NTCIR-13 Core Task:
Short Text Conversation (STC-2)

Lifeng Shang[1], Tetsuya Sakai[2], Zhengdong Lu[1], Hang Li[1],
Ryuichiro Higashinaka[3], and Yusuke Miyao[4]

[1]Noahs Ark Lab of Huawei [2]Waseda University
[3]Nippon Telegraph and Telephone Corporation
[4]National Institute of Informatics
Email: stc-org@list.waseda.jp

August 17, 2016@NTCIR-13 Kickoff

What is STC?

Non-task-oriented,
man-system, single-turn dialogues

Tetsuya Sakai (酒井哲也) @tetsuyasakai · 5m
The first day in Hawaii. Watching the sunset at the balcony with a big glass of wine in hand.

1

NTCIR STC @ntcirstc · 3m
@tetsuyasakai Enjoy it & don't forget to share your photos!

Possible responses (comments)

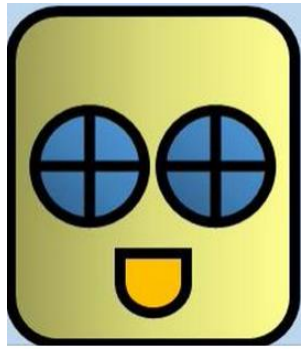
In reply to Tetsuya Sakai (酒井哲也)

NTCIR STC @ntcirstc · 3m
@tetsuyasakai How long are you going to stay there?

3:49 PM - 6 Jul 2015 · Details

STC-1 was the largest task of NTCIR-12(June 2016)!

NTCIR-12 STC task definition



Given a new post, can the system return a “good” response by retrieving a comment to an old post from a repository?

Repository

old post	old comment
old post	old comment
old post	old comment
old post	old comment
old post	old comment

Training data

new post	old comment
new post	old comment
new post	old comment

Graded label (L0-L2) for each comment

Test data

new post
new post

For each new post, retrieve and rank old comments!

Objectives and Methods

- The ultimate objective

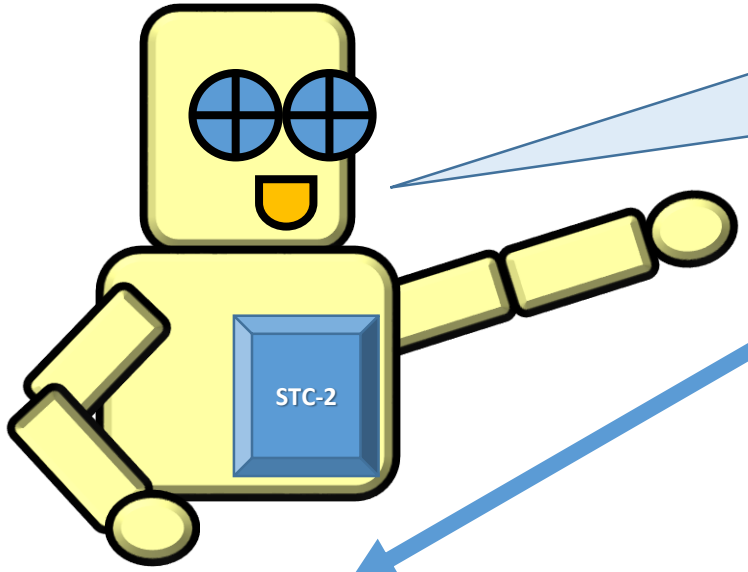
Build an open-domain system that can interact naturally with humans

- **TWO** run types for NTCIR13

1. Retrieval-based runs: Build an IR system that effectively reuses past comments to respond to a post.
2. Generation-based runs: Train a machine learning model that can generate new comments which may not appear in the training set

Fluency: the generated comments should be natural language and should not contain grammatical errors.
Coherence: the post-comment pair makes sense as a consecutive short text exchange between two people.
Usefulness: the comment contains information or an opinion that might be useful to the author of the post.

Retrieval-based runs (same as NTCIR-12)



Given a new post, can a coherent and useful comment be returned by searching a post-comment repository?

post

Search and reuse

post-comment repository

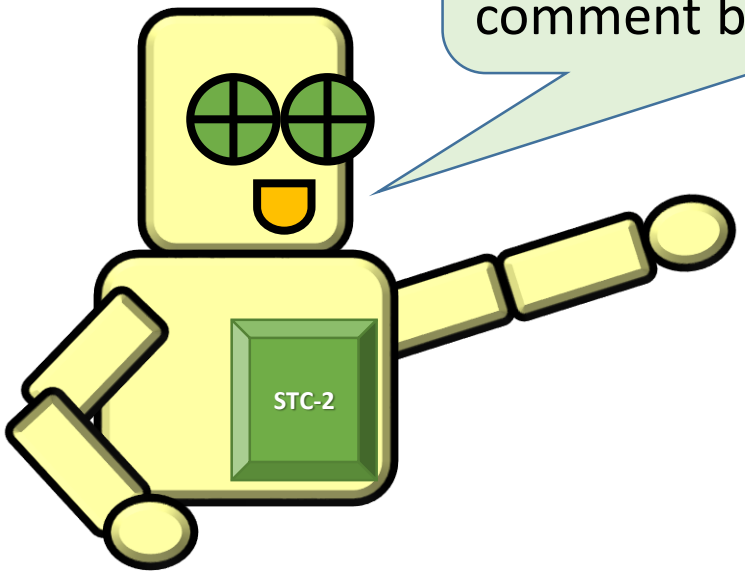
post comment comment

post comment comment

post comment comment

post comment comment

Generation-based runs (new!)



