Research on Collaborative Machine English Translation Using the HIC Technology

Jingjing Lv, Hunan International Business Vocational College, China*

ABSTRACT

Due to the rapid development of smart city, Hybrid Information Centric-Networking (HICN) emerges as a promising technology to enable the power of smart city. One of the most important application is the smart English translation, which becomes more and more popular with the process of internationalization. In this work, the authors focus on studying the intelligent English translation in smart city using the HICN technology. Particularly, a method using collaborative machine learning and quality estimation technique is proposed, which sets a fixed threshold to filter pseudo-parallel data during unsupervised neural machine translation training. The quality estimation is used to evaluate and screen the pseudo-parallel data with high performance generated during reverse translation training. The results indicate that the proposed method outperforms the state-of-the-art methods.

KEYWORDS

HICN, Machine English Translation, Quality Estimation, Smart City

1. INTRODUCTION

The Information Centric-Networking (ICN) (Ion et al. 2013) is a novel and promising network architecture, which focuses on information and pays attention to the content itself. Instead, the location of the content storage is not the target. As the unique identification of the content, the content name is used for the positioning, routing and transmission of the content. Intra-network cache is one of the most important features of ICN network architecture (Kutscher et al. 2016). If the information requested by the user is matched in the router, the router sends the information directly to the user. Instead of parsing the address of the source host as in the traditional IP network, the source host must forward the message to the user, which greatly reduces the transmission delay and network traffic and improves the network performance (Eum et al. 2018). As you can see from figure 1, the biggest difference between ICN and the traditional TCP/IP model is that IP is replaced by Content Chunk at the "thin waist", where the every node in the middle means the router or the switcher used to forward traffic from source to destination. From a network point of view, the naming of content is used instead of the naming of physical entities.In addition, the build-in storage function in the network is used to cache the passing data, to shorten the response time for other users to access the same data, and to greatly reduce the traffic in the network (Li et al. 2018).

ICN has been recognized by scientific research institutes and even some companies around the world (such as Cisco and Huawei) because of its unique characteristics of supporting on-net cache and IP address decoupling (Trossen et al. 2015). In order to speed up the process of ICN, many new networking modes, including data center network, software-defined network, mobile social network, satellite communication network and vehicle network, as well as cloud / fog / edge computing (Yi et

*Corresponding Author

Figure 1. Comparison of protocol structures between traditional IP and ICN
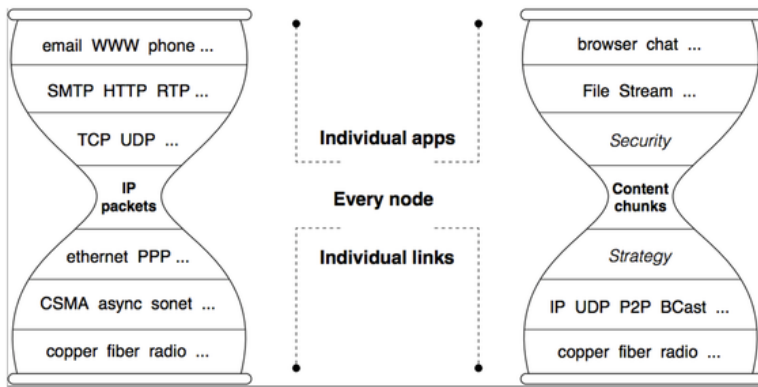
al. 2021), 6G/5G, big data and network function virtualization are introduced to enhance ICN. This important integration with ICN is called the Hybrid ICN (HICN) (Muscariello et al. 2018). Although HICN has improved ICN to a great extent, it also faces some challenges that need to be solved urgently.

Now, smart city (Su et al. 2017), as an important application scenario of HICN, presents an important trend in the process of continuous research and development. It makes full use of the new generation of information technology in various industries in the city to create an advanced form of urban informatization based on the knowledge society (Neirotti et al. 2014). At the same time, with the development and acceleration of the process of internationalization, the popularization of English has become inevitable. Then, the combination of the two is the trend of the times, and it has become an important application of smart cities to accelerate and optimize English translation based on new information technology. Therefore, the intelligent English translation technology based on HICN has become the current research hotspot (Naeem et al. 2018).

