

Zhang, P., Guan, Z., Liu, B., Ding, X. (S.), Lu, T., Gu, H. and Gu, N. (2022) Building user-oriented personalized machine translator based on user-generated textual content. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2), 280. (doi: 10.1145/3555171)

Publisher's URL: https://dl.acm.org/doi/10.1145/3555171

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Copyright © 2022 Association for Computing Machinery. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. ACM on Human-Computer Interaction, 6(CSCW2), 280.

http://eprints.gla.ac.uk/286779/

Deposited on: 09 March 2023

Enlighten – Research publications by members of the University of Glasgow <u>http://eprints.gla.ac.uk</u>

PENG ZHANG^{*}, School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China

ZHENGQING GUAN*, School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China

BAOXI LIU, School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China

XIANGHUA (SHARON) DING, (1) School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China; (2) School of Computing Science, University of Glasgow, Glasgow, UK

TUN LU[†], School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China

HANSU GU, Seattle, USA

NING GU, School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China

Machine Translation (MT) has been a very useful tool to assist multilingual communication and collaboration. In recent years, by taking advantage of the exciting developments of neural networks and deep learning, the accuracy and speed of machine translation have been continuously improved. However, most machine translation methods and systems are data-driven. They tend to select a consensus response represented in training data, while a user's preferred linguistic style, which is important for translation comprehension and user experience, is ignored. For this problem, we aim to build a user-oriented personalized machine translation model in this paper. The model aims to learn each user's linguistic style from the textual content that is generated by her/him (User-Generated Textual Content, UGTC) in social media context and generate personalized translation results utilizing several state-of-the-art deep learning techniques like Transformer and pre-training. We also implemented a user-oriented personalized machine translation using Weibo as a case of the source of UGTC to provide a systematical implementation scheme of a user-oriented personalized machine translation system based on our model. The translator was evaluated by automatic evaluation in

*Both authors contributed equally to this research.

[†]Corresponding author.

Authors' addresses: Peng Zhang, zhangpeng_@fudan.edu.cn, School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China; Zhengqing Guan, zqguan20@fudan.edu.cn, School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China; Baoxi Liu, liu_baoxi@foxmail.com, School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China; Xianghua (Sharon) Ding, xianghua.ding@glasgow.ac.uk, (1) School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China; and (2) School of Computing Science, University of Glasgow, Glasgow, UK; Tun Lu, lutun@fudan.edu.cn, School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China; Hansu Gu, hansug@acm.org, Seattle, USA; Ning Gu, ninggu@fudan.edu.cn, School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China; Hansu Gu, hansug@acm.org, Seattle, USA; Ning Gu, ninggu@fudan.edu.cn, School of Computer Science & Shanghai Key Laboratory of Data Science, Fudan University, Shanghai, China;

combination with human evaluation. The results suggest that our model can generate more personalized, natural and lively translation results and enhance the comprehensibility of translation results, which makes its generations more preferred by users versus general translation results.

CCS Concepts: • Human-centered computing \rightarrow Collaborative and social computing.

Additional Key Words and Phrases: machine translation, personalized, linguistic style, user-generated textual content, Weibo

ACM Reference Format:

Peng Zhang, Zhengqing Guan, Baoxi Liu, Xianghua (Sharon) Ding, Tun Lu, Hansu Gu, and Ning Gu. 2022. Building User-oriented Personalized Machine Translator based on User-Generated Textual Content. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 280 (November 2022), 26 pages. https://doi.org/10.1145/3555171

1 INTRODUCTION

Multilingual communication and collaboration is becoming increasingly common and important with the emergence and prevalence of multilingual communities, multilingual projects, distributed work and international conferences [13, 54]. Language barrier is still the primary challenge for effective multilingual communication and collaboration, and a variety of tools have been developed to help people overcome language barrier and achieve effective communication and collaboration [16, 53]. Machine Translation (MT), which aims to translate a text from one natural language to another using computer-aided tools, has been a very useful tool to assist multilingual communication and collaboration and collaboration [34]. In recent years, an increasing number of multilingual communities, projects and organizations are proposing machine translation for communication and collaboration support [53], and there also emerge different kinds of machine translation systems or tools like Google Translate ¹, Microsoft Translator ² and Amazon Translate ³. As machine translation's crucial role in multilingual communication and collaboration, it has long been researched in the areas of artificial intelligence, Computer-Supported Collaborative Work (CSCW) and Human-Computer Interaction (HCI) [18, 54].

