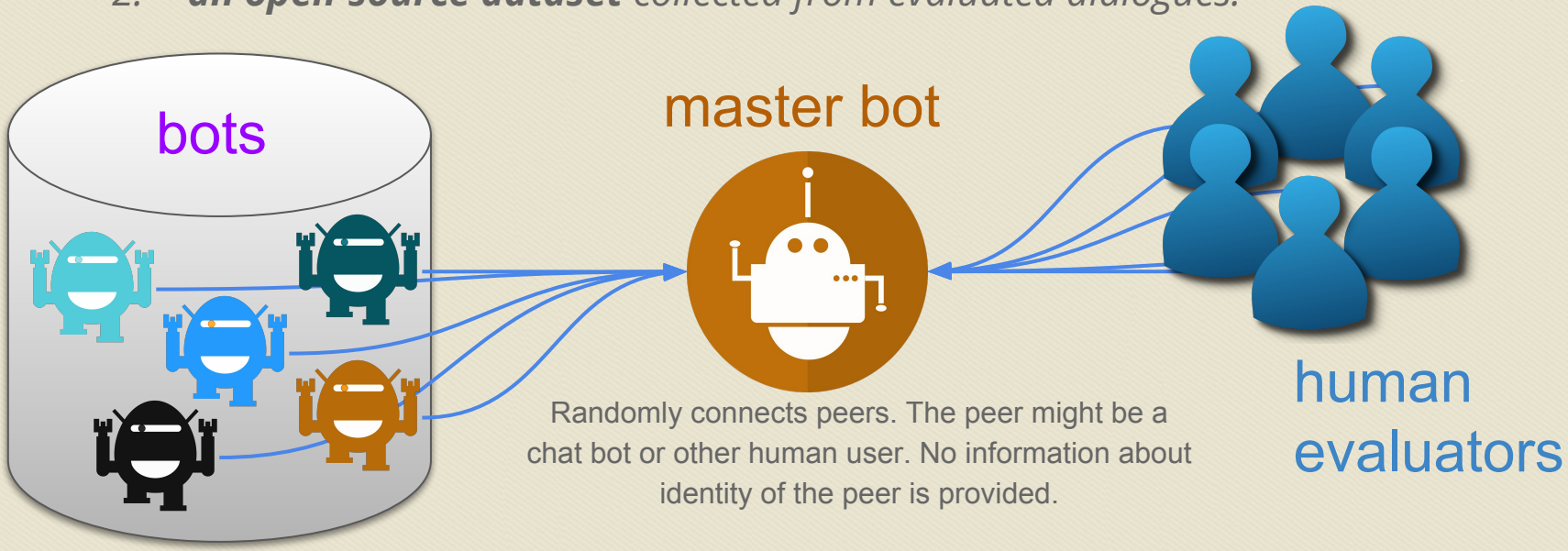
03:00 PM - 03:30 PM Coffee break + Poster session.

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

We expect the competition to have two major outcomes:

1. *evaluate **quality of state-of-the-art dialogue systems**, and*
2. ***an open-source dataset** collected from evaluated dialogues.*



The Conversational Intelligence Challenge

NIPS 2017 Live Competition

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

1. Read a text
2. Talk to a peer (**human** or **bot**) about the text
3. Evaluate bot's quality:
   a.  and  for utterances
   b. **1** to **5** for dialogues - overall, breadth and engagement

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

Results of Pre-NIPS Round

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

Results of Pre-NIPS Round

	BOT	FINAL SCORE	Organisation
1	**poetwannabe**	2.386	University of Wroclaw
2	**bot#1337**	2.318	MIPT
3	**RLLChatBot**	2.226	McGill University
4	**kAIb**	2.027	KAIST, AIBrain, Crossert
5	**DeepTalkHawk**	1.432	Text Machine, UMass Lowell, Trinity College
6	**PolyU**	1.284	The Hong Kong Polytechnic University, Fudan University