Recently, the method of Neural Machine Translation (NMT) (Wu et al. 2016) stands out in the field of machine translation and has made great progress. NMT is usually composed of two sub-neural networks. The encoder network encodes the source language sentences into context vectors, and then the decoder network iteratively decodes the context vectors encoded by the encoder network into target language sentences with the same meaning (Sennrich et al. 2015). Usually in a supervised environment, a large number of bilingual parallel sentence pairs are needed to train the model. However, most language pairs have almost no parallel data. The Unsupervised Neural Machine Translation (UNMT) successfully breaks this limitation, and conventional machine translation tasks can be accomplished only by training with monolingual corpus in two languages. Unsupervised neural machine translation is based on neural machine translation, and combines denoising automatic encoder and reverse translation (Shakhnovich et al. 2018) training dual model initialization and cross-language embedding (Asghari et al. 2019) to achieve the goal of machine translation. However, in the process of unsupervised neural machine translation model training, reverse translation will produce a large number of pseudo-parallel data, and the quality of these pseudo-parallel data becomes very important in reverse translation training. Therefore, controlling the quality of pseudo-parallel data in the process of training reverse translation is a key to improve the quality of unsupervised neural machine translation (Ren et al. 2019).

In this paper, a method using collaborative machine learning and quality estimation technique is proposed, which sets a fixed threshold to filter pseudo-parallel data during unsupervised neural

machine translation training, which effectively improves the effect of unsupervised neural machine translation. Then, based on the generation countermeasure network (Fotohi et al. 2020), reverse translation and language modeling are used as generators and discriminators respectively. The quality estimation is used to evaluate and screen the pseudo-parallel data with high HTER generated during reverse translation training. At the same time of controlling the quality of pseudo-parallel data, this method enriches the generating sample of the generator, makes the discriminator collect more slowly, and makes the generation countermeasure network trained more fully, thus improving the translation effect of unsupervised neural machine translation.

The organizational structure of this paper is as follows: the second section introduces the related research work. The third section introduces the proposed model and method. The fourth section analyzes the experimental results. The fifth section summarizes this work.

2. RELATED WORK

At present, several methods have been proposed to train NMT models without using bilingual parallel corpus directly. The method (Shatsky et al. 2018) used by is to train independent translation from the source language to the pivot language, and then from the pivot language to the target language. Although this method is ingenious, it is highly dependent on the pivot language. at the same time, it invisibly introduces the translation error of the third language or even the fourth language, and the error introduced by the pivot language can not be eliminated (Meyer et al. 2015).

Due to the success of cross-language embedding, Lample et al and Artetxe et al also proposed an unsupervised neural machine translation method based on pre-training cross-language embedding training. This method only uses monolingual corpus of two languages to train the embedding of two languages independently, and learns linear transformation (Abdel et al. 2016) and confrontation training (Zhang et al. 2018) to map them to shared space. The resulting cross-language embedding is used to initialize the shared encoder of the two languages, and the whole system uses a combination of denoising automatic encoder, reverse translation and confrontation training (Aiguier et al. 2015). Yang et al. further improved this method by using two language-specific encoders (Sen et al. 2019), sharing only a subset of their parameters, and combining local and global generation to train unsupervised neural machine translation.However, when training the model, these methods do not take into account the quality of pseudo-parallel data generated by reverse translation, resulting in insufficient training of the model (Rudebeck et al. 2019).

At present, reverse translation technology is often used in unsupervised neural machine translation. Wang et al. used the method of quality estimation (Martins et al. 2017), that is, using OpenKiwi (Kepler et al. 2019) to screen the pseudo-parallel data generated by reverse translation and the square method of combining the real parallel data with the real parallel data to train NMT, which improved the performance of machine translation. Li et al. used reverse translation to expand parallel data using monolingual corpus (Bustamante et al. 2020), and adopted a new data enhancement method to further improve the robustness of NMT to noise and translation effect while keeping the model small. Miguel et al. adopted the scheme of synthetic data generation and reconstructed the reverse translation within the cross entropy optimization norm of the NMT module (Sennrich et al. 2017), elucidating its basic mathematical hypothesis and approximate values other than heuristic usage. The basic problems of the sampling-based method are pointed out by using the formula, and the problem of neural machine translation data enhancement is solved by disabling the label smoothing from the target to the source model and sampling from the restricted search space (Janson et al. 2016). However, Wang et al. and Miguel et al. all use supervised methods to solve the problem of insufficient data and control the quality of pseudo-parallel data, and do not solve the problem of training machine translation tasks without bilingual parallel corpus.

The method in this paper not only relies on using the collaborative machine learning, but also based on using the unsupervised NMT method proposed by Gu et al. 2017. It only uses two kinds of

monolingual corpus to screen pseudo-flat data to train reverse translation, while using less data to train and improve the performance of unsupervised neural machine translation.

3. PROPOSED ALGORITHM

In order to fulfill a better English translation framework, we first establish the whole system structure for English translation, as shown in Figure 2. In particular, there are three main modules which are the feature extraction module, the feature integration module and the machine translation module. In particular, the first module is in charge of extracting different features, and the second module is used to integrate different features. Based on these information, the last module can carry out the translation process easily.