Machine Translation is continuously being improved and refined as the exciting developments of neural networks and deep learning. Neural Machine Translation (NMT) has been shown to be a promising end-to-end learning approach that can conduct translation accurately and quickly [51]. The key benefit of this approach is that a translation model can be trained directly on a large volume of source and target text, in an end-to-end fashion, the mapping from input text to associated output text. By taking advantage of NMT, the accuracy and speed of machine translation are all continuously improved. For example, from 2019.5 to 2020.5, Google Translate obtains an average 5 point increase on BLEU (Bilingual Evaluation Understudy) and human evaluation metrics across all languages [50]. However, most machine translation methods and systems are data-driven. They tend to select the response with the greatest likelihood, i.e., a consensus response represented in training data [24], while users' personal preferences in linguistic style are ignored. As suggested by previous research [27, 28], people often cannot comprehend machine translation results in part due to the difficulty of sense making, and the future design of machine translation systems should forage more useful information to enhance sense making. Linguistic style is one of the key components of natural languages, and it can significantly affect people's sense making of languages [21]. In reality, each person has her/his preferred linguistic style in communication, and the ability of language systems to exhibit a user's preferred style is helpful for the systems to interact with the user in a natural way that the user is familiar with, which can promote sense making and gain

¹https://translate.google.cn/

²https://www.microsoft.com/en-us/translator//

³https://aws.amazon.com/cn/translate/

the user's confidence and trust [59]. So considering a user's linguistic style preference in machine translation can be helpful for the systems to generate translations that are easily comprehended by the user and improve user experience.

However, building a machine translator that can consider users' preferences in linguistic style (user-oriented personalized machine translator) is a challenging task. First, the essence of useroriented personalized machine translation is to generate multiple translations for the same source text, wherein the multiple translations are characterized with the same semantics but differ in linguistic styles regarding different users. Building such a machine translation model is challenging as the definition of the linguistic style is vague, and linguistic style and semantics are generally interweaved in natural languages [9]. It is not easy to disentangle linguistic style from semantics and change the linguistic style while retaining the semantics. Second, there lacks sufficient persona data (a composite of elements that can reflect a user's preference) related to linguistic style to train a user-oriented personalized machine translation model. As mentioned above, most machine translation models are data-driven [24]. Personalized translation results can be generated only if sufficient samples that contain linguistic style-related persona profiles are utilized for model training. In related personalized language generation tasks like personalized dialog generation, personalized comment generation and personalized recipe generation, two kinds of persona profiles including explicit persona profiles and implicit persona profiles are generally utilized for personalized language model training [59]. Explicit persona profiles are observable personal traits like gender, age and personal interests, while implicit persona profiles mean the characteristics that are extracted from users' historical activities. For example, in personalized recipe generation systems [30], gender and preferred foods a user self-reports on the platform can be utilized as explicit persona profiles, and her/his visited historical recipes can be utilized for implicit persona extraction. In nowadays machine translation systems, there is no explicit persona profile nor implicit persona profile, which brings challenges to the training of a user-oriented personalized machine translation model. Third, in translation tasks, what kinds of sentences and words need to be personalized is unknown. As suggested by previous research [60], real-world languages are not always persona-related. So there is no need to make personalization for each sentence or word to be translated. For example, for the formal descriptions like news or scientific articles, the personalized demand might be minor as people tend to express them similarly, while for some informal descriptions like novels or stories, the personalized demand might be higher.

