

# Tailored Sequence to Sequence Models to Different Conversation Scenarios

Hainan Zhang, Yanyan Lan, Jiafeng Guo, Jun Xu and Xueqi Cheng

University of Chinese Academy of Sciences, Beijing, China

CAS Key Lab of Network Data Science and Technology,

Institute of Computing Technology, Chinese Academy of Sciences

zhanghainan@software.ict.ac.cn, {lanyanyan, guojiafeng, junxu, cxq}@ict.ac.cn

## Abstract

Sequence to sequence (Seq2Seq) models have been widely used for response generation in the area of conversation. However, the requirements for different conversation scenarios are distinct. For example, customer service requires the generated responses to be specific and accurate, while chatbot prefers diverse responses so as to attract different users. The current Seq2Seq model fails to meet these diverse requirements, by using a general average likelihood as the optimization criteria. As a result, it usually generates safe and commonplace responses, such as *'I don't know'*. In this paper, we propose two tailored optimization criteria for Seq2Seq to different conversation scenarios, i.e., the maximum generated likelihood for specific-requirement scenario, and the conditional value-at-risk for diverse-requirement scenario. Experimental results on the Ubuntu dialogue corpus (Ubuntu service scenario) and Chinese Weibo dataset (social chatbot scenario) show that our proposed models not only satisfies diverse requirements for different scenarios, but also yields better performances against traditional Seq2Seq models in terms of both metric-based and human evaluations.

## 1 Introduction

This paper focuses on the problem of the single-turn dialogue generation, which is critical in many natural language processing applications such as customer services, intelligent assistant and chatbot. Recently, sequence to sequence (Seq2Seq) models (Sutskever et al., 2014) have been widely

used in this area. In these Seq2Seq models, a recurrent neural network (RNN) based encoder is first utilized to encode the input post to a vector, and another RNN decoder is then used to automatically generate the response word by word. The parameters of the encoder and decoder are learned by maximizing the averaged likelihood of the training data.

It is clear that the requirements for generated responses are distinct in different dialogue scenarios. For instance, in the scenario of customer service or mobile assistant, users mainly expect the system to help them solve a problem. Therefore, the responses should be specific and accurate to provide useful assistance. For example, if the user asks a question *'How can I get the AMD driver running on Ubuntu 12.10?'*, the system is expected to reply *'The fglrx driver is in the repo. But it may depend on your exact chipset.'*, rather than *'I do not know about the package.'*, even though the latter can also be viewed as relevant for the proposed question. We called this kind of scenario as specific-requirement scenario. While in other scenarios such as chatbot, users are interacting with the dialogue system for fun. Therefore, the generated responses should be diverse to attract different users. Take the post *'Can you recommend me a tourist city?'* as an example. If the user prefers the magnificent mountains and rivers, it is better to reply *'You may like the Bernina Express to the Alps'*. While if the user loves literature, it is better to reply *'Paris is a beautiful city with full of the literary atmosphere'*. This kind of scenario is called diverse-requirement scenario.

However, the current generation model Seq2Seq (Sutskever et al., 2014) usually tend to generate common responses, such as *'I don't know'* and *'What does this mean?'* (Li et al., 2016a,b; Zhou et al., 2017), which fails to meet diverse requirements for different conversation

scenarios. Intrinsically, conversation is a typical one-to-many application, i.e., multiple responses with different semantic meanings are correspondent to a same post. That means there are various post-response matching patterns in the training data. Seq2Seq optimizes an averaged likelihood, so it can only capture the common matching patterns, leading to common responses.

The purpose of this paper is to propose two tailored optimization criteria for Seq2Seq models to accommodate different conversation scenarios, i.e. specific-requirement scenario and diverse-requirement scenario. The key idea is to how capture the required post-response matching patterns. For the specific-requirement scenario, we define the maximum generated likelihood as the objective function. With this kind of criterion, we just require one ground-truth response to be close to the given post, instead of requiring the average of multiple ground-truth responses to be close to the post. Therefore, the most significant post-response matching pattern will be learned from the data, to facilitate generating a specific response. While for the diverse-requirement scenario, the conditional value-at-risk (CVaR) is used as the objective function. CVaR is a risk-sensitive function widely used in finances (Rockafellar and Uryasev, 2002; Alexander et al., 2006; Chen et al., 2015), defined to assessing the likelihood (at a specific confidence level) that a specific loss will exceed the value at risk. With CVaR as the objective function, the worst  $1-\alpha$  responses are required to be close to the post, therefore various post-response patterns can be captured, and the learned model has the ability to generate diverse responses.

We use public data to evaluate our proposed models. For the specific-requirement scenario, the experiments on public Ubuntu dialogue corpus (Ubuntu service) show that optimizing the maximum generated likelihood produces more specific and accurate responses than traditional Seq2Seq models. While for the diverse-requirement scenario, the experiments on the public Chinese Weibo dataset (social chatbot) show that optimizing CVaR produces diverse responses, as compared with Seq2Seq and the variants.