Given a new post, can a fluent, coherent and useful comment be generated?

post

The Trained Generator

Understanding

Generating

generated comment

generated comment

generated comment

Used to train the generator

post-comment repository

post comment comment

post comment comment

post comment comment

post comment comment

Subtasks of STC-2

- Chinese Subtask
  - Still use post-comment pairs from Weibo.
  - Dataset: We will randomly select half of the post-comment pairs from the repository used at NTCIR-12, and then strictly follow the method described in [3] to construct the other new half.
  - http://ntcirstc.noahlab.com.hk/STC2/stc-cn.htm
- Japanese Subtask
  - Still use post-comment pairs from Twitter.
  - Dataset: We are preparing for a new data source in addition to Twitter.
  - http://ntcirstc.noahlab.com.hk/STC2/stc-jp.htm

Update

- Chinese subtask
  - The repository of post-comment pairs has been constructed
  - The training data set used for retrieval-based method has been constructed and will be labeled soon
- Japanese subtask
  - The repository of post-comment pairs will be constructed from the new data source
  - The training data for retrieval-based methods will be labeled

Evaluation Measures

- As in STC@NTCIR-12, evaluation will be conducted by standard IR effectiveness measures
- Pooling and graded relevance assessments

L2: coherent and useful

L1: coherent but not useful

L0: not coherent (and therefore not useful either)

- Evaluation measures (basically one good comment is enough):

nG@1 (normalised gain at rank 1)

ERR (expected reciprocal rank)

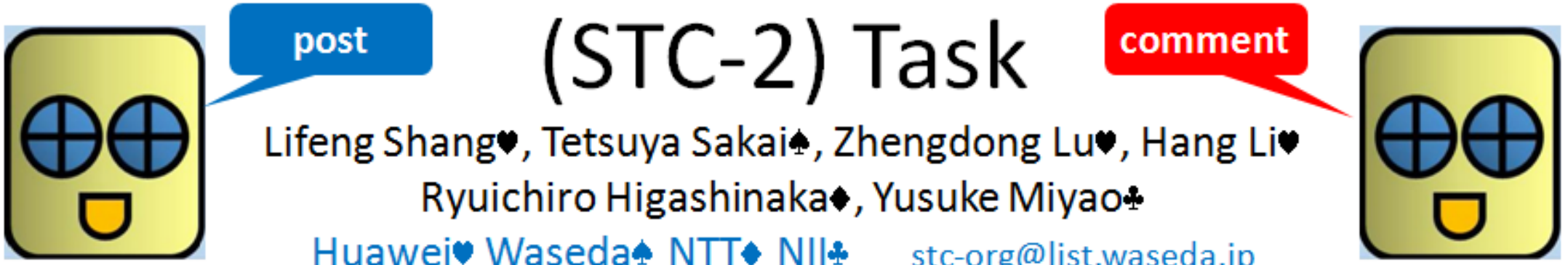
P+ (Similar to Q-measure, suitable for navigational intents)

[Sakai14PROMISE]

Schedule

Jul-Aug 2016	Post-comment pairs released to registered participants
Oct 2016-Jan 2017	Training data released
Apr 2017	Task registration due
May 2017	STC run submission deadline
Jun-Jul 2017	Relevance assessments
Sep 1, 2017	Results and Draft Task overview released to participants
Oct 1, 2017	Participants papers due
Dec 2017	NTCIR-13 Conference

NTCIR-13 Short Text Conversation (STC-2) Task



Lifeng Shang♥, Tetsuya Sakai♠, Zhengdong Lu♥, Hang Li♥
Ryuichiro Higashinaka♦, Yusuke Miyao♣
Huawei♥ Waseda♠ NTT♦ NII♣ stc-org@list.waseda.jp

http://ntcirstc.noahlab.com.hk/STC2/stc-cn.htm

http://ntcirstc.noahlab.com.hk/STC2/stc-jp.htm

Twitter: @ntcirstc

References

[Sakai14PROMISE] Tetsuya Sakai: Metrics, Statistics, Tests, PROMISE Winter School 2013: Bridging between Information Retrieval and Databases (LNCS 8173), Springer, 2014.

https://waseda.box.com/sakai14PROMISE

[Shang16] Lifeng Shang, Tetsuya Sakai, Zhengdong Lu, Hang Li, Ryuichiro Higashinaka and Yusuke Miyao: Overview of the NTCIR-12 Short Text Conversation Task, NTCIR-12 Proceedings, 2016.

http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings12/pdf/ntcir/OVERVIEW/01-NTCIR12-OV-STC-ShangL.pdf