The above backgrounds motivate us to build a user-oriented personalized machine translator in this paper. The characteristics of our work lie in: 1) To overcome the problem that there is a lack of sufficient persona data to train a user-oriented personalized machine translation model, we build a user-oriented personalized machine translation model by learning each user's linguistic style from the textual content that is generated by her/him (User-Generated Textual Content, UGTC) on social media platforms like social network sites and online communities. The theories in sociolinguistics indicate that people tend to perform persona when using language to socialize, which leads UGTC of social media to be an ideal corpus with diversified personality traits [46, 59, 60]. 2) We build our model based on state-of-the-art machine translation techniques including Transformer and pre-training to generate multiple linguistic styles associated with the same semantics. Firstly, Transformer and existing public machine translation data sets are utilized to train a translation module that generates a general translation result. After then, the linguistic style of the general translation result will be processed to be personalized by a Transformer-based module that is pre-trained on all users' UGTC posts and tuned on each user's own posts. 3) Based on the proposed user-oriented personalized machine translation model, we present a user-oriented personalized machine translation prototype using Weibo as a case of the source site of UGTC. The prototype provides a systematical implementation scheme of a user-oriented personalized machine translation system, including how to introduce UGTC into the system with user consent, how to train and use our user-oriented personalized machine translation model, etc. 4) To evaluate our user-oriented personalized machine translation model and prototype, we conduct both automatic evaluation and human evaluation. In the automatic evaluation, personalization degree, BLEU and perplexity are set as metrics, and in the human evaluation, we intend to evaluate the comprehensibility of the translation results and users' real experience of using the user-oriented personalized machine translator. The evaluation results indicate that our model can generate more personalized, natural and lively translation results and enhance the comprehensibility of translation results, which makes its translations more preferred by users versus general translations. However, we also find performing linguistic style personalization can slightly hurt translation fluency, accuracy and perplexity. To conclude, our main contributions are:

- We aim to build a user-oriented personalized machine translator in this paper. To the best of our knowledge, it is the first work that considers users' linguistic style preferences in machine translation.
- We propose a user-oriented personalized machine translation model which can learn a user's linguistic style from her/his UGTC of social media and generate personalized translation results.
- We implement a user-oriented personalized machine translation prototype which provides a systematical implementation scheme of a user-oriented personalized machine translation system based on our model.
- We conduct automatic evaluation and human evaluation for our model and uncover its strengths and weaknesses compared with general machine translation models. Based on the results, we propose several insights for the future design of machine translation systems.

The rest of this paper is organized as follows. In Section 2, we review related research of linguistic style, machine translation and personalized language generation methods. In Section 3, we show the architecture of our model. A prototype is presented in Section 4, and the evaluation procedure and results are exhibited in Section 5. In Section 6, we propose insights for future study and highlight the limitations of our work. Finally, the conclusion is given in Section 7.

2 RELATED WORK

2.1 Linguistic Style

Linguistic style is one of the key components of natural languages [21]. It reflects how people use a variety of linguistic transformations like lexical and grammatical transformations, formality adjustment and catchy phrase selection for language organization [20]. As suggested by previous research [8, 42, 59], each person has her/his preferred linguistic style and a unique stylistic tendency in communication. In these studies, many features are proposed to represent linguistic style, including character-based features (e.g. the number of special characters and the number of white-space characters), word-based features (e.g. the total number of words and vocabulary richness), function words(e.g. the number of article words and the number of pro-sentence words), etc [8].