## 2 Related Work

The basic neural-based Seq2Seq framework for dialogue generation is inspired by the studies of statistical machine translation. Sutskever et

al. (Sutskever et al., 2014) proposed the original Seq2Seq framework (Seq2Seq), which used a multilayered Long Short-Term Memory (LSTM) to map the input sequence to a fixed dimension vector and then used another LSTM to decode the target sequence from the vector. Then Cho et al. (Cho et al., 2014) followed the above architecture, and proposed to feed the last hidden state of encoder to every cell of decoder (RNN-encdec), which enhanced the influence of contexts in generating each word of the targets. To further alleviate the long dependency problem, Bahdanau et al. (Bahdanau et al., 2015) introduced the attention mechanism into the neural network and achieved encouraging performances (Seq2Seq-att). Many studies (Shang et al., 2015; Vinyals and Le, 2015) directly applied the above neural SMT models to the task of dialogue generation, and gained some promising performances.

Although the current Seq2Seq model is capable to generate fluent responses, these responses are usually general. Therefore, many researchers focused on how to improve the generation quality and specification. Li et al. (Li et al., 2016a) proposed a mutual information model (MMI) to tackle this problem. However, it is not a unified training model, instead it still trained original Seq2Seq model, and used the Maximum Mutual Information criterion only for testing to rerank the primary top-n list. Mou et al. (Mou et al., 2017) proposed a forward-backward keyword method which used a pointwise mutual information to predict a noun as a keyword and then used two Seq2Seq models to generate the forward sentence and the backward sentence. Xing et al. (Xing et al., 2017) proposed a joint attention mechanism model, which modified the generation probability by adding the topic keywords likelihood to the generated maximum likelihood with extra corpus. The recent works such as seqGAN (Yu et al., 2017) and Adver-REGS (Li et al., 2017) try to use Generative Adversarial Networks (GAN) for generation, where the discriminator scores are used as rewards for reinforcement learning.

For the study of generating diverse responses, Vijayakumar et al. (Vijayakumar et al., 2016) introduced a diverse beam search which decoded a list of diverse outputs by optimizing for a diversity-augmented objective, which can control for the exploration and exploitation of the search space. Zhou (Zhou et al., 2017) proposed to apply

a hidden state as a generating style (Mechanism). They make an assumption that some latent responding mechanisms can generate different responses, and model these mechanisms as latent embedding. With these latent embedding in the mid of Seq2Seq, the mechanism-aware Seq2Seq can generate different mechanism responses.

However, most of these models are using an averaged approach for optimization, similar to that in Seq2Seq. This paper proposes two new criteria for different conversation scenarios. For the specific-requirement scenario, the maximum generated likelihood is used as the objective function. While for the diverse-requirement scenario, CVaR is used for optimization.

### 3 Sequence to Sequence Models

We first introduce the typical LSTM-based Seq2Seq framework (Bahdanau et al., 2015) used in dialogue generation.

Given a post  $X = \{x_1, \dots, x_M\}$  as the input, a standard LSTM first maps the input sequence to a fixed-dimension vector  $h_M$  as follows.

$$\begin{aligned} i_k &= \sigma(W_i[h_{k-1}, w_k]), & f_k &= \sigma(W_f[h_{k-1}, w_k]), \\ o_k &= \sigma(W_o[h_{k-1}, w_k]), & l_k &= \tanh(W_l[h_{k-1}, w_k]), \\ c_k &= f_k c_{k-1} + i_k l_k, & h_i &= o_k \tanh(c_k), \end{aligned} \quad (1)$$

where  $i_k$ ,  $f_k$  and  $o_k$  are the input gate, the memory gate, and the output gate, respectively.  $w_k$  is the word embedding for  $x_k$ , and  $h_k$  stands for the vector computed by LSTM at time  $k$  by combining  $w_k$  and  $h_{k-1}$ .  $c_k$  is the cell at time  $k$ , and  $\sigma$  denotes the sigmoid function.  $W_i$ ,  $W_f$ ,  $W_o$  and  $W_l$  are parameters.

Then another LSTM is used as the decoder to map the vector  $h_M$  to the ground-truth response  $Y = \{y_1, \dots, y_N\}$ . Typically, the decoder is trained to predict the next word  $g_i$ , given the context vector  $h_M$  and the previous generated words  $\{g_1, \dots, g_{i-1}\}$ . In other words, the decoder defines a probability over the output  $Y$  by decomposing the joint probability into the ordered conditionals by chain rule in the probability theory:

$$\begin{aligned} P(Y|X) &= \prod_{i=1}^N p(y_i | h_M, y_1, \dots, y_{i-1}) \\ &= \prod_{i=1}^N g(h_M, y_{i-1}, h'_i), \end{aligned}$$

where  $g_\theta$  is a softmax function,  $h'_i$  is the hidden state in the decoder LSTM.

