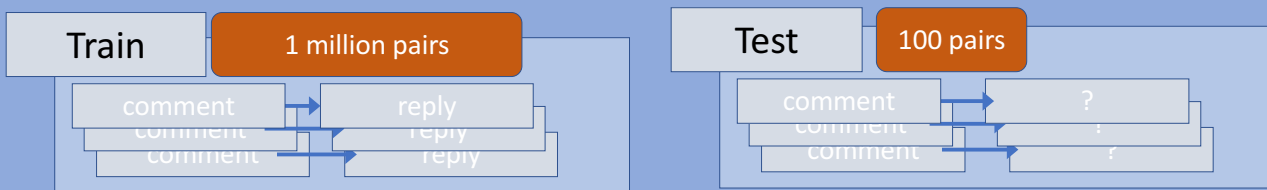# Report on Japanese subtask for NTCIR-13 STC-2 from mnmlb

Sotaro Takeshita, Ryuji Tamaki, Yasuhiro Minami, Takeru Kazama, Masato Nakamura
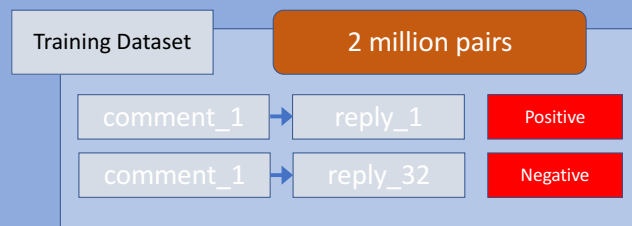The University of Electro-Communications

Abstract:
We report a Japanese subtask for NTCIR-13 STC-2 for which we made a dialogue system and introduced neural network-based retrieval models (LSTM, ESIM and CNN) to rank the dialogue replies in the training dataset. We used data from Yahoo! News comments data and introduced LSTM and ESIM to effectively capture sequential information from the given comments. To evaluate the effectiveness, we compared systems using LSTM or ESIM with systems that use CNN. We also introduced an n-gram-based statistical filter into our systems to reduce the number of reply candidates.

STC-2 Japanese subtask Description:

Train	1 million pairs
comment	reply
comment	reply
comment	reply

Test	100 pairs
comment	?
comment	?
comment	?

For the training data, we were given 894,998 comment and reply pairs. In the test phase, we were required to generate comment to each of 100 comments.
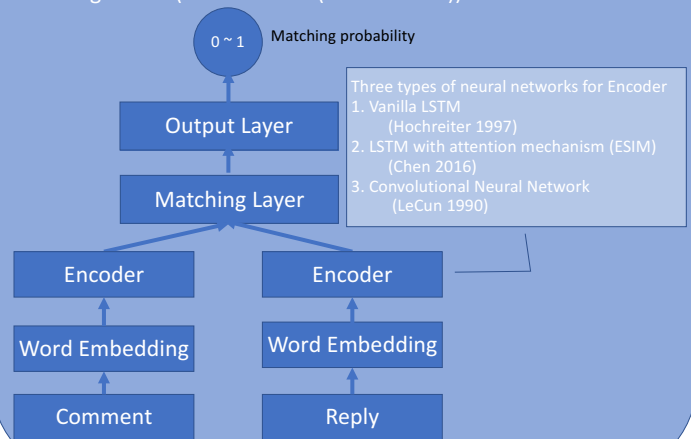
System Framework:

Training Dataset	2 million pairs	
comment_1	reply_1	Positive
comment_1	reply_32	Negative

1. Train a model to predict positive (1) or negative (0) for a given comment-reply pair
2. Calculate matching score for a given comment with all reply candidates (or filtered candidates)
3. Return the reply which scores the highest in 2

Method: Candidate raking

Ranking Model (Dual Encoder (R Lowe 2017))

0 ~ 1 Matching probability

Output Layer

Matching Layer

Encoder — Encoder

Word Embedding — Word Embedding

Comment — Reply

Three types of neural networks for Encoder
1. Vanilla LSTM (Hochreiter 1997)
2. LSTM with attention mechanism (ESIM) (Chen 2016)
3. Convolutional Neural Network (LeCun 1990)

Method: Reply candidates filtering
1. Candidates Filtering with simple rules
    1. Remove the candidates that contain the same words more than two times
    2. Remove the candidates that contain more than two sentences
2. N-gram based candidates filtering
    1. Extract top 100 tri-gram combinations which appears at the end of the sentence
    2. Extract top 100 sentences containing ones in list from 1.

Result:

Model	Mean Acc l1, l2 @1
ESIM with filter	0.5400
ESIM without filter	0.6540
LSTM with filter	0.4020
LSTM without filter	0.4200
CNN with filter	0.4520

We can see attention mechanism is improving the result, but the candidate filtering we proposed decreased the accuracy.
Probability calculation method was inappropriate.

References:
Sepp Hochreiter and Ju¨rgen Schmidhuber. Long short-term memory. Neural Comput., 9(8):1735–1780, November 1997.
Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, and Hui Jiang. Enhancing and combining sequential and tree LSTM for natural language inference. CoRR, abs/1609.06038, 2016.
Yann LeCun, Bernhard E. Boser, John S. Denker, Donnie Henderson, R. E. Howard, Wayne E. Hubbard, and Lawrence D. Jackel. Handwritten digit recognition with a back-propagation network. In D. S. Touretzky, editor, Advances in Neural Information Processing Systems 2, pages 396–404. Morgan-Kaufmann, 1990.
Ryan Lowe, Nissan Pow, Iulian Vlad Serban, Laurent Charlin, Chia-Wei Liu, and Joelle Pineau. Training end-to-end dialogue systems with theubuntu dialogue corpus. Dialogue and Discourse, 10.5087/dad.2017.102, 2017.