Figure 2. System framework and structure
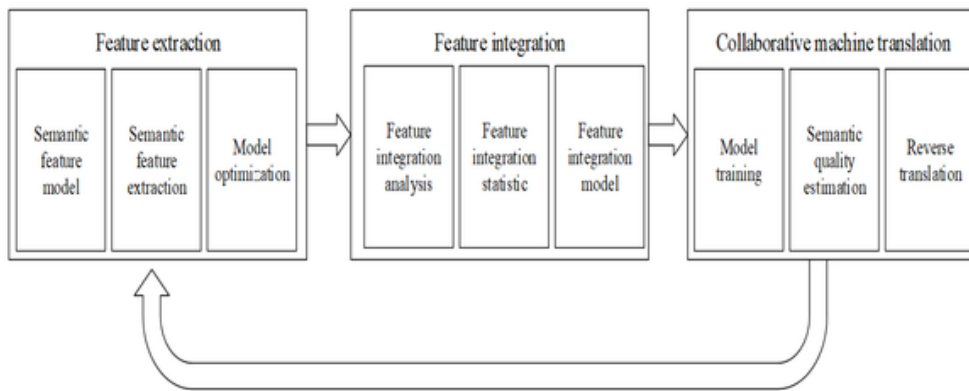
3.1 Feature Extraction

1) Semantic Feature Model

On the basis of the above construction of the overall structure model of English-Chinese machine automatic conversion translation, the semantic features of English-Chinese machine automatic conversion translation are analyzed, and the semantic relevance mapping and association rules reorganization are adopted. The expression of semantic relevance degree model of English-Chinese machine automatic translation is as follows:

$$A = P(\frac{a}{2} + b)^2 \tag{1}$$

where a is the balance factor of the semantic feature distribution of English machine translation; b is the adaptive analytical factor of English machine translation, and P is the entity probability density distribution set of English machine automatic translation control.

Now, we need to establish a fuzzy information fusion model, and the joint probability density space of n common notional words is obtained. Through the automatic transformation of the mapping results of semantic information in translation by the English translation machines, using linear weighting, the weighted control quantity is obtained as follows:

$$B = \frac{(a+b) - A\sqrt{k}}{(a-b)^2} \tag{2}$$

where k is the keyword weight of the translated text. The feature extraction of English machine automatic transformation translation is carried out by using semantic mapping and relevance feature reorganization. According to the feature extraction results of English machine automatic transformation translation, the joint parameter distribution set of semantic mapping is obtained. Through fuzzy measure decomposition, the mapping expression of joint feature distribution is obtained as follows:

$$C = \frac{\log_k (B+k) + 1}{\log_k B} \tag{3}$$

Combined with the fuzzy decision of English-Chinese machine automatic translation, the similarity fusion component of each subset translation is obtained, that is, $\left\{ e_1, e_2, \cdots, e_i, \cdots \right\}, \forall i \in \left[1, r \right]$ where i means the i-th word of the text to be translated and r is the total length of the text. By using subspace sequence fusion, the template learning function of English-Chinese machine automatic conversion translation is:

$$D = QN + \frac{C}{u} \tag{4}$$

Among them, Q is the semantic relevance distribution set of English machine automatic translation; N is the joint feature parameter distribution set, and u is the self-similar fusion distribution coefficient of English machine automatic translation. The fuzzy analytic feature clustering matrix is as follows:

$$E = \begin{bmatrix} \omega^0 & \omega^0 & \omega^0 & \cdots & \omega^0 \\ \omega^0 & \omega^1 & \omega^2 & \cdots & \omega^{k-1} \\ \omega^0 & \omega^2 & \omega^4 & \cdots & \omega^{2(k-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega^0 & \omega^{k-1} & \omega^{2(k-1)} & \cdots & \omega^{(k-1)^2} \end{bmatrix} \tag{5}$$

where ω represents the feature similarity corresponding to each word. Combined with feature distributed fusion and joint parameter analysis, a joint feature analysis model of English machine automatic conversion translation is constructed to improve the output adaptability and feature detection and analysis ability of English machine automatic conversion translation.

2) Semantic Feature Extraction

The ontology fusion method of semantic unit is used to construct the extraction of semantic context in English-Chinese machine translation.By using the method of semantic modification, the joint feature distribution set An of automatic machine conversion between English and Chinese is constructed, and the piecewise function of automatic machine conversion between English and Chinese is constructed by combining the fuzzy comprehensive decision method. The semantic structure distribution of the translation output of automatic machine conversion between English and Chinese is obtained:

$$\delta\left(t_1, t_2\right) = E^T + \frac{2\left(t_1 + t_2\right)}{D} \tag{6}$$

