Balancing Quality and Human Involvement: An Effective Approach to Interactive Neural Machine Translation

Tianxiang Zhao,[1]* Lemao Liu,[2] Guoping Huang,[2] Zhaopeng Tu,[2] Huayang Li,[2]
Yingling Liu,[3] Guiquan Liu,[3] Shuming Shi[2]
[1]Penn State University, [2]Tencent AI Lab
[3]University of Science and Technology of China
tkz5084@psu.edu, {gqliu, ylliu22}@ustc.edu.cn
{redmondliu, donkeyhuang, zptu, alanili, shumingshi}@tencent.com

Abstract

Conventional interactive machine translation typically requires a human translator to validate every generated target word, even though most of them are correct in the advanced neural machine translation (NMT) scenario. Previous studies have exploited confidence approaches to address the intensive human involvement issue, which request human guidance only for a few number of words with low confidences. However, such approaches do not take the history of human involvement into account, and optimize the models only for the translation quality while ignoring the cost of human involvement. In response to these pitfalls, we propose a novel interactive NMT model, which explicitly accounts the history of human involvements and particularly is optimized towards two objectives corresponding to the translation quality and the cost of human involvement, respectively. Specifically, the model jointly predicts a target word and a decision on whether to request human guidance, which is based on both the partial translation and the history of human involvements. Since there is no explicit signals on the decisions of requesting human guidance in the bilingual corpus, we optimize the model with the reinforcement learning technique which enables our model to accurately predict when to request human guidance. Simulated and real experiments show that the proposed model can achieve higher translation quality with similar or less human involvement over the confidence-based baseline.

Introduction

Recent years have witnessed a breakthrough in neural machine translation (NMT) (Bahdanau, Cho, and Bengio 2015; Vaswani et al. 2017), but its translation quality is still incapable of satisfying the requirements in many industrial applications. Interactive machine translation (IMT) (Foster, Isabelle, and Plamondon 1997; Langlais, Foster, and Lapalme 2000), where human and machines collaborate to translate in a joint strategy, has thereby drawn much research attention (Green et al. 2014; Wuebker et al. 2016; Knowles and Koehn 2016; Peris, Domingo, and Casacuberta 2017). In the conventional IMT, a user and machine collaboratively generate the translation from left-to-right: at each time the user validates all words in a prefix and corrects one of them, and then the machine generates new words based on the corrected prefix until the translation process is finished.

Despite its appealing performance, the user has to validate *all* words generated by a machine although most of them are actually correct, leading to considerable human involvement (Ueffing and Ney 2005). Many efforts (González-Rubio, Ortiz-Martínez, and Casacuberta 2010b; Cheng et al. 2016; Knowles and Koehn 2018) thereby have been made to reduce human involvement based on the notion of confidence estimation (Blatz et al. 2004; González-Rubio, Ortiz-Martínez, and Casacuberta 2010a; Lam, Kreutzer, and Riezler 2018). They query for human guidance for *few* of those words with low confidences so that prevent manually validating other words with high confidences. However, in these confidence based approaches, the confidence model is either an external classifier (Ueffing and Ney 2005; Cheng et al. 2016) or inherited from the standard translation models (Knowles and Koehn 2018) and they do not take the human guidance into account at all. Moreover, their learning objectives for both the confidence and translation models completely ignore human guidance. As a result, their performance could be limited, failing to catch the long-term influence of human involvement.

In this paper, we propose a novel interactive machine translation to jointly predict a word and a decision on when to request human guidance, which not only reduces human involvement but also generates excellent translations. We cast the query action as a special token in the output dictionary. At each time, the proposed NMT model predicts either a target word or the special token which indicates a request for the correct word from human. Furthermore, our NMT model is trained towards two objectives, i.e., improving translation quality and reducing human involvement required. Since there is no explicit signals on the decisions of requesting human guidance in the bilingual corpus, we employ the reinforcement learning (RL) technique to optimize the joint model which is able to predict when to request human guidance. Unfortunately, the reward includes two objectives which are somehow contradictory with each other: a higher translation quality often requires more human guidance, the standard RL learning algorithm (Bahdanau et

*This work was mainly done when Tianxiang Zhao was an intern at Tencent AI Lab. Lemao Liu is the corresponding author.

al. 2017) produced highly unstable results and was found not applicable in our preliminary experiments. To put RL into practice in our scenario, we additionally propose simple yet effective techniques into the standard RL training algorithm. Simulated and real experiments demonstrate that our proposed approach obtains higher BLEU while using less human involvement compared to the confidence based approach.

This paper makes the following contributions:

- It proposes a joint interactive NMT model for the paradigm which takes historical human involvement into account in architecture and is optimized to human involvement besides translation quality.
- It studies the behavior of reinforcement learning on the setting where its reward balances two contradictory objectives and highlights two important techniques in making RL successful.
- The proposed approach obtains better balance between translation quality and human involvement on both simulated and real experiments than the confidence based approach.