Usually the attention mechanism is further introduced to the above Seq2Seq framework in real applications. Instead of using  $h_M$  as the context vector in the decoder, we let the context vector, denoted as  $s_i$ , to be dependent on the sequence  $(h_1, \dots, h_M)$ . Each  $h_k$  contains information about the input sequence with a strong focus on the parts surrounding the  $k$ -th word of the input sentence. The context vector  $s_i$  is then computed as a weighted sum of these  $h_k$ :

$$s_i = \sum_{k=1}^M \alpha_{ik} h_k.$$

The weight  $\alpha_{ik}$  of each representation  $h_k$  is computed by:

$$\begin{aligned} \alpha_{ik} &= \frac{\exp(e_{ik})}{\sum_{j=1}^M \exp(e_{ij})}, \\ e_{ik} &= v^T \tanh(W_1 h'_{i-1} + W_2 h_k), \end{aligned}$$

where  $v^T$ ,  $W_1$  and  $W_2$  are learned parameters.  $e_{ik}$  is an alignment model which scores how well the inputs around position  $k$  and the output at position  $i$  match. The score is based on the LSTM hidden state  $h'_{i-1}$  (just before emitting  $y_i$ ), and  $h_k$  of the input sentence.

Given a set of training data  $\mathcal{D}$ , Seq2Seq assumes that data are i.i.d. sampled from a probability  $P$ , and uses the following log likelihood as the objective for maximization:

$$\mathcal{L} = \sum_{(X,Y) \in \mathcal{D}} \log P(Y|X). \quad (2)$$

### 4 Tailored Sequence to Sequence Models

We can see that a general averaged likelihood of the training data is used as the objective function in Seq2Seq. However, this objective function is usually criticized for generating common responses, such as ‘*I don’t know*’ and ‘*What does this mean?*’. Clearly, this kind of responses cannot satisfy either the specific or the diverse requirements. The underlying reason is not difficult to understand. Intrinsically, conversation is a typical one-to-many application, i.e., multiple responses with different semantic meanings are correspondent to a same post. That means there are various post-response matching patterns in the

training data. If we optimize an averaged likelihood, we can only capture the common matching patterns, which leads to generating common responses. Therefore, if we want to generate specific responses, we need to capture the most significant matching pattern; while if we want to generate diverse responses, we need to define a criteria which has the ability to capture the various matching patterns. Motivated by this idea, we propose two optimization criteria, i.e. maximum generated likelihood, and CVaR, to adapt two different scenarios.

#### 4.1 Maximum Generated Likelihood Criteria

To meet the specific requirement, we need to capture a specific matching pattern between post and response, rather than the common matching pattern. Therefore, instead of optimizing the averaged likelihood, we turn to use the maximum generated likelihood (MGL) as the objective function. Mathematically, for a given post  $X$  and its associated ground-truth responses  $(Y_X^{(1)}, Y_X^{(2)}, \dots, Y_X^{(m_X)})$ , the objective function is defined as:

$$\mathcal{L} = \sum_X \max_{k=1}^{m_X} \log P(Y_X^{(k)}|X).$$

From the definition, we can see that we aim to capture the most significant post-response matching pattern in the training data. Therefore, the learned model can output specific responses for a given post. Since there is a max operator in the objective function, which is difficult for accurate optimization, we approximate it by the softmax function. Then the objective function becomes the following form:

$$\mathcal{L} = \sum_X \sum_{k=1}^{m_X} \log \frac{P(Y_X^{(k)}|X)}{\sum_{j=1}^{m_X} P(Y_X^{(j)}|X)}.$$

If the probability for one ground-truth  $Y_X^{(k)}$  is small, it contributes little to the objective function. That is to say, we just require the top ground-truth responses with relative large probabilities to be close to the post.

#### 4.2 CVaR Criteria

To meet the diverse requirements, we need to capture various matching patterns between post and its multiple ground-truth responses. Therefore, instead of optimizing the averaged likelihood, we turn to optimize the conditional value-at-risk, named CVaR for short. CVaR is a prominent risk measure used extensively in finance, and

it is proved to be coherent (Artzner et al., 1999) and numerically effective (Krokhmal et al., 2002; Uryasev, 2013).

The definitions of VaR and CVaR are as follows. For a confidence level  $\alpha \in [0, 1]$ , and a continuous random cost  $Z$  whose distribution is parameterized by a controllable parameter  $\theta$ , the  $\alpha$ -VaR of the cost  $Z$ , denoted by  $\nu_\alpha(\theta)$ , is defined as:

$$\nu_\alpha(\theta) = \inf\{\nu \in \mathbb{R} | P(Z \leq \nu) \geq \alpha\}.$$

$\alpha$ -VaR denotes the maximum cost that might be incurred with probability at least  $\alpha$ , or can be simply regarded as the  $\alpha$ -quantile of  $Z$ . And the  $\alpha$ -CVaR, denoted by  $\Phi_\alpha(\theta)$ , is defined as:

$$\Phi_\alpha(\theta) = \frac{1}{1-\alpha} \int_\alpha^1 \nu_r(\theta) dr = \mathbb{E}^\theta[Z | Z \geq \nu_\alpha(\theta)].$$

It can be viewed as the expected cost over the  $(1-\alpha)$  worst outcomes of  $Z$ .