Among them, the t_1 is the output time series of English machine automatic conversion translation; t_2 is the sampling time interval of English machine automatic conversion translation, and the joint features of English machine automatic conversion translation are:

$$F = \delta\left(t_1, t_2\right) + uC \tag{7}$$

By using the method of semantic probability set fusion, the semantic feature distribution nodes of English-Chinese machine translation are obtained. Denoting them by l1 and l2 respectively, then the distance between them is defined as the Euclidean distance, based on which we create the joint feature components of English-Chinese Machine automatic conversion translation that is denoted by Δx . In the interval of $\left[-\dfrac{\Delta x}{2}, \dfrac{\Delta x}{2}\right]$, the constraint optimization problem of English-Chinese machine automatic conversion translation is constructed, and the joint constraint features of English-Chinese machine automatic conversion translation are obtained, as follows.

$$G = \begin{cases} 0, \left| F - \dfrac{l_1 + l_2}{2} \right| \le \Delta x \\ 1, \left| F - \dfrac{l_1 + l_2}{2} \right| > \Delta x \end{cases} \tag{8}$$

The vector model structure of the semantic unit of the English-Chinese translation model is constructed, and the semantic feature extraction of English-Chinese machine translation is carried out.

3)Model optimization

A remote fuzzy control model of English-Chinese machine automatic conversion translation is constructed, and linear statistical features and feature analysis are used to detect. The statistical functions of quality evaluation of English-Chinese machine automatic conversion translation are obtained as follows:

$$H = (\sqrt{G} + v)^2 + \beta F \tag{9}$$

where v is expressed as the set of constraint index parameters of English-Chinese machine automatic conversion translation, which is a standard normal distribution function, and β is the output stable characteristic quantity of English-Chinese machine automatic conversion translation. By using the

method of relevance mapping analysis, the mapping of the statistical feature sets R and S of English-Chinese machine translation are carried out, and the optimal solution of English-Chinese machine conversion translation text is obtained.:

$$K = \frac{H\sqrt{\beta}}{R} \times \frac{Hv}{S} \qquad (10)$$

The semantic fusion model of English-Chinese machine automatic translation is established, and the spectral information clustering and attribute judgment are used to realize the automatic English-Chinese machine translation.

3.2 Feature Integration

The feature integration is carried out in three steps which are the feature analysis, statistics, and the integration model building, . The details are as follows.

1) Feature Integration Analysis

Syntactic feature method, also known as dependency parsing or dependency grammar, focuses on the association of words in long sentences. In English grammar, the common relations of sentence components are subject-predicate relationship, verb-object relationship, juxtaposition relationship and so on. In a sentence, the verb is generally regarded as the core word in the sentence, and the other words in the sentence are directly or indirectly related to the core word. For the syntactic feature analysis of a sentence, the analysis process is usually represented by a directed graph. The words in the sentence are individual nodes, and the syntactic relationship between the core word and the dependent word is marked with a directed arrow, and the relationship is explained above the arrow.

In the syntactic feature model proposed in this paper, it is stored according to the unit, and the syntactic feature unit model will be established below. In this model, the syntactic relations between words are directly stored in the corresponding syntactic units, and the unit construction rules are as follows:

$$M_i = \left\{ Mx_i, MPx_i, MCx_i, MBx_i \right\} \qquad (11)$$

where M_i represents the storage syntactic unit of the first word in a sentence. MPx_i, MCx_i, and MBx_i represent the position of the first parent node word, child node word and adjacent node word in the sentence.

Although the syntactic feature model can judge the sentence components, the translation model can also achieve a better accuracy.However, due to the limitations of the model itself, the model can not better learn the weak features of sentences. Therefore, it is still necessary to add a statistical model on the basis of syntactic features to strengthen the semantic relationship of each word in a long sentence, so as to further generate correct results.

2) Feature Integration Statistic

In the process of cutting long sentences, not any long sentence can be divided into appropriate short sentences. Only when the segmented sentence has an independent syntactic feature structure can it be considered meaningful for the segmentation of long sentences. Therefore, the conditional random field model is introduced to make a meaningful segmentation of words and commas in long sentences. In the conditional random field model, all corpus should be properly trained to judge whether the sentence segmentation in the corpus set is reasonable or not.

The subordinate classification of conditional random field is undirected graph model, which has the characteristics of maximum entropy and hidden Markov chain. This statistical model is widely used in the field of natural language processing. The conditional random field can be defined as a conditional probability event, X is used to represent the observation sequence set condition, and Y is used to mark the sequence set condition, then the conditional random field model can be represented by conditional probability P(Y|X). The conditional random field is defined according to the statistical mathematical model.

Definition 1: Suppose that an undirected graph is $T(V, E)$, where V is the set of vertices and E is the set of edges. Suppose that $Y=\{Y_v \mid v \hat{I} V\}$, that is, each individual element in the vertex set has a variable Y_v. Let X satisfy the condition of Y_v, then the variable Y_v can satisfy the following formula:

$$p(Y_v \mid X, Y_u, u \neq v) = p(Y_v \mid X, Y_u, u : v) \tag{12}$$

where u, v indicates any two nodes in the graph T and the pair (X, Y) is a strip random field.