Preliminaries

NMT Model

We consider the problem of learning to generate the output sentence $\mathbf{Y} = \langle y_0, \cdots, y_{|\mathbf{Y}|-1} \rangle$ with length T being the target language based on an input $\mathbf{X} = \langle x_0, \cdots, x_{|\mathbf{X}|-1} \rangle$ in the source language. Typical methods are based on the encoder-decoder framework with attention mechanism. The encoder first maps the source sentence $\mathbf{X}$ into a set of representations by mixing the information in different tokens. Then, at each decoding step t, the next output token is predicted according to $P(y_t \mid \mathbf{Y}_{<t}, \mathbf{X})$. $\mathbf{Y}_{<t} = \langle y_0, \cdots, y_{t-1} \rangle$ is the hypothesis generated by the model, containing output tokens at each time-step, and $\mathcal{V}$ is the target-side vocabulary containing all candidate output tokens. Models are usually trained in a teacher-forcing manner, based on the maximum likelihood estimation (MLE) loss along the ground-truth reference.

Currently, the model showing the highest performance is Transformer, proposed by Vaswani et al. (2017) and thus we employ it as the testbed in our experiments. Different from previous models such as RNN-based (Cho et al. 2014) and Convolution-based (Gehring et al. 2017), it relies purely on the self-attention structure in both encoding and decoding process, which can mix the features in different time-step positions more efficiently. Our methods utilized this model as the backbone structure.

Confidence-based Baseline INMT

In this paper, we follow the conventional interactive machine translation to set the baseline which generates a translation from the the left-to-right (Foster, Isabelle, and Plamondon 1997). Conventional interactive machine translation requires human translators to validate all words in the prefix and then decide to correct a word accordingly. To further reduce human involvement, confidence based approaches have been proposed which focus on some of those words with low confidence and request human translator to type the correct words (Ueffing and Ney 2005; González-Rubio, Ortiz-Martínez, and Casacuberta 2010b; Cheng et al. 2016).

We develop our confidence baseline on top of the advanced Transformer. Similar to (Knowles and Koehn 2018), the generative story of our confidence based IMT can be described as follows. Suppose $\mathbf{Y}_{<t}$ has already been generated, and the next token is obtained through two steps:

1. Generate a token y_t by the NMT model;
2. Reset the token y_t via an oracle O (for example, a human translator) if the $P(y_t \mid \mathbf{Y}_{<t}, \mathbf{X}) \leq \eta$, where η is a predefined threshold to control the frequency of requesting human guidance.

Unfortunately, there are at least two pitfalls in the baseline which might lead to limited performance. On one hand, the NMT model is insensitive to the human guidance in the prefix $\mathbf{Y}_{<t}$, i.e. it completely ignores which tokens in $\mathbf{Y}_{<t}$ have been modified by the oracle O. On the other hand, the training objective of the model is the standard MLE, which does not take the human guidance into account.

Proposed INMT Methodology

In this section, we illustrate how we embed each query of human guidance as a special token in the output dictionary, and accordingly design our INMT model which takes historical queries of human guidance into account.

Joint INMT Model

To account human guidance in the architecture, we propose an improved INMT model which models it explicitly. The improved model is on top of the oracle O besides $\mathbf{X}$ and jointly predicts both translation and interaction. Specifically, we append a special token "$\langle$orc$\rangle$" to the target vocabulary: $\mathcal{V}' = \mathcal{V} \cup \{\langle \text{orc} \rangle\}$. The joint model parameterized by θ is defined as follows:

$$P(\mathbf{Y}' \mid \mathbf{X}, O; \theta) = \prod_t P\Big(y'_t \mid \mathbf{Y}'_{<t}, O(\mathbf{Y}'_{<t}), \mathbf{X}; \theta\Big) \quad (1)$$

where y'_t can be a real word in the vocabulary $\mathcal{V}$ or the special token "$\langle$orc$\rangle$", and $O(\mathbf{Y}'_{<t})$ denotes the token sequence which is obtained by the oracle after specifying all "$\langle$orc$\rangle$" in $\mathbf{Y}'_{<t}$ in the incremental manner to facilitate the left-to-right decoding strategy. For example, in Table 1, $\mathbf{Y}'_{<3}$ ="He is $\langle$orc$\rangle$" and $O(\mathbf{Y}')_{<3}$ =" He is my " . By introducing the special token, our model can jointly translate and predict when to query for human guidance: if the predicted y'_t is in $\mathcal{V}$, and our model does not to query the oracle, otherwise it needs the query from the oracle O.

To further define the network architecture of our model in Eq. (1), we mainly employ the same networks as in Transformer (Vaswani et al. 2017), to encode $\mathbf{X}$ into the representation vectors using self attention but with two significant differences. Firstly, since $\mathbf{Y}'_{<t}$ may include "$\langle$orc$\rangle$" for multiple times, it is inferior to predict y'_t by directly feeding $\mathbf{Y}'_{<t}$ into the model compared to standard transformer, and

Notation	Value
$\mathbf{Y}'$	He is ⟨orc⟩ colleague ⟨orc⟩ whom I traveled for ⟨orc⟩ three weeks .
$O(\mathbf{Y}')$	He is my colleague with whom I traveled for about three weeks .