Applying CVaR to generating diverse responses, we can define the random cost  $Z$  as  $-\log P(Y|X)$ , the corresponding CVaR is:

$$\Phi_\alpha(\theta) = \frac{1}{1-\alpha} \int_\alpha^1 \nu_r(\theta) dr,$$

where  $\nu_r(\theta) = \inf\{\nu \in \mathbb{R} | P(-\log P(Y|X) \leq \nu) \geq r\}$ , and  $\theta$  are parameters of the Seq2Seq model. Therefore, we have:

$$\nu_r(\theta) = \inf\{\nu \in \mathbb{R} | P(P(Y|X) \geq e^\nu) \geq r\}.$$

Therefore, for a given post  $X$  and its ground-truth responses  $(Y_X^{(1)}, Y_X^{(2)}, \dots, Y_X^{(m_X)})$ , optimizing CVaR is equivalent to maximizing the following objective function:

$$\mathcal{L} = \sum_X \frac{1}{1-\alpha} \sum_{Y_X^{(k)} \in \mathcal{Y}_{1-\alpha}} P(Y_X^{(k)}|X),$$

where  $\mathcal{Y}_{1-\alpha}$  is a collection of ground-truth responses such that:

$$\sup\{P(Y_X^{(i)}|X) : Y_X^i \in \mathcal{Y}_{1-\alpha}\} \leq \alpha.$$

We can see that maximizing the above objective function requires the worst  $1-\alpha$  responses to be close to the post. Therefore, we aim to capture each distinct post-response matching pattern by optimizing the CVaR criteria, which can meet the requirement for generating diverse responses.

## 5 Experiments

In this section, we conduct experiments on both specific-requirement and diverse-requirement scenarios, to evaluate the performances of our proposed methods.

### 5.1 Experimental Settings

#### 5.1.1 Datasets

We use two public datasets in our experiments. For the specific-requirement scenario, we use the Ubuntu dialogue corpus<sup>1</sup> extracted from Ubuntu question-answering forum, named Ubuntu (Lowe et al., 2015). The original training data consists of 7 million conversational post-responses pairs from 2014 to April 27,2012. The validation data are conversational pairs from April 27,2014 to August 7,2012, and the test data are from August 7,2012 to December 1,2012. We set the number of positive examples as 4,000,000 in the Github to directly sample data from the whole corpus. Then we construct post and response pairs based on the period from both context and utterance. We also conduct some data pro-processing. For example, we use the official script to tokenize, stem and lemmatize, and the duplicates and sentences with length less than 5 or longer than 50 are removed. Finally, we obtain 3,200,000, 100,000 and 100,000 for training, validation and testing, respectively.

For the diverse-requirement scenario, we use the Chinese Weibo dataset, named STC (Shang et al., 2015). It consists of 3,788,571 post-response pairs extracted from the Chinese Weibo website and cleaned by the data publishers. We randomly split the data to training, validation, and testing sets, which contains 3,000,000, 388,571 and 400,000 pairs, respectively.<sup>2</sup>

#### 5.1.2 Baseline Methods

Six baseline methods are used for comparison, including traditional Seq2Seq (Sutskever et al., 2014), RNN-encdec (Cho et al., 2014), Seq2Seq with attention(Seq2Seq-att) (Bahdanau et al., 2015), mutual information(MMI) (Li et al., 2016b), Adver-REGS (Li et al., 2017) and Mechanism model (Zhou et al., 2017). Here are some empirical settings. We first introduce the input em-

beddings. For STC, we utilize character-level embeddings rather than word-level embeddings, due to the word sparsity, segmentation mistakes and unknown Chinese words which may lead to inferior performance (Hu et al., 2015). For Ubuntu, we use word embeddings trained by word2vec on the training dataset. In the training process, the dimension is set to be 300, the size of negative sample is set to be 3, and the learning rate is 0.05. For fair comparison among all the baseline methods and our methods, the number of hidden nodes is all set to 300, and batch size is set to 200. Stochastic gradient decent (SGD) is utilized in our experiment for optimization, instead of Adam, because SGD yields better performances in our experiments. The learning rate is set to be 0.5, and adaptively decays with rate 0.99 in the optimization process. We run our model on a Tesla K80 GPU card with Tensorflow framework. All the methods are pretrained with the same Seq2Seq model. For maximum generated likelihood(MGL) model, some people may argue that the specific results may be due to the usage of single post-response pair. Thus we also implement the baseline of using a single post-response pair, by random selecting the response from the ground-truth for each post, denoted as Single Model.