3) Feature Integration Model

The model based on syntactic features can not learn the weak features of sentences. Therefore, a translation model based on syntactic features is established by combining the translation method based on language template and the translation method based on statistics. This model can strengthen the semantic relationship of each word in long sentences and further improve the translation quality of long sentences. The combination of the two simple models uses parallel execution, that is, sentence analysis based on syntactic features; sentence analysis based on conditional random field, and then two ways of long sentence segmentation are obtained. Then the two methods are fused, and the fusion methods include merging, de-duplication and so on. Finally, the translation is carried out in the translation engine.

Before using the fusion syntactic feature model for sentence segmentation, we should first train the conditional random field model according to definition 1. The process of model training is as follows:

- The corpus set merging is selected for pre-processing, and the sentences in the corpus are pre-processed, including the removal of repetitive sentences, the removal of sentence special symbols and so on.
- Extract the features of the sentences in the corpus set, then use the syntactic feature filter to analyze the components of the sentences and establish the dependency digraph, and then complete the sentence feature extraction.
- Input the sentence features into the conditional random field model for training. After the model training is completed, the training results of the conditional random field in the pre-processing module will be input to the conditional random field decoder in the segmentation process for decoding. At the same time, the sentences processed by syntactic features are compared with the sentences processed by syntactic features, and the operations such as merging and de-duplication are completed, and finally the processed sentences are fed into the translation model for translation.

3.3 Collaborative Machine Translation

1) Model Training

The process of model training is separated into the following steps:

Step1: Initialization

Although most of the previous studies rely on bilingual dictionaries, the method of pre-training cross-language embedding proposed is more simple and effective. First of all, the Byte-Pair Encoding (BPE) is done for the corpus of the two languages. This process reduces the size of the vocabulary and eliminates the unknown words in the output translation. Secondly, through the joint processing of the corpus of the two languages, the need for bilingual dictionaries is eliminated by using part of the shared words of the joint BPE. Finally, learn tag embedding on the same two-language joint corpus and use these embedding to initialize lookup tables in encoders and decoders.

Step2: Language model establishment

Language modeling is realized through the denoising automatic encoder, which optimizes and updates the neural network parameters for the source language and the target language respectively, so that the pseudo-parallel data generated by reverse translation is getting better and better. Minimize the objective function can be described as follows:

$$L^{lm} = E_{x \sim S}[-\log P_{s \to s}(x \mid C(x))] + E_{y \sim T}[-\log P_{t \to t}(y \mid C(y))] \tag{13}$$

Among them, C is a noise model in which a partial single word is deleted. $P_{s \to s}$ and $P_{t \to t}$ targets are composed of encoders and decoders working on the source side and the target side, respectively.

Step3: Reverse translation

In the process of reverse translation, by using a common encoder and decoder, the target function is as follows.

$$L^{back} = E_{y \sim T}[-\log P_{s \to t}(y \mid u^*(y))] + E_{x \sim S}[-\log P_{t \to s}(x \mid v^*(x))] \tag{14}$$

2) Semantic Quality Establishment

In order to solve the problem that neural machine translation needs a large number of bilingual parallel corpus, unsupervised neural machine translation uses the method of reverse translation to train the translation model. The method of reverse translation has always been widely used in the field of unsupervised neural machine translation. Usually, when minimizing the loss function, it is not back-propagated through the model that generates the data. This is to train the simplicity of reverse translation, but it ignores the importance of using data quality in neural machine translation training. Therefore, this paper proposes a square method of using quality estimation to control the quality of pseudo-parallel lines generated in the process of reverse translation training.

In this paper, the quality estimation model OpenKiwi is selected and trained. In order to train the OpenKiwi quality estimation model, we chose the quality estimation test set, verification set and training set provided by the WMT2019 shared task. In order to simplify the complexity of the system loading model, we only train the English-German and English-Czech quality estimation models,

which avoids the problem of introducing errors due to the use of different models in reverse translation and screening pseudo-parallel data. Finally, the sentence-level quality estimation models of English-German and English-Czech with Pearson coefficients of 58.51 and 54.91 are used in reverse translation.

3) Reverse Translation

The method of reverse translation (also known as round-trip translation) mainly uses encoders and decoders to train independent translation models on the monolingual corpus of the two languages. Therefore, during the training process, the unsupervised neural machine translation will train the same encoder and decoder iteratively in two directions, and finally apply the encoder and decoder to conventional machine translation.