Table 1: The examples. $\mathbf{Y}'$ denotes a translation from our joint model, in which "⟨orc⟩" indicating that this token should be corrected by an oracle O. $O(\mathbf{Y}')$ denotes the corrected translation by O.

thus we feed $O(\mathbf{Y}'_{<t})$ into the decoder instead. In addition, to take the human guidance in the history into account, we represent $\mathbf{Y}'_{<t}$ with a binary sequence to indicate whether $y'_i = \langle\text{orc}\rangle$ or not for all $i < t$, and then we employ the technique to encode this binary sequence similar to position embedding in (Gehring et al. 2017). For example, in Table 1, $\mathbf{Y}'_{<5}$ = 'He is ⟨orc⟩ colleague ⟨orc⟩", then $O(\mathbf{Y}'_{<4})$ ="He is my colleague with"0 0 1 0 1", both of which are actually encoded in our network to predict y'_5.

Proposed INMT Protocol

The generative story of our INMT for translation is achieved by the following two steps:

- generate a token y'_t from the joint model defined in Eq. (1).
- reset y'_t by requesting the oracle O if $y'_t = \langle\text{orc}\rangle$.

In our protocol, we predict when to request human involvement by adding a special token in the target dictionary, and thus enlarge the output dimensions of the actor by 1 accordingly. our protocol advantages can be summarized: 1) it only introduces three extra embedding parameters (one for the special tokens and two for binary code) compared to the standard transformer; 2) takes the cost of human involvement from the history into account.

Training via Reinforcement Learning

In the confidence based INMT baseline, its translation model is trained only for translation quality and thus is insensitive to the cost of human involvement. In this section, we thereby propose to optimize the joint INMT model towards both objectives (i.e., translation quality and human involvement). Since there is no explicit signals on the decisions of requesting human guidance in the bilingual corpus, we optimize the model with the actor-critic algorithm such that our model is able to predict when to request human guidance. To the best of our knowledge, it is the first time to train translation models towards such both goals for interactive machine translation.

However, it is far from trivial to train the INMT model via the actor-critic algorithm, because in our scenario the two goals are contradictory somehow: reducing the cost human involvement typically leads to worse translation quality. In the rest of this section, we will presents the components of our actor-critic algorithm and particularly we will analyze the issues suffered in our preliminary experiments in details and propose effective techniques to address them.

Pretraining

Suppose we are given a bilingual training corpus $\{\langle\mathbf{X}^n, \mathbf{Y}^n\rangle \mid n = 1, \cdots, N\}$, where $\mathbf{Y}^n$ is the reference of source sentence $\mathbf{X}^n$ and N is the size of the corpus. For each bilingual sentence $\langle\mathbf{X}^n, \mathbf{Y}^n\rangle$, there is no special token "⟨orc⟩" in the reference $\mathbf{Y}^n$, and thus we can not train the joint INMT to predict "⟨orc⟩" by using the bilingual corpus alone.

To address this, we propose to train it in a simple knowledge-transfer way. To this end, we firstly train a standard Transformer model on the given bilingual corpus, and then use the trained model to modify the reference sentence $\mathbf{Y}^n$ to $\mathbf{Y}'^n$ which includes "⟨orc⟩". Specifically, we use the trained model to rescore each token y^n_t in the reference $\mathbf{Y}^n$. If the model score of y^n_t is less than or equal to η, we set $y'^n_t = \langle\text{orc}\rangle$; otherwise $y'^n_t = y^n_t$. This procedure is similar to the baseline INMT protocol except that the former rescores the reference translation $\mathbf{Y}$ instead of a generated translation in the latter. Table 1 shows an example for $\langle\mathbf{Y}', \mathbf{Y}, \mathbf{A}\rangle$ according to the reference. In this way, we can convert the bilingual corpus including a set of $\{\langle\mathbf{X}^n, \mathbf{Y}^n\rangle \mid n = 1, \cdots, N\}$ into the corpus $\{\langle\mathbf{X}^n, \mathbf{Y}^n, \mathbf{Y}'^n\rangle \mid n = 1, \cdots, N\}$. Note that in the converted corpus we memorize $\mathbf{Y}^n$ such that it can recover "⟨orc⟩" in $\mathbf{Y}'^{,n}$ by looking up $\mathbf{Y}^n$ instead of query the oracle O.

Finally we pretrain our critic model over the converted corpus in a teacher-forcing manner by maximize the following objective:

$$\sum_t \log P(y'^n_t \mid \mathbf{Y}'^n_{<t}, \mathbf{Y}^n_{<t}, \mathbf{X}^n; \theta). \tag{2}$$

where the above model P is defined in Eq. (1) but it is conditioned on $\mathbf{Y}^n_{<t}$ rather than $\mathbf{Y}'^n_{<t}$ and thus it does not need the oracle during the pretraining. We expect this pretrained model to capture the intrinsic structure in deciding when the human-input is needed, which may provide a reasonable initialization for our RL algorithm.

Simulated Oracle and Reward

During the training, we have to sample a translation $\mathbf{Y}'$ for each source bilingual sentence $\langle\mathbf{X}^n, \mathbf{Y}^n\rangle$ following the generative story presented in our INMT protocol last section. Since it is too costly to employ a human translator as the oracle O, we instead provided a simulated oracle to mimic a human translator. Suppose at the timestep t, all tokens in $\mathbf{Y}'_{<t}$ are not equal to ⟨orc⟩ but $y'_t = \langle\text{orc}\rangle$, we simulate O to reset y'_t as a new token in $\mathcal{V}$ such that the following holds:

$$y'_t = \arg\max_{y \in \mathcal{V}} \text{pBLEU}(\mathbf{Y}'_{<t} \circ y)$$

where pBLEU denotes the partial BLEU score (Liu and Huang 2014), and $\mathbf{Y}_{<t} \circ y$ denotes the new prefix which extends $\mathbf{Y}_{<t}$ with a token y.