Sat Dec 9th 08:00 AM -- 06:30 PM @ S1

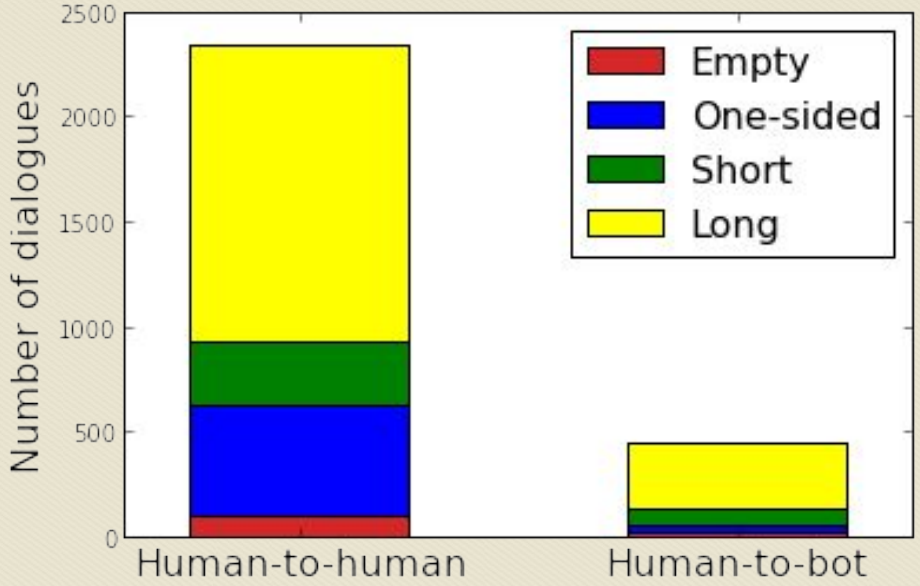
Machine Learning Challenges as a Research Tool

Isabelle Guyon · Evelyne Viegas · Sergio Escalera · Jacob D Abernethy

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

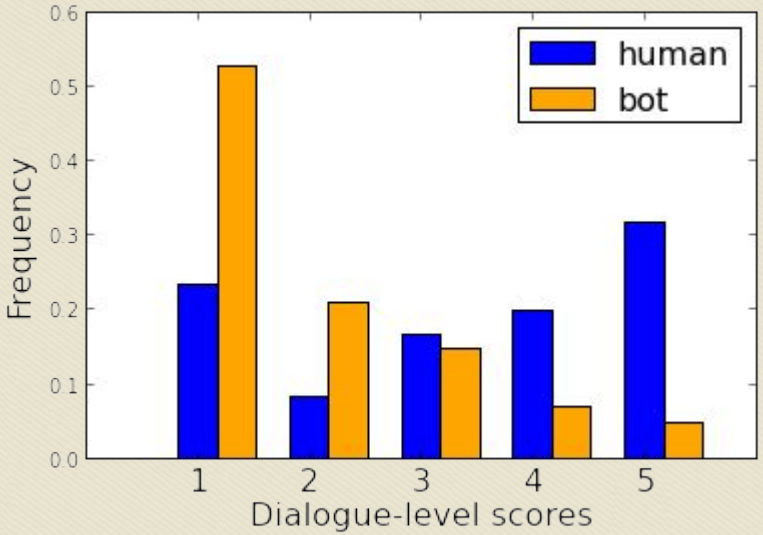
2,778 dialogues: human-to-bot (84%) and human-to-human (16%)



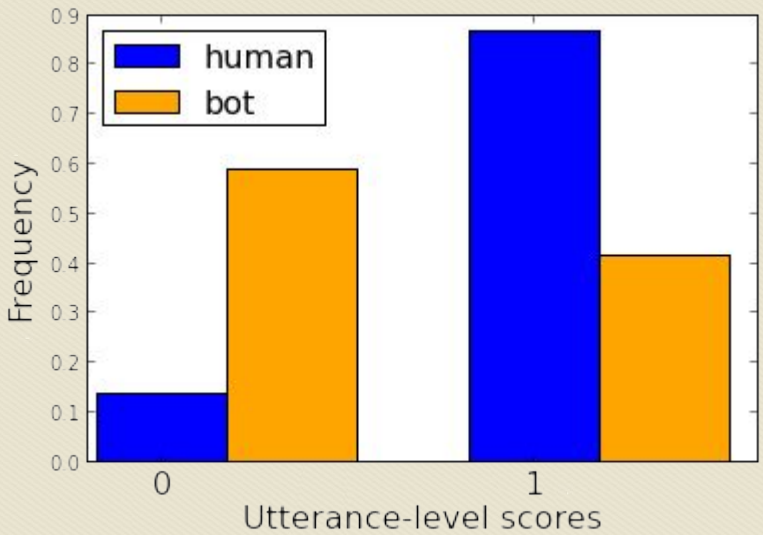
The Conversational Intelligence Challenge

NIPS 2017 Live Competition

Dialogue-level quality



Utterance-level quality



Correlation of dialogue-level and utterance-level scores: **0.6**

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

Data collection insights:

- dialogue-level *quality*, *breadth* and *engagement* are correlated;
- evaluating dialogues on the fly is difficult, turns are not rated;
- paragraphs from Wikipedia are too long and usually not used for the dialogue;
- humans are irresponsible;
- humans prefer talking to bots and break dialogue when understand that peer is human.