The method of conventional reverse translation is a process of translating the source language to the target language and then back to the source language. Assuming that the source language sentence is S and the target language sentence is T, then the conventional reverse translation is S "T" S. The method used in this paper needs to use the QE model in the process of reverse translation. first, the conventional reverse translation process is divided into two stages: translating the source language sentence into the target language sentence, that is, S translating the target language sentence back to the source language sentence. Then, after the first stage, the QE model is added to evaluate the quality of the pseudo-parallel sentence pairs, and the evaluation result is the HTER value given by the QE model, and then the HTER value given by the QE model is compared with the threshold to filter the data. Finally, the pseudo-parallel data whose HTER value is greater than the threshold is carried out for two stages and the loss is calculated. after optimization, the neural network parameters are updated to end the current reverse translation process.

The final result of the estimator of the quality estimation model is the Human Translation Error Rate (HTER). The HTER metric is the percentage needed to be manually corrected in the translation. The HTER value ranges from 0 to 1, and the closer to 0 generation table, the better the translation, without manual editing. The closer to 1 means the worse the translation, which requires many manual edits. In the sentence-level quality estimation model, the translated parallel sentence pairs are put into the quality estimation model, and then the HTER value is predicted by the quality estimation model for us to select. Finally, the data obtained from the batch evaluation is calculated as the evaluation result of the pseudo-parallel data of the batch. Note that in order to improve the evaluation efficiency, we put a batch of pseudo-parallel data into the model to predict.

$$H_{batch} = \frac{\left(H_1 +, \cdots, +H_i +, \cdots, +H_n \right)}{n} \tag{15}$$

where n is the number of sentences in a batch and H_i is the HTER value of a pair of pseudo-parallel data in a batch. The threshold for screening pseudo-parallel data is to take the average value of the highest value (H_{max}) and the lowest value (H_{min}) of HTER for generating sentence pairs through many experiments, as follows:

$$\delta = \frac{(H_{max} + H_{min})}{2} \tag{16}$$

where is the threshold for filtering parallel data (we run the test ten thousand times and obtain the average value that can maximize the profits, that is, the threshold is generally set to 0.46 and 0.49 respectively). Finally, the pseudo-parallel data generated by the reverse translation are screened

through the fixed threshold, and then the pseudo-parallel data satisfying the threshold are used to train and update the neural network parameters.

As for the setting of the threshold, because the quality of generating pseudo-parallel data in the process of reverse translation training is getting better and better, and the threshold and HTER values are not fixed, if the method of dynamic threshold can not control the retention ratio of pseudo-parallel data, it is impossible to analyze the experimental results, so we did not choose to use the method of setting dynamic threshold. Instead, the method of setting a fixed threshold is adopted to control the quality of pseudo-parallel data.

4. PERFORMANCE EVALUATION

4.1 Setup

The parameters used for experimental simulation are summarized and shown in Table 1. In particular, the value of δ is set to 0.5. The batch size is set to 32 and the learning rate is set to 0.004. Moreover, the pre-processing data set used for model training and testing exceeds 100 million. This model has 4 network layers and 512 hidden layers.

Table 1. Parameters

Parameters	Value
Network layer	4
Multi-head attention	8
Word embedding dimension	512
Hidden layer dimension	512
Learning rate	0.004
δ	0.5
Batch Size	32
dropout	0.1
Pre-processing training set	>100 million

4.2 Experimental Results

The chapter control distribution of machine translation text can reflect the relationship between translation results and semantics. When the distribution of chapter control points is loose, it shows that the keywords of the translated text are not close enough to the semantic context, and the distribution of section control points is close. It shows that the translated text is correct, contextual and coherent. The analysis indicates that the designed English-Chinese machine automatic translation system has good accuracy in capturing key words, coherent and accurate translation, and suitable for semantics. The node control points of the other two systems are loosely distributed, and the keyword capture is not enough, and the semantic representation of the translated text is insufficient. The accuracy of automatic translation between English and Chinese by different systems is tested, and the contrastive results are shown in Table 2. In particular, the compared methods are Baseline1 (Sennrich et al. 2015), Baseline2 (Aiguier et al. 2019), Baseline3 (Gu et al. 2017), where 1) Baseline1 implements the translation by using the neural machine learning of rate words with sub-word units; 2)Baseline2 implements the translation by using the shared encoder and language-specific decoders; 3)Baseline3 implements the translation by using the non-auto-regressive neural machine learning.

According to the analysis of Table 2, the system can exceed 93% after 100 iterations. After the iteration of 500, the 100% accurate translation can be achieved, which is much higher than that of other methods. The highest accuracy of the other methods can also reach above 96% with large number of

Table 2. Translation accuracy (%)

Iterations	Proposed	Baseline1	Baseline2	Baseline3
100	**93.3**	83.6	85.2	86.0
200	**95.6**	86.7	87.7	90.4
300	**98.1**	89.1	91.3	92.7
400	**99.2**	93.7	94.3	93.2
500	**99.7**	93.9	94.7	93.9
600	**100**	94.9	95.2	94.0
700	**100**	95.0	95.4	98.7
800	**100**	95.8	95.6	100
900	**100**	98.2	97.6	100
1000	**100**	99.8	98.5	100

iterations. In this way, the more iterations also lead to more resource consuming, which is obviously a drawback. For example, only when the number of iteration exceeds 800, baseline3 can achieve the 100% accuracy. However, for the other two methods, their translation accuracy approaches 100% even when the number of iteration becomes 1000, which presents the benefits of the proposed method.