Traditionally, reward $\mathbf{R}$ is defined using BLEU score. However, since our goal is to maximize translation quality

while minimizing the number of human involvement, we introduce a penalty term to balance the accuracy and human involvement. In our scenario, we implement it in a most intuitive and simple way:

$$\mathbf{R}(\mathbf{Y}', \mathbf{Y}^n, O) = \text{BLEU}(O(\mathbf{Y}'), \mathbf{Y}^n) - \lambda \times \sum_t \delta\big(\langle \text{orc} \rangle, y'_t\big)$$

where $O(\mathbf{Y}')$ denotes the translation hypothesis where all "$\langle$orc$\rangle$" have been reset by the simulated oracle O, $\delta(y, y')$ returns 1 if $y = y'$ or 0 otherwise, and BLEU denotes sentence-wise BLEU+1 as in (Bahdanau et al. 2017), and λ is a hyperparameter to balance the both factors.

Critic Model and Its Updating

The reward can only be computed after the hypothesis sentence is finished, but if these rewards are delayed until the end, the algorithm will degrade, as reported in (Bahdanau et al. 2017; Nguyen, Daumé, and Boyd-Graber 2017). Therefore, a proper way to estimate the reward for intermediate steps is important for this method to coverage. As directly applying Monte Carlo search often yields high variance and results in instability during training when the search space is large, we use a critic model to approximate the average future reward.

Formally the critic model is defined by

$$V\big(\mathbf{Y}'_{<t}, O(\mathbf{Y}'_{<t}), \mathbf{Y^n}; \phi\big)$$

where $\mathbf{Y}'$ is a translation of the source sentence $\mathbf{X^n}$ and ϕ is the parameter of V. V is defined by the network which is almost the same as the joint model defined in Eq. 1 except two differences: its inputs include $\mathbf{Y}^n$ rather than the source sentence $\mathbf{X}^n$; in addition, its output is not a distribution over a vocabulary but a number which estimates the average future rewards at current time step t.

The critic model is trained to approximate the reward of the proposal with respect to ground-truth using minimal square loss as follows:

$$\sum_t G_t(\mathbf{Y}', \mathbf{Y}^n; \phi)^2 \tag{3}$$

where G_t is defined by:

$$G_t(\mathbf{Y}', \mathbf{Y}^n, O; \phi) = \mathbf{R}(\mathbf{Y}', \mathbf{Y}^n, O) - V\big(\mathbf{Y}'_{<t}, O(\mathbf{Y}'_{<t},), \mathbf{Y^n}; \phi\big).$$

Note that as critic model is only needed at the training phase, it will not cause a problem that the input contains ground-truth $\mathbf{Y}^n$ during testing phase.

Updating Actor Model

With critic model ϕ fixed, the actor model θ (i.e. our joint NMT model in Eq.(1)) can be updated following the standard advantage policy-gradient criteria (Sutton, Barto, and others 1998). However, since training INMT in our scenario is much difficult than training automatic NMT due to the two contradictory goals, we found that the standard actor-critic algorithm in (Bahdanau et al. 2017; Wu et al. 2018) fails in our preliminary experiments and we observed two severe issues leading to its failure.

The first issue is that our actor model suffers from so-called 'catastrophic forgetting': the change of the actor model after RL training begins is drastic, and consequently it soon loses the ability to give credential predictions in the following batches. This contributes to its ineffectiveness in exploring, and the actor will gradually degrade and get lower performance. To address this issue, we introduce two MLE auxiliary objectives and update actor model θ via the regularized policy gradient:

$$\nabla_\theta \Big[\log P\big({y'}_t \mid \mathbf{Y}'_{<t}, O(\mathbf{Y}'_{<t}), \mathbf{X}^n; \theta\big) G_t(\mathbf{Y}', \mathbf{Y}^n; \phi) + \lambda_1 \sum_t \log P(y_t^n \mid \mathbf{Y}'^n_{<t}, \mathbf{Y}^n_{<t}, \mathbf{X}^n; \theta) + \lambda_2 \sum_t \log P(y_t'^n \mid \mathbf{Y}'^n_{<t}, \mathbf{Y}^n_{<t}, \mathbf{X}^n; \theta) \Big]. \tag{4}$$

where P parameterized by θ is defined in Eq. (1), and $\langle \mathbf{X}^n, \mathbf{Y}^n, \mathbf{Y}'^n \rangle$ is the converted data from $\langle \mathbf{X}, \mathbf{Y} \rangle$ during pretraining. The loss corresponding to λ_1 can guide the actor to produce correct tokens and the loss The loss corresponding to λ_1 can guide the query for human's guidance at proper juncture respectively, making the updated θ more meaningful in exploration.

The second issue is that during training the actor model tends to change drastically after the RL training begins and thus the critic model V will soon be outdated, and incapable of giving accurate estimation of future rewards. To address this, we adopt a strategy of asynchronous updates between critic and actor models: we update critic model for three times while updating actor model once. This simple method shows a much more stable training trajectory. We attribute it to the reason that the value model can see more samples in the second way, and the variance introduced by sampling is alleviated.