#### 5.1.3 Evaluation Measures

We use both quantitative metrics and human judgements to evaluate the proposed MGL model and the CVaR model. Specifically, we use two kinds of metrics for quantitative comparisons. The first one kind is the traditional metric, including PPL and Bleu score (Xing et al., 2017). They are both widely used in natural language processing, and here we use them to evaluate the quality of the generated responses. The other kind is to evaluate the specific degree<sup>3</sup> in (Li et al., 2016a,b). It measures the specific degree of the generated responses, by calculating the number of distinct unigrams and bigrams in the generated responses, denoted as *distinct*. If a model usually generates common responses, the *distinct* will be low.

For the diverse-requirement scenario, we define two measures to evaluate the performance. Specifically, we set the beam as 10. Group-diversity is

<sup>3</sup>Though it is named as diversity in Li’s paper, this diversity is not the same as that used in our paper. This diversity measures the specific degree of the generated responses over all generations. While the diversity used in our paper means that the responses are required to be relevant to a post from different aspects.

<sup>1</sup><https://github.com/rkadlec/ubuntu-ranking-dataset-creator>

<sup>2</sup><https://github.com/zhanghainan/TailoredSeq2Seq2-DifferentConversationScenarios>

model	distinct-1	distinct-2	BLEU	PPL
Seq2Seq	0.140	1.11	1.231	51.26
RNN-encdec	0.125	1.24	1.231	46.97
Seq2Seq-att	0.351	4.36	1.294	47.84
MMI	0.283	4.84	1.297	42.52
Adver-REGS	0.268	5.07	1.279	37.71
Single	0.324	5.27	1.342	30.36
MGL	<b>0.358</b>	<b>6.30</b>	<b>1.354</b>	<b>26.34</b>
CVaR	0.294	5.52	1.290	30.03

Table 1: The metric-based evaluation results(%) of different models on Ubuntu.

defined to calculate the difference between each two generations for one post, denoted as *divrs*. Group-overlap is defined to calculate the overlap between each two generations for one post, denoted as *overlap*. The detailed definitions are shown as follows.

$$divrs = \frac{1}{N} \sum_{i=1}^N \sum_{X_i} cosine(G_{i1}, G_{i2}),$$

$$overlap = \frac{1}{N} \sum_{i=1}^N \sum_{X_i} overlap(G_{i1}, G_{i2}),$$

where  $G_{i1}$  and  $G_{i2}$  are the generated responses from the model for post  $X$ ,  $cosine(G_{i1}, G_{i2})$  is the cosine similarity, and the  $overlap(G_{i1}, G_{i2})$  is defined as the intersection divided by union.

For human evaluation, given 200 randomly sampled post and it’s generated responses, three annotators, randomly selected from a class of computer science majored students(48 students), are required to give 3-graded judgements. The annotation criteria are defined as follows:

1. the response is nonfluent or has wrong logic; or the response is fluent but not related with the post;
2. the response is fluent and weak related, but it’s common which can reply many other posts;
3. the response is fluent and strong related with its post, which is like following a real person’s tone.

## 5.2 Specific-Requirement Scenario

We demonstrate the experimental results on the specific-requirement scenario, based on the Ubuntu dataset.

### 5.2.1 Metric-based Evaluation

The quantitative evaluation results are shown in Table 1. From the results, we can see that both

model	human score distribution(%)			Ave.	Kappa
	1	2	3		
Seq2Seq-att	46.5	38.6	14.9	1.684	0.387
MMI	42	38	20	1.78	0.395
Adver-REGS	42	26	32	1.9	0.379
Single	49	14	37	1.88	0.383
MGL	33	16	51	2.18	0.372
CVaR	40	12	48	2.08	0.381

Table 2: The comparisons of different models by human evaluation on Ubuntu.

MMI and Adver-REGS outperform Seq2Seq baselines in terms of BLEU, PPL and *distinct* measures. That’s because both MMI and Adver-REGS further consider some reward functions in the optimization process to encourage specific results. Specifically, MMI uses a predefined reward function to penalize generating common responses, and Adver-REGS uses a learned discriminator to define the reward function. Our MGL model obtains higher BLEU and lower PPL than baseline models. Take the BLEU score on Ubuntu dataset for example, the BLEU score of MGL model is 1.354, which is significantly better than that of MMI and Adver-REGS, i.e., 1.297 and 1.279. These results indicate that our MGL generates responses with higher quality. When compared with the Single model, MGL is also better because MGL considers more data in the model computation process. The *distinct* scores of MGL are also higher than baseline models, which indicate that our model can generate more specific responses. That’s because it has the ability to learn the significant matching pattern between post and responses, by optimizing the maximum generated likelihood rather than the averaged one. In summary, our maximum generated likelihood model produces more fluent and specific results, as compared with baseline methods.