Table 3 shows the results of the data filtering comparison between the English translation Baseline models and the proposed model, as well as the comparison of the BLEU metric of the first epoch test. In particular, the parameter HTER only works for the proposed method. Then, as can be seen from Table 3, the square method described in this paper successfully uses the quality estimation model to control

Table 3. The date filtering ratio (%)

Iterations	HTER	Ratio (%)	BLEU
Proposed	**0.46**	**53.21**	**2.43**
Baseline1	—	92.8	1.02
Baseline2	—	89.1	1.13
Baseline3	—	100	1.43

the quality of pseudo-parallel data generated by unsupervised neural machine translation training.

By using the square method described in the algorithm section, the quality estimation model trained in this paper evaluates the pseudo-parallel data. Two kinds of translation are used, that is, the translation between English and Japaneses (E-J), the translation between English and Chinese (E-C). In particular, compared with the Baseline1 model, our model enriches the samples generated by reverse translation, makes the discriminator converge more slowly, and makes the generation countermeasure network trained more fully, thus improving the translation effect of unsupervised neural machines. The threshold obtained from the method described in the algorithm section is 0.46. As can be seen from Table 4, when selecting sentences, the proposed model reduces the number of sentences and increases the quality of model training (BLEU metric) at the same time. When the threshold is 0.36, the number of proposed model is 9.32% higher than that when the threshold is 0.46, but the BLEU

metric is lower than that when the threshold is 0.46. When the threshold is 0.56, the number of data removed by model screening is relatively large, which is not conducive to model training. The experimental results show that when the ratio of the filtered data to the retained data is close to 1: 1, the BLEU metric of the model is higher, and neither too much data can be filtered nor retained.

Table 4. The results achieved under E-J and E-C respectively

Model	E-J (HTER>0.46)		E-C (HTER>0.49)	
	BLEU	**PPL**	**BLEU**	**PPL**
Proposed	**18.11**	**44.56**	**12.01**	**53.2**
Baseline1	17.13	46.60	10.45	58.6
Baseline2	16.82	44.99	11.97	55.24
Baseline3	16.48	47.21	11.68	56.71

5. CONCLUSION

Unsupervised neural machine translation aims to use less monolingual corpus. To solve the problems of insufficient bilingual parallel corpus and language expansion in machine translation. The method proposed in this paper solves the existing unsupervised neural machine turning. It is impossible to control the quality of pseudo-parallel data when training reverse translation. We divide the conventional reverse translation into two stages. Firstly, the segment uses quality estimation to score the pseudo-parallel data generated by it. In contrast, the pseudo-parallel data with higher than the threshold value are screened out to complete the reverse translation. Secondly, through experiments, it is found that our method can enrich the network and quality based on generating confrontation at the same time by controlling the pseudo-parallel number according to the prime quantity. The results indicate that the proposed method can achieve better performance.

FUNDING AGENCY

REFERENCES

Abdel-Aziz, Y. I., Karara, H. M., & Hauck, M. (2015). Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. *Photogrammetric Engineering and Remote Sensing*, *81*(2), 103–107.

Aiguier, G., Oboeuf, A., Cobbaut, J. P., & Vanpee, D. (2015). Activity confrontation methods: A reflexive and metacognitive approach for interprofessional collaboration training. *Journal of Interprofessional Care*, *29*(5), 457–463.

Asghari, H., Fatemi, O., Mohtaj, S., Faili, H., & Rosso, P. (2019). On the use of word embedding for cross language plagiarism detection. *Intelligent Data Analysis*, *23*(3), 661–680.

Bustamante, G., Oncevay, A., & Zariquiey, R. (2020, May). No data to crawl? monolingual corpus creation from PDF files of truly low-resource languages in Peru. In *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 2914-2923). Academic Press.

Eum, S., Nakauchi, K., Murata, M., Shoji, Y., & Nishinaga, N. (2012, August). CATT: potential based routing with content caching for ICN. In *Proceedings of the second edition of the ICN workshop on Information-centric networking* (pp. 49-54). Academic Press.

Fotohi, R., & Bari, S. F. (2020). A novel countermeasure technique to protect WSN against denial-of-sleep attacks using firefly and Hopfield neural network (HNN) algorithms. *The Journal of Supercomputing*, *23*(4), 1–27.

Gu, J., Bradbury, J., Xiong, C., Li, V. O., & Socher, R. (2017). *Non-autoregressive neural machine translation.* arXiv preprint arXiv:1711.02281.