The full training algorithm for our model is shown in Algorithm 1, in which the inputs are a threshold, two auxiliary weights and a bilingual dataset and the output is the actor model θ used for inference.

Experiment

We conduct both simulated experiments and real experiments. For the simulated experiments, we employ the bilingual datasets from IWSLT14 German-English (De-En), IWSLT14&15 Chinese-English (Zh-En), and IWSLT17 French-English (Fr-En); while for real experiments, we only use the De-En dataset. Note that in the simulated experiments, we use the simulated oracle with partial BLEU as described in previous section to mimic human translators while in the real experiments, we use the human translators as the oracle.

Setup

Data and Preprocessing For the IWSLT14 De-En dataset, we follow the procedures in (Ott et al. 2019) and use BPE (Sennrich, Haddow, and Birch 2016) to process the

Algorithm 1 RL for Proposed INMT Model

Require:
$\{\langle \mathbf{X}^{\mathrm{n}}, \mathbf{Y}^{\mathrm{n}} \rangle \mid n = 1, \cdots, N\}$.
Initialize the actor model θ by pretraining;
Set asynchronous update variable $\alpha = 0$;
while Not Converged **do**
  $\alpha = \alpha + 1$;
  Receive a random converted example $\langle \mathbf{X}^n, \mathbf{Y}^n, \mathbf{Y}'^n \rangle$;
  Sample $\mathbf{Y}'$ for $\mathbf{X}^n$ from the actor model θ;
  Update the critic model ϕ by one-step gradient over Eq. (3);
  if $\alpha \mod 3 == 0$ **then**
    Update the actor model θ by one-step gradient according to Eq. (4);
  end if
end while

source and target sentences. Data is tokenized and cleaned using Moses toolkit (Koehn et al. 2007). Similar steps are performed for IWSLT17 Fr-En. For the Zh-En dataset, we mainly adopt the same data-split method as Nguyen, Daumé, and Boyd-Graber (2017) and Li et al. (2019), except that we use both their 'Supervised training' and 'Bandit training' sets as the training set. We use the Stanford Chinese word segmenter (Chang, Galley, and Manning 2008) to segment Chinese sentences. Besides, BPE is also adopted for both source and target sentences.

Baseline Since our aim is to demonstrate that our joint model with RL training is able to obtain better translation with less human involvement than the confidence based approach, we implement both approaches on top of the same NMT framework and employ the left-to-right interactive decoding strategy. The confidence baseline INMT protocol is the one we proposed in Section 2, which is directly based on the standard transformer. In addition, to further show the effectiveness our RL training, we use the pretrained joint model as our second baseline, which learns when to interact in a knowledge-transfer way, as defined in Section 4.

Model Configuration Both the NMT model and the critic model are built upon 'transformer_iwslt_de_en' setting as defined in (Ott et al. 2019) for all three datasets, which is recommended for IWSLT datasets. Concretely, we use six self-attention layers for both encoding and decoding, and the embedding dimension is set to 512. For training our model, the BLEU score in our reward is the sentence-wise BLEU+1 which is a common practice when sentence-level BLEU is considered (Bahdanau et al. 2017). We train our models by the Adam Optimizer (Kingma and Ba 2015) with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and set the maximum tokens in a batch to 4000. For the pretraining process, we adopt a warm-up of 4000 steps and set the initial learning rate to 0.0003. Before the RL training begins, the actor model is initialized with parameters of the pretrained model, and the reward model is pretrained with learning rate 0.005 for one epoch. Then, we continue the training with learning rate 0.0001 for the actor and 0.0005 for the reward model. The scale λ of the query penalty in the reward function is set to 0.015(we found the improvement to be stable when it is between 0.015 and 0.02). And weight λ_1, λ_2 of two auxiliary losses are initialized to 0.1 at the beginning and lowered by approximately 0.6 after each epoch.

Evaluation The comparison between our approach and baseline approaches is not straightforward, as it is in essence a multi-goal task and particularly our two goals are contradictory. A quick idea is to take several different interpolation coefficients and evaluate against the interpolated goal, but this idea is not applicable because our approach can directly optimize towards it by casting this interpolation coefficient as λ in the reward, making the comparison unfair. Inspired by Duh et al. (2012), we employ the following criterion to conclude that one approach A is better than the approach B if and only if the approach A achieves higher BLEU while using less human involvement than the approach B. Since the baselines and our approach adopt the different ways to control the level of human involvement, it is not easy to maintain the similar human involvement for three different approaches. Therefore, in the simulated experiments, we independently running all three approaches with different hyperparameters and draw a piece-wise linear curve similar to ROC curves (Bradley 1997) for each approach in terms of both goals, then we validate whether the criterion is satisfied by picking points from the curve [1]. To reduce the cost in the real experiments, we pick one setting for the baseline and one setting for our approach from the simulated experiments, and then validate whether the criterion is satisfied during the real human-computer interaction.

Results and Analysis

Simulated Scenario The performance of each approach on three datasets are shown in Figure 1. It can be seen that in IWSLT14 De-En, our RL-based method shows an improvement of average 2 BLEU points over the baselines when using the same amount of human guidance. The improvement is more significant when query frequency is below 18 percent, with a gap of over 3 points. When the human effort is involved in over 22 percent of tokens, the gap drops a little, becoming 1.5. This result shows that despite the baseline approaches are already 11.5 BLEU points better than the traditional transformer when having an average guidance frequency of 0.16, their interaction policies are still sub-optimal. Our approach, which uses RL to directly trade off between human effort and translation quality, can outperform them with a large margin. Similar results can be found in Zh-En and Fr-En datasets, with an average improvement of 1.1 and 1.7 BLEU scores under the same query frequency respectively, as also shown in Figure 1.