In linguistic style studies, one emerging research topic named linguistic style transfer is related to our user-oriented personalized machine translation task. It aims to change the linguistic style of the current text to another style while retaining the semantics, e.g. transfer between formal language and informal language. The linguistic styles studied in a linguistic style transfer task are more specific than those of the aforementioned linguistic style research as there are generally two predefined styles (informal and formal, etc.) in the task. For example, [20] proposed a model to transform text from modern English to Shakespearean English, [36, 49] focused on the style transfer between formal language and informal language and presented corpus, benchmarks and

metrics for model building, etc. Our task differs from these linguistic style transfer tasks in the following two aspects. First, there are generally two styles (informal and formal, etc.) in a linguistic style transfer task. So the goal of that is to change the linguistic style of a text to another style. Our task aims to generate a personalized linguistic style that is consistent with each user's style preference. The linguistic styles contained in our task are more diverse and vague. Second, as the linguistic style transfer task just contains a certain number of styles, it is not difficult to obtain a large number of training samples (parallel samples or non-parallel samples). While in our task, we need to prepare sufficient data to reflect each user's linguistic style preference.

2.2 Machine Translation

Machine Translation has long been researched in both artificial intelligence and the areas of CSCW and HCI. In the artificial intelligence area, although many different kinds of machine translation methods have been proposed, they can be summarized into two categories: rule-based approach and corpus-based approach [34]. The rule-based approach makes syntax analysis and semantic analysis for the source text and then generates the target text according to a finite set of gigantic dictionaries and sophisticated linguistic rules. The training and development costs are very high for rule-based machine translation systems, and the translation results generally lack fluency as they are generated just by pre-defined hard rules. To overcome these drawbacks, researchers propose the data-driven machine translation approach - corpus-based approach. It relies on a large parallel corpus that contains a large number of pairs of the source text and target text. The most representative corpus-based approach is Statistical Machine Translation (SMT) technology [17, 47]. As the superiority of neural networks and deep learning, many researchers have attempted to incorporate deep learning techniques like RNN (Recurrent Neural Network) and attention mechanism into SMT and proposed many neural machine translation methods. NMT models can be directly trained on a large volume of source and target text, in an end-to-end fashion, the mapping from input text to associated output text [51]. They have been shown to be a promising end-to-end learning approach that can conduct translation quickly and accurately, which makes them dominant in machine translation. The core of NMT models is the Seq2Seq (Sequence to Sequence) structure which includes an encoder and a decoder [40]. The encoder takes the sequence of the source text as input and generates a representation (context vector) for that, and then the decoder reads the context vector and generates the target text word by word.

Machine translation has also gathered lots of attention in the areas of CSCW and HCI, and several studies have focused on machine translation-mediated communication and collaboration. First, some studies intend to investigate the role and impact of machine translation in multilingual communication and collaboration, e.g. its effects on conversational efficiency and content [54], the effects on idea exchange [44] and the impact on conversation style [16]. On one hand, it has been suggested that machine translation is helpful to promote idea production and collaboration experience in multilingual communication and collaboration. On another hand, these studies find that there exist some drawbacks (e.g. inconsistency of the same term in translation [54], benefit asymmetries between native and non-native speakers [44] and difficulties in establishing common ground in machine translation-mediated groups [53]) of machine translation-mediated communication and collaboration due to limitations in state-of-the-art machine translation methods, tools and systems, based on which several design insights are proposed for the future machine translation. Second, based on these insights, many studies aim to propose new machine translation methods, tools and systems to promote machine translation-mediated communication and collaboration. For example, [15, 18, 19] proposed new machine translation systems to support human editing, [14, 52] improved machine translation by showing two outputs, and [3] presented a new machine translation system with human-in-the-loop interpretation.