Ion, M., Zhang, J., & Schooler, E. M. (2013, August). Toward content-centric privacy in ICN: Attribute-based encryption and routing. In *Proceedings of the 3rd ACM SIGCOMM workshop on Information-centric networking* (pp. 39-40). ACM.

Janson, L., Schmerling, E., Clark, A., & Pavone, M. (2015). Fast marching tree: A fast marching sampling-based method for optimal motion planning in many dimensions. *The International Journal of Robotics Research*, *34*(7), 883–921.

Kepler, F., Trénous, J., Treviso, M., Vera, M., & Martins, A. F. (2019). *OpenKiwi: An open source framework for quality estimation.* arXiv preprint arXiv:1902.08646.

Kutscher, D., Eum, S., Pentikousis, K., Psaras, I., Corujo, D., Saucez, D., ... Waehlisch, M. (2016). Information-centric networking (ICN) research challenges. *Internet Research Task Force (IRTF)*, 1-38.

Li, B., Huang, D., Wang, Z., & Zhu, Y. (2016). Attribute-based access control for ICN naming scheme. *IEEE Transactions on Dependable and Secure Computing*, *15*(2), 194–206.

Martins, A. F., Junczys-Dowmunt, M., Kepler, F. N., Astudillo, R., Hokamp, C., & Grundkiewicz, R. (2017). Pushing the limits of translation quality estimation. *Transactions of the Association for Computational Linguistics*, *5*, 205–218.

Meyer, K. D., Patil, D. P., Zhou, J., Zinoviev, A., Skabkin, M. A., Elemento, O., & Jaffrey, S. R. et al. (2015). 5¢ UTR m6A promotes cap-independent translation. *Cell*, *163*(4), 999–1010.

Muscariello, L., Carofiglio, G., Auge, J., & Papalini, M. (2018). *Hybrid information-centric networking.* Internet Engineering Task Force, Internet-Draft draft-muscariello-intarea-hicn-00.

Naeem, M. A., Ali, R., Kim, B. S., Nor, S. A., & Hassan, S. (2018). A periodic caching strategy solution for the smart city in information-centric Internet of Things. *Sustainability*, *10*(7), 2576.

Neirotti, P., De Marco, A., Cagliano, A. C., Mangano, G., & Scorrano, F. (2014). Current trends in Smart City initiatives: Some stylised facts. *Cities (London, England)*, *38*, 25–36.

Ren, S., Zhang, Z., Liu, S., Zhou, M., & Ma, S. (2019, July). Unsupervised neural machine translation with smt as posterior regularization. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 241–248.

Rudebeck, P. H., Rich, E. L., & Mayberg, H. S. (2019). From bed to bench side: Reverse translation to optimize neuromodulation for mood disorders. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(52), 26288–26296.

Sen, S., Gupta, K. K., Ekbal, A., & Bhattacharyya, P. (2019, July). Multilingual unsupervised NMT using shared encoder and language-specific decoders. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3083-3089). Academic Press.

Sennrich, R., Firat, O., Cho, K., Birch, A., Haddow, B., Hitschler, J., . . . Nădejde, M. (2017). *Nematus: A toolkit for neural machine translation.* arXiv preprint arXiv:1703.04357.

Sennrich, R., Haddow, B., & Birch, A. (2015). *Neural machine translation of rare words with subword units.* arXiv preprint arXiv:1508.07909.

Shakhnovich, V. (2018). It's time to reverse our thinking: The reverse translation research paradigm. *Clinical and Translational Science*, *11*(2), 98.

Shatsky, I. N., Terenin, I. M., Smirnova, V. V., & Andreev, D. E. (2018). Cap-independent translation: What's in a name? *Trends in Biochemical Sciences*, *43*(11), 882–895.

Su, K., Li, J., & Fu, H. (2011, September). Smart city and the applications. In 2011 international conference on electronics, communications and control (ICECC) (pp. 1028-1031). IEEE.

Trossen, D., Reed, M. J., Riihijärvi, J., Georgiades, M., Fotiou, N., & Xylomenos, G. (2015, June). IP over ICN-the better IP? In *2015 European conference on networks and communications (EuCNC)* (pp. 413-417). IEEE.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., . . . Dean, J. (2016). *Google's neural machine translation system: Bridging the gap between human and machine translation.* arXiv preprint arXiv:1609.08144.

Yi, B., Wang, X., & Huang, M. (2021). Computation Migration Oriented Resource Allocation in Mobile Social Clouds. *IEEE Transactions on Cloud Computing*, *1*(1), 1–12.

Zhang, X., Liu, G., Yang, C., & Wu, J. (2018). Research on air confrontation maneuver decision-making method based on reinforcement learning. *Electronics (Basel)*, *7*(11), 279.