Figure 1 has shown the gains in directly learning when to query for human guidance. To better understand the learning process, we also provide an example of the curves over

[1]For both baselines, the variant hyperparameter was η taken from $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$. For our approach, we initialized it with different pretrained models.

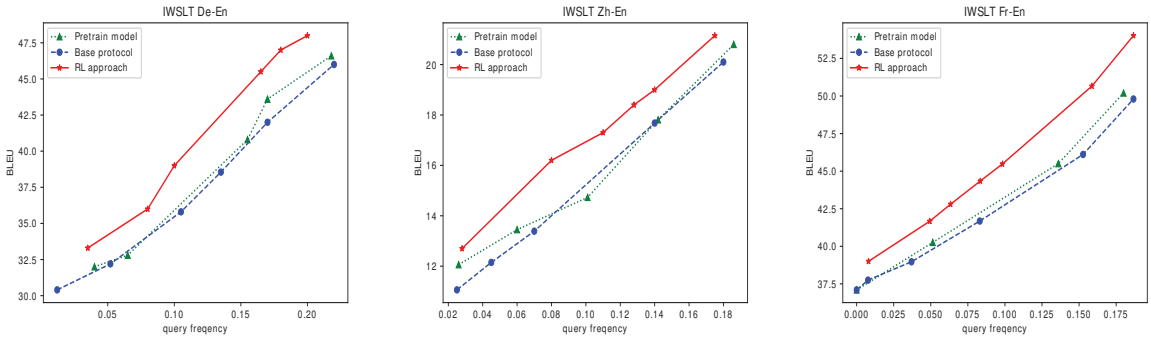

Figure 1: Comparison of different approaches. The x-axis refers to the frequency of making queries, with 1 meaning guidance is required at every time-step. The y-axis is the BLEU score.

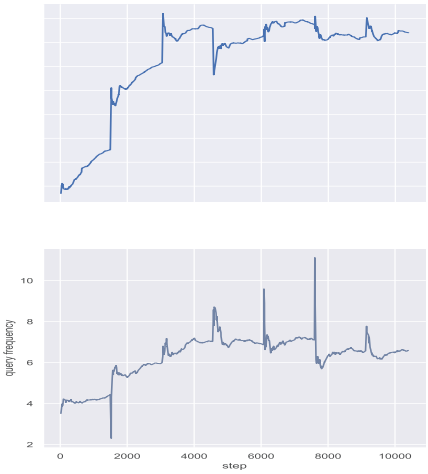

Figure 2: Progresses of the BLEU scores (upper) and the query times (lower) during training.

BLEU scores and query times during training in Figure 2. We can see clearly that the change of model's behavior in the first several epochs is drastic. The model tends to optimize towards relying more on human guidance in the beginning, and the translation quality soars rapidly. After arriving at certain point, it will stop this tendency, and seek to reduce the human efforts while keeping a high translation quality.

methods	BLEU	query freq
NMT	39.48	0
Base protocol	46.06	0.10
RL approach	49.33	0.10

Table 2: Comparison of different approaches in real IMT scenario, where human translators serve as the oracle.

Real Scenario To validate whether our RL-based method can really reduce the amount of human involvement, we also perform an experiment to test its usage in the real human-computer interaction scenario. Since it is costly to conduct IMT experiments for all test sentences on the real scenario, we sample 400 samples from IWSLT14 De-En and then ask two human translators to conduct the human-computer translation. Each human translator has to interact with two INMT systems [2] (i.e. confidence-based and our RL-based) to translate one source sentence twice. To make the comparison fairer, when translating one source sentence we supply both systems to each translator in a random order such that the translator does not perceive which system it is. We record the BLEU scores of translation and query frequencies for each human translator, and find that the result from one translator correlate well with that from the other. Then we average both the BLEU scores and query frequencies as our final result, which is shown in Table 2. It could be seen that our RL approach can get higher BLEU score with similar human involvement.

Analysis on Queried Tokens

We try to look into what sorts of tokens are queried more frequently. We split all words into two sets, frequent tokens and infrequent tokens, and show the policy over several samples in Table 3 and 4. Here, we use recall to measure the ability of each model to successfully predict the ground-truth token, i.e. the possibility that this token is generated by a translation model during decoding. Here, TFM refers to standard transformer model without human intervention.

First, it can be seen that the policy is closely related to the effectiveness of the translation model. For tokens that a traditional transformer can translate well, the query frequency is low. While for those which has a low recall by TFM, the query frequency becomes high. This further illustrates the efficiency of our interaction approach.

Second, it is shown in Table 3 that the model is more uncertain over its output when the ground-truth of that token does not have a specific meaning. It is not because they are under-represented in the dataset, but because they can be used in many conditions and have complex usages. On

[2]Since the pretrained model is comparable to the confidence-based baseline, and thus we only compare with the latter in the real scenario.