Besides the above research, some studies have begun to investigate personalized machine translation mainly from the perspective of translation authors [31, 32, 35, 48]. They hold that good translation is expected to preserve an author's subtle and implicit personality and characteristics, and then propose several methods for author-aware personalized machine translation. Most of these methods define this problem as a domain a daptation task [31, 32, 35, 48], where the domain represents a specific author trait. By designing domain-specific models, it can preserve the original author traits in machine translation. Compared with these author-aware machine translation studies, our user-oriented personalized machine translation conducts personalization from the opposite perspective, which is a novel task. On one hand, author-aware machine translation studies and our task have different research aims and evaluating criteria. The former aims to promote objective indicators like translation accuracy and fluency by considering author traits, while our user-oriented personalized machine translation optimizes machine translation from the view of users with the aim of promoting translation comprehension and user experience. On another hand, our task faces different challenges. First, the author-aware machine translation needs to preserve the traits of some given authors in translation, while for our user-oriented personalized machine translation task, it aims to mine linguistic style preferences for a large number of users and generate translations with different linguistic styles regarding different users. Persona profile mining and personalized translation generation for so many users are non-trivial. Second, previous author-aware personalized machine translation research mainly uses explicit and simple author traits like gender as persona profiles [31, 32, 35]. In our task, there is no explicit persona profile nor implicit persona profile, aggravating the difficulty of building a user-oriented personalized machine translation model.

2.3 Personalized Language Models

Building human-like language systems is a long-standing research focus in recent years. One of the main challenges is to generate personalized content for each user in order to promote comprehension and gain her/his confidence and trust [59, 60]. As mentioned above, existing personalized language models can be classified into two kinds: explicit personalization and implicit personalization [59]. Explicit personalization models directly introduce explicit personalized responses, while implicit personalization models extract persona profiles from the user's historical activities. The following reviews recent studies of personalized language models in terms of these two branches.

From the perspective of explicit personalization, many user traits like gender and age have been considered by researchers to generate personalized responses. A major challenge of explicit personalization is how to introduce such explicit persona profiles into general language models. Previous research has proposed several strategies to introduce them into the encoder and decoder of the Seq2Seq model. From the perspective of the encoder, there are three common-used strategies: additional vocabulary, embedding combination and user selection [25, 60, 61]. According to the additional vocabulary strategy [61], a structured vector representing user traits is set as the initial token of each input in the encoder, and its embedding will be processed together with other word embeddings. Different from that, embedding combination utilizes the user vector in each time step and combines it with each of the word embeddings by concatenation or addition [60]. While in reality, not all sentences and words are equipped with personality traits. Such hard integration strategies cannot be able to decide which personality traits to express on each word or sentence. To overcome this problem, the user selection strategy builds gates based on user traits and utilizes them to control the personality of each word [25]. From the perspective of the decoder, there are also two strategies to introduce user traits. The first one is to consider user traits in the decoder procedure, i.e., the user vector is combined with the original input of the decoder or the decoder

state of each token [56, 60], while the second strategy incorporates user traits in the output layer of the decoder [59].

As explicit user traits are sometimes difficult to obtain, researchers have been exploring implicit personalization methods in recent studies. A direct approach is to train a language model for each user based on her/his historical activities. [33] proposed a personalized review generation model by regarding words written by different reviewers as different vocabularies. [58] first pre-trained a general response generation model on large-scale conversational data, and then made the responses to be in agreement with a user's preference by tuning the model on the user's personal conversation data. Another thought of implicit personalization research is to extract user preference from historical activities and then introduce the preference representation into the model by utilizing similar strategies that explicit personalization methods use. Most of such work considers user preference in the decoder module. On one hand, some studies incorporate user preference into the output layer of the decoder [25, 26, 30]. On another hand, some research considers user preference in the core layers of the decoder, like the persona-based neural conversation models [23, 24]. Besides the above work, some research attempts to incorporate user traits into the encoder. For example, [7] proposed a personalized model named Template-based Personalized EDM Subject generation (TemPEST) by considering users' preferred style sentences and templates in the encoder to generate a personalized article representation and subject generation.

Although a few studies have focused on different kinds of personalized language systems, our work differs with such tasks in two aspects. First, existing personalized language systems intend to generate personalized responses that are consistent with a user's content preference, and they generally conduct personalization without distinction of semantics and linguistic style. For our user-oriented personalized machine translation task, the translations associated with the same source text should correspond to the same semantics but different linguistic styles, which brings a new challenge for model construction. Second, the contexts studied in previous research like dialogue systems and recipe recommendation platforms have some explicit persona profiles and implicit persona profiles that can be utilized for personalized model training, while in nowadays machine translation systems, there is no explicit persona profile nor implicit persona profile. So the existing personalized language models cannot be utilized for our task.