### 5.2.2 Human Evaluation

The human evaluation results are shown in Table 2, in which the percentage of sentences belonging to each grade and the averaged grade are demonstrated to evaluate the quality of generated responses. Kappa (Fleiss, 1971) value is presented to demonstrate the consistency of different annotators. From the results, we can see that MGL significantly outperforms baseline methods. The averaged score of MGL Model is 2.18, which is much higher than that of MMI and Adver-REGS, i.e., 1.78 and 1.9, respectively. The percentage of

post	Install grub to the hdd where the bios is set to boot.
response1	So I would not just install it to /boot?
Seq2Seq-att	A log file a bug with the install cd.
MMI	What do you mean?
Adver-REGS	How do I install grub?
Single	Do you delete grub?
MGL	I have boot with the window from the grub and it go from the boot menu.
CVaR	If you have grub on the same pc, it is 0 0 boot partition.
post	I want to have a backup , sad I use up all my space with dual boot and game.
response1	Oh you can get a sata to usb adapt : )
Seq2Seq-att	What about your computer ?
MMI	It is not in the install , you can do that.
Adver-REGS	Game game game game.
Single	And other user will do too ?
MGL	And do you buy direct driver from the kernel and it cause run ?
CVaR	Yes , you can also use a text file to your file .
post	Take a look at install .
response1	I am to cd to the directories contain the folder.
Seq2Seq-att	How do you install it?
MMI	I think it be a good idea to do that.
Adver-REGS	I have no idea what I am looking for.
Single	I think it is a bite , but I do not know a good thing to do that. I am use.
MGL	I think so, I have a lot of nautilus. I am already install.
CVaR	I just install it from synaptics, but I want to install it on the same repository.

Table 3: The generated responses from different models on Ubuntu.

strongly related sentences (i.e., the grade ‘3’) of MGL Model is 51%, which is also higher than that of MMI, Adver-REGS and Single Model, i.e., 20% , 32% and 37%. In summary, our maximum generated likelihood model produces better responses compared with baselines. As compared with MMI and Adver-REGS, both the metric-based improvements and human evaluation improvements of MGL are significant on Ubuntu datasets (p-value < 0.01).

### 5.2.3 Case Study

Here we show some generated responses for demonstration. Specifically, Table 3 gives one example post and its ground-truth responses from Ubuntu. We also list the generated responses from different models. We can see that Seq2Seq-att, MMI and Adver-REGS all produce common responses, such as ‘What do you mean?’, ‘I have no idea what I am looking for.’ and ‘What about your computer?’. Our models give interesting responses with specific meanings. Take the post ‘Install grub to the hdd where the bios is set to boot.’ as an example, our model conveys more specific information by replying ‘I have boot with the window from the grub and it go from the boot menu.’. And in another case, for the given post ‘I want to have a backup , sad I use up all my space with dual boot and game.’, our MGL model generates a question for the post ‘And do you buy direct driver from the kernel and it cause run?’, which is more intelligent. Similar observations have been obtained for many other posts, and we omit them for space limitations.

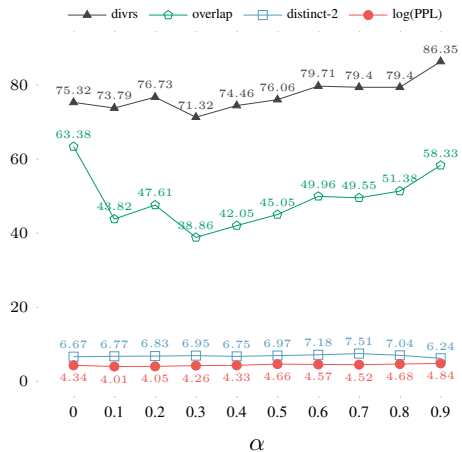


Figure 1: Influences of different  $\alpha$  in CVaR.

model	BLEU	PPL	overlap	divrs
Seq2Seq	1.616	132.93	67.26	87.83
RNN-encdec	1.636	130.56	65.72	87.85
Seq2Seq-att	1.620	76.95	63.38	85.32
Adver-REGS	1.635	84.77	57.96	84.94
Mechanism	1.642	90.48	57.67	84.64
MGL	<b>1.703</b>	<b>36.25</b>	66.92	86.22
CVaR	1.652	70.94	<b>38.96</b>	<b>71.38</b>

Table 4: The metric-based evaluation results(%) of different models on STC.

## 5.3 Diverse-Requirement Scenario

Now we introduce the experimental results for the diverse-requirement scenario, based on STC.