Token	thank	but	ago	tell	said	up	down	out	@@	ing
Query freq	0.006	0.046	0.054	0.057	0.060	0.437	0.444	0.468	0.468	0.469
recall by TFM	0.986	0.912	0.907	0.697	0.817	0.237	0.190	0.254	0.155	0.220
recall by Ours	0.986	0.946	0.898	0.786	0.872	0.502	0.479	0.436	0.409	0.485

Table 3: Examples of the interaction policy over frequent tokens.

Token	gulf	mechanic	empathy	constant	reward	eland	tives	ayer	escent	tish
Query freq	0	0	0	0	0	1	1	1	1	1
recall by TFM	1	0.67	1	0	0.67	0	0	0	0	0
recall by Ours	1	0.67	1	0	0.67	0.5	0.5	0.5	1	1

Table 4: Examples of the interaction policy over infrequent tokens.

the contrary, for tokens have less ambiguity, the model will safely believe its prediction. For infrequent tokens, as shown in table 4, similar behavior can be observed. The model become confused and does not know what to predict when coming across meaningless tokens generated by BPE. For tokens with a clearer meaning, although they are also rare during training, the model can successfully predict them by itself.

Related Work

Interactive-predictive MT reaches back to early IBM-type (Foster, Isabelle, and Plamondon 1997) and phrase-based MT (Barrachina et al. 2009; Green et al. 2014). These methods seek to reduce human's effort in correcting model's translation by making human and machine collaborate on a joint iterative strategy. Advances have been introduced by utilizing a richer variety forms of feed-backs, like using judgements on the quality of partial translation results Lam, Kreutzer, and Riezler (2018), using guidance on the segments instead of on the prefix Peris, Domingo, and Casacuberta (2017), using human's correction of the left-most wrong word Azadi and Khadivi (2015), etc. Another track of works dedicates to exploit more information from the feed-back. For example, Koehn (2009) propose to suggest more than one suffix for users to validate, Peris and Casacuberta (2018b) adopts online learning techniques to improve the system with the user feedback, and Peris and Casacuberta (2018a) uses active learning to choose the sentences that can gain more knowledge from users. However, few of these works address the problem of 'when to query'.

In active learning domain, there are a few works seeking to evaluate the model's uncertainty towards its translation result. For example, González-Rubio, Ortiz-Martínez, and Casacuberta (2012) uses the distribution of confidence score, and Peris and Casacuberta (2018a) propose to use the attention coverage and distraction. However, their methods are all based on heuristic criterion, instead of directly balancing the gain and penalty in obtaining human's guidance.

Ibraheem, Altieri, and DeNero (2017) is close to our work in the sense that they also employ reinforcement learning to learn an interaction policy. However, they fix the translation model and only model the human actions as a binary variable over the attention vector.Therefore, their training process is more stable, but also resulting in the gains of only about 12 BLEU points over the standard TFM with 80% query frequency. On the other hand, we train both the translation and human action in a unified model and design a more complex guidance form, leading to a large action space and making our approach more difficult than theirs. But, the benefit of our work is that it can obtain more BLEU improvement with much less human involvement (with query frequency about 20%).

Conclusion and Future Work

Confidence-based interactive machine translation is effective to reduce the human involvement because it only requires human translators to correct few of those words with low confidence while avoiding the validation for other words. In this paper, we propose a novel approach to IMT which does not require human translators to validate all output words but only focus on some words. Our approach relies on a novel neural model which considers human involvement in its architecture and is optimized towards both translation quality and the cost of human involvement via reinforcement learning. Simulated and real experiments show that the proposed approach outperforms the confidence baseline with a margin in translation quality by using similar or less human involvement. Since the training efficiency is relatively low compared to that of the standard NMT model, in the future, we plan to accelerate our approach and then apply it to large scale translation tasks.

Acknowledgments

We would like to thank all the anonymous reviewers for their valuable suggestions. T. Zhao, Y. Liu and G. Liu were supported in part by the Anhui Sun Create Electronics Company Ltd., under Grant KD1809300321, the National Key R&D Program of China under Grant 2018YFC0832101, the National Key New Product Plan of China under Grant 2014GRC30006, and STCSM18DZ2270700.

References

Azadi, F., and Khadivi, S. 2015. Improved search strategy for interactive predictions in computer-assisted translation. In *Proceedings of MT Summit*, 319–332.

Bahdanau, D.; Brakel, P.; Xu, K.; Goyal, A.; Lowe, R.; Pineau, J.; Courville, A. C.; and Bengio, Y. 2017. An actor-critic algorithm for sequence prediction. *CoRR* abs/1607.07086.

Bahdanau, D.; Cho, K.; and Bengio, Y. 2015. Neural machine translation by jointly learning to align and translate. *CoRR* abs/1409.0473.

Barrachina, S.; Bender, O.; Casacuberta, F.; Civera, J.; Cubel, E.; Khadivi, S.; Lagarda, A. L.; Ney, H.; Tomás, J.; Vidal, E.; and Vilar, J. M. 2009. Statistical approaches to computer-assisted translation. *Computational Linguistics* 35:3–28.

Blatz, J.; Fitzgerald, E.; Foster, G. F.; Gandrabur, S.; Goutte, C.; Kulesza, A.; Sanchís, A.; and Ueffing, N. 2004. Confidence estimation for machine translation. In *Proceedings of COLING*.

