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

Lifeng Shang¹, Tetsuya Sakai², Zhengdong Lu¹, Hang Li¹, Ryuichiro Higashinaka³, and Yusuke Miyao⁴

¹Noahs Ark Lab of Huawei  ²Waseda University  ³Nippon Telegraph and Telephone Corporation  ⁴National 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

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

In the end, we received Chinese runs from 16 teams and Japanese runs from 7 teams. 22 unique active teams. We are the biggest #ntcir12 task!
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?

For each new post, retrieve and rank old comments!

Repository

<table>
<thead>
<tr>
<th>old post</th>
<th>old comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>old post</td>
<td>old comment</td>
</tr>
<tr>
<td>old post</td>
<td>old comment</td>
</tr>
<tr>
<td>old post</td>
<td>old comment</td>
</tr>
</tbody>
</table>

Training data

<table>
<thead>
<tr>
<th>new post</th>
<th>old comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>new post</td>
<td>old comment</td>
</tr>
<tr>
<td>new post</td>
<td>old comment</td>
</tr>
</tbody>
</table>

Test data

<table>
<thead>
<tr>
<th>new post</th>
</tr>
</thead>
<tbody>
<tr>
<td>new post</td>
</tr>
</tbody>
</table>

Graded label (L0-L2) for each comment
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?

Search and reuse

post-comment repository
Generation-based runs (new!)

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

The Trained Generator

Used to train the generator

post-comment repository

post

comment

comment

post

comment

comment

post

comment

comment

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.

• Japanese Subtask
  • Still use post-comment pairs from Twitter.
  • Dataset: We are preparing for a new data source in addition to Twitter.
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

<table>
<thead>
<tr>
<th>Date Range</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul-Aug 2016</td>
<td>Post-comment pairs released to registered participants</td>
</tr>
<tr>
<td>Oct 2016-Jan 2017</td>
<td>Training data released</td>
</tr>
<tr>
<td>Apr 2017</td>
<td>Task registration due</td>
</tr>
<tr>
<td>May 2017</td>
<td>STC run submission deadline</td>
</tr>
<tr>
<td>Jun-Jul 2017</td>
<td>Relevance assessments</td>
</tr>
<tr>
<td>Sep 1, 2017</td>
<td>Results and Draft Task overview released to participants</td>
</tr>
<tr>
<td>Oct 1, 2017</td>
<td>Participants papers due</td>
</tr>
<tr>
<td>Dec 2017</td>
<td>NTCIR-13 Conference</td>
</tr>
</tbody>
</table>
NTCIR-13 Short Text Conversation (STC-2) Task

Lifeng Shang, Tetsuya Sakai, Zhengdong Lu, Hang Li, Ryuichiro Higashinaka, Yusuke Miyao
Huawei, Waseda, NTT, NII

Twitter: @ntcirstc
References

https://waseda.box.com/sakai14PROMISE