### 5.3.1 Parameters Setting

First, we study the influences of different parameter  $\alpha$  in CVaR. Specifically, we show the validation result with  $\alpha$  ranging from 0 to 0.9 with step 0.1, to see the change of CVaR performances. Figure 1 show the results of different  $\alpha$  in terms of divrs , overlap, distinct-2 and PPL. From the results, we can see that the performances of divrs , overlap and PPL are all changing in a similar trend, i.e. first drop and then increase. The best  $\alpha$  for CVaR is 0.3, which is used in the following experiments.

### 5.3.2 Metric-based Evaluation

The quantitative evaluation results are shown in Table 4. From the results, we can see that both Adver-REGS and Mechanism outperform Seq2Seq models in terms of BLUE and PPL measures. That’s because they both use some techniques to enhance the generation ability. Adver-REGS uses a learned discriminator to define the reward function, while Mechanism uses a style

model	human score distribution(%)			Ave.	Kappa
	1	2	3		
Seq2Seq-att	54.5	21	24.5	1.7	0.452
MMI	56	15.5	28.5	1.725	0.447
Adver-REGS	48.5	20	31.5	1.83	0.436
Mechanism	52.5	17.5	30	1.775	0.427
MGL	37	11	52	2.15	0.451
CVaR	44.5	11.5	44	1.995	0.437

Table 5: The comparisons of different models by human evaluation on STC.

hidden state to describe the generation mechanism. Both MGL and CVaR obtain better results in terms of BLUE and PPL, compared with other baselines. These results indicate that our proposed models generate more fluent responses in the diverse-requirement scenario. As for the evaluation for the diversity, we can see that CVaR model obtains the lowest overlap and divrs among all the baseline models. Take the overlap score on STC for example, the overlap score of CVaR model is 38.86, which is significantly lower than that of Adver-REGS, Mechanism and GLM, i.e., 57.96, 57.67 and 66.92. These results indicate that our CVaR model can generate responses with higher diversity. That’s because it has the capability to capture various matching patterns in the training data, by optimizing the worst  $1 - \alpha$  costs. Therefore, our CVaR model produces both fluent and diverse results, as compared with baseline methods.

### 5.3.3 Human Evaluation

The human evaluation results are shown in Table 5. From the results, we can see MGL and CVaR models achieve comparable results, which are significantly better than baseline methods. Specifically, the averaged score of MGL and CVaR is 2.15 and 1.995, which is significantly higher than that of Adver-REGS and Mechanism, i.e., 1.83 and 1.775, respectively. The percentage of strongly related sentences (i.e., the grade ‘3’) of MGL Model and CVaR are 52% and 44%, which are also significantly higher than that of Adver-REGS and Mechanism, i.e., 31.5% and 30%. We conducted significant test for the improvement. As compared with Adver-REGS and Mechanism, both the metric-based improvements and human evaluation improvements of CVaR are significant on STC datasets ( $p$ -value  $< 0.01$ ).

### 5.3.4 Case Study

Here we show some generated responses for demonstration. Specifically, Table 6 gives one

post	总决赛继续等待韦德。(Waiting for Wade in the final games.)
response	每个人都有每个人的喜爱(Everyone has his favorite stars.)
response	分析比新浪分析的好多了(Analysis is much better than Sina)
response	等待闪电侠彻底爆发! (Waiting for the explosion of Mr.Flash)
Mechanism	韦德真伟大啊!支持!(Wade is really great! Support him!)
Mechanism	韦德越来越好。(Wade will be better)
Mechanism	韦德威武!(Wade is mighty)
CVaR	决赛一定要去看看的!(I must go and see the final games)
CVaR	詹皇怎么还能这么快啊!(James is so fast)
CVaR	决赛是一种对对手的打击,热火加油 (The final games is a blow for the opposite. Heat come on)
post	神舟八号飞船与天宫一号第二次对接成功。 (Shenzhou 8 spacecraft and Tiangong-1 has the second successful docking)
response	希望其它方面也赶快领先世界(Hope other aspects will soon lead the world)
response	过硬的技术,紧密地配合促成了这次成功。 Strong technology and close cooperation contributed to this success
response	下一次就是载人对接啦。(Next will have a human in it)
Mechanism	周末报还会直播?(Will broadcast in the Weekend Newspaper?)
Mechanism	不错,就是不错(It is really good)
Mechanism	不错,就是在一起。(It is really good, they should be together)
CVaR	是啊,还是要坚持在一起。(Yes, they should insisted on being together)
CVaR	您这是在看头版吗?(Are you see it in the front page of the newspaper?)
CVaR	不错,有空推荐给爸爸! (It is really good, you could recommend it to your father if you have time)

Table 6: The generated responses from different models on STC.

post and its three ground-truth responses from STC. We also give three generated responses from Mechanism and CVaR model. We can see that Mechanism produces responses with the same meaning, such as ‘Wade is so amazing’ and ‘It is really good’. However, our CVaR models give specific responses with diverse meanings. Take the post ‘Waiting for Wade in the final games.’ for example, CVaR’s responses are related to different topics. The response ‘I must go and see the final games’ focuses on the game, while another response of ‘James is so fast’ focuses on the person, James. For the other case, the post is about the docking of two spacecrafts and the CVaR responses are related to different users, such as the supporter of the event, the newspaper reader and the children who have a father concerned with the current news. We have obtained similar observations for many other posts, but we have to omit them for space limitations.