More details and data on - http://convai.io/data/

https://github.com/facebookresearch/ParlAI/tree/master/parlai/tasks/convai_chitchat

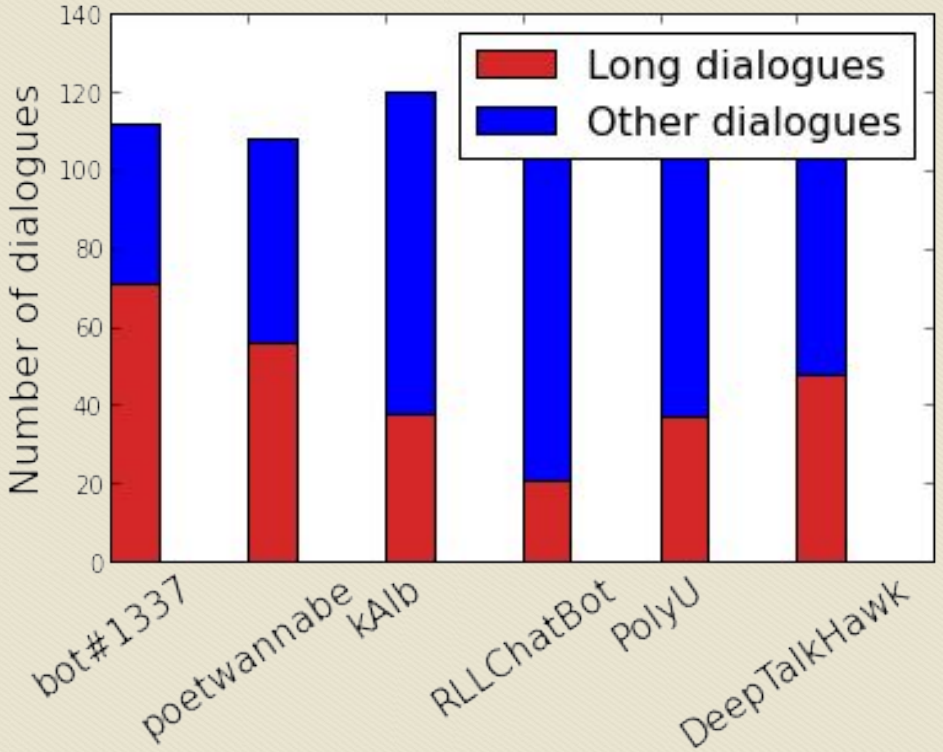
ParlAI

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

NIPS data collection:

- 739 dialogues overall
- 292 long dialogues
- number of long dialogues depends on bot's quality



The Conversational Intelligence Challenge

NIPS 2017 Live Competition

Grand Conclusions

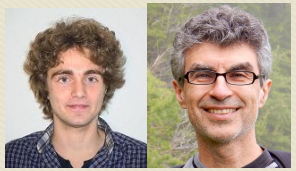
Everybody is fascinated with Conversational AI!

Nobody wants to to spend time chatting with bots on irrelevant topics!

1. **We should continue**
2. **Motivate people**

MIPT
MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY

Université de Montréal



McGill

CARNEGIE MELLON UNIVERSITY
PITTSBURGH PENNSYLVANIA 1900

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

Jan 29 - Feb 4
2018

DeepHack.Babel

/ always
in your
language /

Travel grants!!!

The main goal of hackathon is **semi-supervised Machine Translation**: Machine Translation which makes use not only from (often scarce) parallel data, but also from (potentially unlimited) untranslated texts in source and target languages. The participants will need to train an MT model on a very small parallel corpus and boost its performance using monolingual non-parallel data.

Timeline

1st December, 2017 — **Registration is open**

Registered teams are allowed to submit applications and solutions for the Qualification Round

8th January, 2018 — **Qualification Round is closed**

9th January, 2018 — **Notification of acceptance.**

Top of 50 participants are selected

29th January - 4th February, 2018 — **DeepHack.Babel**

A 24/7 week-long sci-hack school

http://deephack.me

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

6

DeepTalkHawk	1.229	Text Machine, UMass Lowell, Trinity College

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

5

PolyU	1.5	The Hong Kong Polytechnic University, Fudan University

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

4

McGill

RLLChatBot	1.905	McGill University

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

3

kAIb	2.105	KAIST, AIBrain, Crossert

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

1-2

Uniwersytet Wrocławski

bot#1337	2.746	MIPT
poetwannabe	2.536	University of Wroclaw

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

Task:

between task-oriented and chit-chat dialogue

Both human evaluators and dialogue agents complete the same task.

1. Both parties are given a text of a wikipedia article.
2. Discuss content of the article with the peer as long as you wish.
3. Choose another article and/or anonymous peer.

The Conversational Intelligence Challenge

NIPS 2017 Live Competition

Evaluation

1. Evaluator will not be given any information about identity of the peer.
2. Members of the team will be automatically excluded from evaluation of their own submission and each other.
3. The quality of every response is subjectively evaluated by “thumbs up” & “thumbs down” binary scale. This evaluation could be skipped by evaluator.
4. The quality of the dialog as a whole as well as its breadth and engagement are evaluated on the 1 to 5 range. This evaluation is obligatory.
5. Final rating is calculated as an average of evaluation values accumulated by submitted system during the NIPS Round of Competition.