3 USER-ORIENTED PERSONALIZED MACHINE TRANSLATION MODEL

3.1 Technical Background of Model Building

We build our user-oriented personalized machine translation model based on state-of-the-art neural machine translation models. In this section, we introduce the crucial thoughts and technical details of neural machine translation. NMT models are also called end-to-end translation models as they can be directly trained on a large volume of source and target text [51]. The core of NMT models is the Seq2Seq structure which includes an encoder and a decoder [40]. For the source text which is represented as a word index sequence $X = X_1, \dots, X_t, \dots, X_T$ (*T* is the sequence length), the encoder aims to generate hidden states to represent *X* using neural networks. RNN models are the most commonly used methods for this task. Using RNN, the hidden states are generated as $h_t = f(h_{t-1}, X_t)$, where h_t is the hidden state corresponding to the *t*th word, and *f* is a nonlinear function which is generally implemented as LSTM (Long Short-Term-Memory) [37] or GRU (Gated Recurrent Unit network) [55]. After that, the encoder generates a context vector *c* (e.g. setting $c = h_T$ or computing it using attention mechanisms) as the final representation of the source text. In the decoder, the target text is generated word by word generally using RNN models. At the *t*th step of the target text, it first generates a hidden state as $s_t = f(s_{t-1}, [y_{t-1}, c])$, where s_t is the corresponding hidden state, y_{t-1} is the output at the (t-1)th step, and f is a nonlinear function similar to that of the encoder. Finally, s_t is utilized to map to a distribution over the vocabulary to determine the *t*th word of target text by using the maxout activation function [58].

According to the aforementioned thoughts and procedures, many Seq2Seq models have been proposed, and the most famous one is Transformer. Transformer is a state-of-the-art model for dealing with sequences. The most prominent application of it is in natural language processing tasks especially machine translation. Transformer is essentially a stack of encoder and decoder layers [43]. Compared with traditional Seq2Seq models, it has two characteristics that improve its performance on accuracy and efficiency. First, there are six encoders with the same structure in the encoding module and six decoders in the decoding module, which reflects the thought of deep learning. Second, it incorporates the multi-head self-attention mechanism into both encoders and decoders. By taking advantage of the self-attention mechanism, each input word is represented as its embedding in combination with the embeddings of the other relevant words, while masking the words that contain irrelevant information. Multi-head means multiple self-attention mechanisms are implemented in parallel using the parallel computing offered by GPUs. Transformer can process multiple words simultaneously, which reduces the running time of model training and prediction. More technical details about Transformer can be found in [43].

3.2 Model Building

In this section, we build a user-oriented personalized machine translation model by setting Englishto-Chinese translation (i.e., translating English into Chinese) as a sample task. However, our model is not specific to such a translation task. As we elaborate later, it can be easily applied to other crosslanguage translation tasks. The architecture of our user-oriented personalized machine translation model and its training procedure are shown in Figure 1. The main modules are elaborated below.



Fig. 1. User-oriented personalized machine translation model.

UGTC in a social media site. As summarized in related work, existing personalized language systems generate personalized responses using the explicit persona profiles or the implicit persona profiles mined from historical activities. While most machine translation systems focus on the translation functionality itself, and users' profiles like gender or age and activities like translation result feedback are ignored. So we cannot obtain persona data from machine translation systems to build the user-oriented personalized translation model. In our paper, we utilize the UGTC of

social media platforms like social network sites and online communities as the corpus to learn users' linguistic style preferences. The reasons are three-fold. First, the theories in sociolinguistics suggest that people tend to perform persona when using language to socialize. So the social media UGTC is an ideal corpus with diversified personality traits [59, 60]. Second, users generate a large volume of content on social media platforms every day, which provides us sufficient data to train a personalized translation model. For example, on Twitter, users post over 300,000 tweets every minute, and on Facebook, more than 680,000 posts are published per minute [41]. Third, we only consider the UGTC published publicly. Compared with the personalized language models which utilize users' explicit attributes like gender and age for personalization, our method is characterized by a lower risk of invading user privacy.