## 6 Conclusion

In this paper, we propose two new optimization criteria for Seq2Seq model to adapt different conversation scenario. For the specific-requirement scenario, such as customer service, which requires specific and high quality responses, maximum generated likelihood is used as the objective function instead of the averaged one. While for the diverse-requirement, such as chatbot, which requires diverse and high quality responses even if for the same post, CVaR is used as the objective function for worst case optimization. Experimental results on both specific-requirement



(Ubuntu data) and diverse-requirement scenarios (STC data) demonstrate that the proposed optimization criteria can meet the corresponding requirement, yielding better performances against traditional Seq2Seq models in terms of both metric-based and human evaluations.

The contribution of this paper is to use tailored Seq2Seq model for different conversation scenarios. The study shows that if we want to generate specific responses, it is important to design the model to learn the most significant matching pattern between post and response. While if we want to generate diverse responses, a risk-sensitive objective functions is helpful. In future work, we plan to further investigate the impact of risk-sensitive objective functions, including the relations between model robustness and diverse generations.

## Acknowledgments

This work was funded by the 973 Program of China under Grant No. 2014CB340401, the National Natural Science Foundation of China (NSFC) under Grants No. 61425016, 61472401, 61722211, 61773362, and 20180290, the Youth Innovation Promotion Association CAS under Grants No. 20144310, and 2016102, and the National Key R&D Program of China under Grants No. 2016QY02D0405.

## References

- S. Alexander, T. F. Coleman, and Y. Li. 2006. Minimizing cvar and var for a portfolio of derivatives. *Journal of Banking and Finance* 30(2):583–605.
- P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. 1999. Coherent measures of risk mathematical finance 9. *Mathematical Finance Theory Modeling Implementation* volume 9(3):203–228(26).
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. *The International Conference on Learning Representations* .
- Youhua Chen, Minghui Xu, and Zhe George Zhang. 2015. Technical note—a risk-averse newsvendor model under the cvar criterion. *Operations Research* 57(4):1040–1044.
- Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *Computer Science* .
- Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *American Psychological Association* .
- Baotian Hu, Qingcai Chen, and Fangze Zhu. 2015. Lc-sts: A large scale chinese short text summarization dataset. *arXiv preprint arXiv:1506.05865* .
- Pavlo Krokhmal, Jonas Palmquist, and Stanislav Uryasev. 2002. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk* 4:11–27.
- Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016a. A diversity-promoting objective function for neural conversation models. *The North American Chapter of the Association for Computational Linguistics* .
- Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, and Dan Jurafsky. 2016b. Deep reinforcement learning for dialogue generation. *The Conference on Empirical Methods in Natural Language Processing* .
- Jiwei Li, Will Monroe, Tianlin Shi, Alan Ritter, and Dan Jurafsky. 2017. Adversarial learning for neural dialogue generation. *The Conference on Empirical Methods in Natural Language Processing* .
- Ryan Lowe, Nissan Pow, Iulian Serban, and Joelle Pineau. 2015. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. *Computer Science* .
- Lili Mou, Yiping Song, Rui Yan, Ge Li, Lu Zhang, and Zhi Jin. 2017. Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation. *The Annual Meeting of the Association for Computational Linguistics* .
- R. Tyrrell Rockafellar and Stanislav Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26(7):1443–1471.
- Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. *The Annual Meeting of the Association for Computational Linguistics* .
- Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *The Annual Conference on Neural Information Processing Systems*. pages 3104–3112.
- Stanislav Uryasev. 2013. Probabilistic constrained optimization: methodology and applications. *Springer Science and Business Media* .
- Ashwin K Vijayakumar, Michael Cogswell, Ramprasath R. Selvaraju, Qing Sun, Stefan Lee, David Crandall, and Dhruv Batra. 2016. Diverse beam search: Decoding diverse solutions from neural sequence models. *arXiv* .

Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *Computer Science* .

Chen Xing, Wei Wu, Yu Wu, Jie Liu, Yalou Huang, Ming Zhou, and Wei-Ying Ma. 2017. Topic aware neural response generation. In *The Association for the Advancement of Artificial Intelligence*. pages 3351–3357.

Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. 2017. Seqgan: Sequence generative adversarial nets with policy gradient. In *The Association for the Advancement of Artificial Intelligence*. pages 2852–2858.

Ganbin Zhou, Ping Luo, Rongyu Cao, Fen Lin, Bo Chen, and Qing He. 2017. Mechanism-aware neural machine for dialogue response generation. In *The Association for the Advancement of Artificial Intelligence*. pages 3400–3407.