Gbot at the NTCIR-13 STC-2 Task

Hainan Zhang, Tonglei Guo, Yanyan Lan, Jiafeng Guo, Jun Xu, Jianing Li, Xueqi Cheng
1. CAS Key Lab of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
2. University of Chinese Academy of Sciences, Beijing, China
{zhanghainan, guotonglei, lijianging}@software.ict.ac.cn, {lanyanyan, guojiafeng, junxu, cxq}@ict.ac.cn

BACKGROUND

Problem:
For a given post X, the goal of STC-2 task is to generate or retrieve a response Y which looks like a real human response. The STC-2 task has two subtask: generation-based method and retrieval-based method. For generation-based subtask, we maximize the likelihood probability \( P(Y|X) \). For retrieval-based subtask, we return the top-k relevant comments with the given query.

Motivation:
For generation-based method, most existing works follow the Seq2Seq which is easy to generate general and safe responses [1-2]. We propose to add three kinds of constraint functions to the original Seq2Seq loss function, in order to guide the generation process.

For retrieval-based method, we apply method based on MatchPyramid, which first builds interactions and then uses a deep model to obtain the representation for the interactions and the relevance score.

GENERATION MODELS

Original Seq2Seq Loss Function

\[
\mathcal{L} = - \sum_{(X,Y) \in D} \log P(Y|X)
\]

Constraint describing the quality of a generation G

Our Constrained Loss Function

\[
\mathcal{L}_m = - \sum_{(X,Y) \in D} \text{cons}_m(X, G) \times \log P(Y|X)
\]

Three kinds of constraints

Cosine Similarity function (SIM):

\[
\text{cos}(X, G) = 1 - \cos(Average(X), Average(G))
\]

where Average(\(X\)) is an embedding which is the mean over the word embeddings in sentence X.

MatchingPyramid function (MP) [3]:

\[
s_{mp}(X, G) = \text{Matching} - \text{Pyramid}\(X, G)\)
\[
\text{cons}_m(X, G) = 1 - \frac{s_{mp}(X, G) - mn}{mx - mn}
\]

where \(mn\) and \(mx\) are the min and max score of the set (min-max normalization). We randomly select five negative generated sentences \(\{G_{m1}, ..., G_{m5}\}\). The score set has six scores

\[
\{s_{mp}(X, G_{m1}), ..., s_{mp}(X, G_{m5})\}
\]

Bi-linear function (BL) [4]:

\[
s_{bl}(X, G) = \text{em}(X) \times W \times \text{em}(G)
\]

\[
BL(X, G) = 1 - \frac{s_{bl}(X, G) - mn}{mx - mn}
\]

where \(\text{em}(X)\) is the embedding of X with GRU encoder. W is a matrix of the transformation. ConsBL uses the same min-max normalization as MP does.

RETrieval MODELS

We take use of MatchPyramid [3].

Converting two 1D text representations of words within them to a typically 2D grid. Represent the input of text matching as a matching matrix \(M\), with each element \(M_{ij}\) standing for the basic interaction, i.e. cosine similarity between word \(w_i\) and \(w_j\).

The body of MatchPyramid is a typical convolutional neural network, use the matching matrix mentioned below as input. The \(k\)-th kernel \(w_{ij,k}\) scans over the whole matching matrix \(M\) to generate a feature map \(z_{ij,k}\):

\[
z_{ij,k} = \sigma(\sum_{s=0}^{r_k-1} \sum_{t=0}^{r_k-1} w_{s,t}(1,k) \cdot x_{i+s,j+t} + b(1,k))
\]

where \(r_k\) denotes the size of the \(k\)-th kernel.

A max-pooling is used to get a fixed length pattern vector. A two-layer DNN to produce the final matching score:

\[
s = W_2\sigma(W_1 z + b_1) + b_2
\]

EXPERIMENTS

<table>
<thead>
<tr>
<th>Table 1 The metric-based evaluation results generated from different models.</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
</tr>
<tr>
<td>Seq2Seq att.</td>
</tr>
<tr>
<td>SIM</td>
</tr>
<tr>
<td>MP</td>
</tr>
<tr>
<td>BL</td>
</tr>
</tbody>
</table>

Comparison results

- BL model has the best generations in metric-based evaluation. The distinct-bigram of BL model is 0.0857, which improves 15.5% compared with Seq2Seq att.
- MP model is better. The distinct-bigram of MP model is 0.0565, which improves 10.2% compared with Seq2Seq att.
- The goal of SIM model is to optimized the Average measure, and it got the best evaluation on Average metric-based measure.

Conclusions

- Our constraint models have better results than Seq2Seq att.
- The generated responses of our constraint models are more coherent to the post and the quality of generation has been improved.

REFERENCES