Bradley, A. P. 1997. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30:1145–1159.

Chang, P.-C.; Galley, M.; and Manning, C. D. 2008. Optimizing chinese word segmentation for machine translation performance. In *WMT@ACL*.

Cheng, S.; Huang, S.; Chen, H.; Dai, X.-Y.; and Chen, J. 2016. Primt: A pick-revise framework for interactive machine translation. In *Proceedings of NAACL-HLT*, 1240–1249.

Cho, K.; van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *EMNLP*.

Duh, K.; Sudoh, K.; Wu, X.; Tsukada, H.; and Nagata, M. 2012. Learning to translate with multiple objectives. In *ACL*.

Foster, G. F.; Isabelle, P.; and Plamondon, P. 1997. Target-text mediated interactive machine translation. *Machine Translation* 12:175–194.

Gehring, J.; Auli, M.; Grangier, D.; Yarats, D.; and Dauphin, Y. 2017. Convolutional sequence to sequence learning. In *ICML*.

González-Rubio, J.; Ortiz-Martínez, D.; and Casacuberta, F. 2010a. Balancing user effort and translation error in interactive machine translation via confidence measures. In *Proceedings of the ACL*, 173–177.

González-Rubio, J.; Ortiz-Martínez, D.; and Casacuberta, F. 2010b. On the use of confidence measures within an interactive-predictive machine translation system. In *Proc. EAMT*.

González-Rubio, J.; Ortiz-Martínez, D.; and Casacuberta, F. 2012. Active learning for interactive machine translation. In *EACL*.

Green, S.; Wang, S. I.; Chuang, J.; Heer, J.; Schuster, S.; and Manning, C. D. 2014. Human effort and machine learnability in computer aided translation. In *EMNLP*.

Ibraheem, S.; Altieri, N. D.; and DeNero, J. 2017. Learning an interactive attention policy for neural machine translation. In *Proceedings of MT Summit*.

Kingma, D. P., and Ba, J. 2015. Adam: A method for stochastic optimization. *CoRR* abs/1412.6980.

Knowles, R., and Koehn, P. 2016. Neural interactive translation prediction. In *Proceedings of AMTA*, 107–120.

Knowles, R., and Koehn, P. 2018. Lightweight word-level confidence estimation for neural interactive translation prediction. In *Proceedings of AMTA 2018 Workshop*, 35–40.

Koehn, P.; Hoang, H.; Birch, A.; Callison-Burch, C.; Federico, M.; Bertoldi, N.; Cowan, B.; Shen, W.; Moran, C.; Zens, R.; Dyer, C.; Bojar, O.; Constantin, A.; and Herbst, E. 2007. Moses: Open source toolkit for statistical machine translation. In *ACL*.

Koehn, P. 2009. A process study of computer-aided translation. *Machine Translation* 23:241–263.

Lam, T. K.; Kreutzer, J.; and Riezler, S. 2018. A reinforcement learning approach to interactive-predictive neural machine translation. *CoRR* abs/1805.01553.

Langlais, P.; Foster, G.; and Lapalme, G. 2000. Transtype: a computer-aided translation typing system. In *ANLP-NAACL 2000 Workshop: Embedded Machine Translation Systems*.

Li, G.; Liu, L.; Huang, G.; Zhu, C.; and Zhao, T. 2019. Understanding data augmentation in neural machine translation: Two perspectives towards generalization. In *Proceedings of EMNLP-IJCNLP*.

Liu, L., and Huang, L. 2014. Search-aware tuning for machine translation. In *Proceedings of EMNLP*, 1942–1952.

Nguyen, K.; Daumé, H.; and Boyd-Graber, J. L. 2017. Reinforcement learning for bandit neural machine translation with simulated human feedback. In *EMNLP*.

Ott, M.; Edunov, S.; Baevski, A.; Fan, A.; Gross, S.; Ng, N.; Grangier, D.; and Auli, M. 2019. fairseq: A fast, extensible toolkit for sequence modeling. In *Proceedings of NAACL-HLT 2019: Demonstrations*.

Peris, Á., and Casacuberta, F. 2018a. Active learning for interactive neural machine translation of data streams. In *CoNLL*.

Peris, Á., and Casacuberta, F. 2018b. Online learning for effort reduction in interactive neural machine translation. *CoRR* abs/1802.03594.

Peris, Á.; Domingo, M.; and Casacuberta, F. 2017. Interactive neural machine translation. *Computer Speech & Language* 45:201–220.

Sennrich, R.; Haddow, B.; and Birch, A. 2016. Neural machine translation of rare words with subword units. *CoRR* abs/1508.07909.

Sutton, R. S.; Barto, A. G.; et al. 1998. *Introduction to reinforcement learning*, volume 135. MIT press Cambridge.

Ueffing, N., and Ney, H. 2005. Application of word-level confidence measures in interactive statistical machine translation. In *Proceedings of EAMT*, 262–270.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is all you need. In *NIPS*.

Wu, L.; Tian, F.; Qin, T.; Lai, J.; and Liu, T.-Y. 2018. A study of reinforcement learning for neural machine translation. In *EMNLP*.

Wuebker, J.; Green, S.; DeNero, J.; Hasan, S.; and Luong, M.-T. 2016. Models and inference for prefix-constrained machine translation. In *Proceedings of ACL*, 66–75.