Translation generation and personalization. This component is utilized to generate personalized translation results for a given user. It contains two Transformer-based models: E2C (Englishto-Chinese) Transformer and C2C (Chinese-to-Chinese) Transformer. E2C Transformer is essentially a general English-to-Chinese translator based on the Transformer structure, wherein the input is the English text that needs to be translated, and the output is the corresponding Chinese text (general translation result). It can be trained utilizing several public Chinese-English translation data sets like WMT Parallel English/Chinese test set ⁴, NIST Chinese-English test sets ⁵ and TED corpus ⁶. The C2C Transformer is utilized to polish the translation result to make it personalized (personalized translation result), i.e., the linguistic style is consistent with the user's preference. So the input of it is the general translation result, and the output is the personalized translation result. Note that C2C Transformer can only change the linguistic style of the general translation result, while the semantics should be retained.

Training corpus preparation. To train a personalized translation model for a given user, we need a parallel training data set that contains many pairs of training samples, wherein each pair consists of the English text to be translated and the ground truth (the corresponding personalized Chinese text). However, to the best of our knowledge, there is no such data set currently. To construct the data set by hand-annotating is also characterized by a labor-intensive nature. In our work, we solve this problem based on back-translation [11, 38]. Back-translation is an effective method for neural machine translation with monolingual data. Its crucial thought is constructing a parallel corpus with back-translations of target language sentences, i.e. automatically translating the target sentence into the source language. Based on this thought, we propose an approach for training corpus preparation by utilizing the C2E (Chinese-to-English) Transformer and UGTC. As mentioned above, a user's UGTC in a social media site can reflect her/his preference in linguistic style. So in each pair of our training samples, we set each UGTC post (e.g. Weibo post) which is written in Chinese as the ground truth, i.e., the expected personalized translation result. The C2E Transformer, which is essentially a Transformer-based Chinese-to-English translator, takes the post as the input and outputs the corresponding English text to be input into the E2C Transformer. The C2E Transformer can also be trained utilizing the existing Chinese-English translation data sets. By taking advantage of the above procedures, it can lead the input of the C2C Transformer and output of that to correspond to the same semantics but reflect different linguistic styles (the input reflects the style of a general translator, and the output reflects a person's preferred style), which can meet the requirement of C2C Transformer. That is why we consider such a method for training corpus preparation.

⁴http://www.statmt.org/wmt20/translation-task.html

⁵https://chinesenlp.xyz/#/docs/machine_translation#span-classtnistspan

⁶http://cs.jhu.edu/~kevinduh/a/multitarget-tedtalks/

Deep learning models generally need a large number of samples for training. A user's posts may not be sufficient for training the C2C Transformer. We solve this problem by pre-training and tuning which is an effective strategy to process prediction tasks with insufficient labeled training data. We first pre-train the C2C Transformer utilizing all users' training samples, and then the model will be tuned to be specific to a given user by using her/his corresponding training samples. After training, the personalized translation model can be characterized with the capability of personalized translation generation. When the user submits some English sentences that need to be translated, the E2C Transformer reads these sentences and generates a general translation result, and then the C2C Transformer further processes the general result and generates the final personalized translation expression.

Above all, we can see such a personalized translation model has the following three characteristics.

- Social media UGTC is utilized as the corpus for model training. Our model aims to learn users' linguistic styles from a large volume of textual content published publicly by users in social media context. It provides a safe and efficient approach to solve the data sparsity problem of user-oriented personalized machine translation.
- It represents the thought of pre-training and tuning. As a user's posts may not be sufficient for personalized translation model training, the C2C Transformer is pre-trained utilizing all users' training samples and then tuned to be personalized using each user's own training samples. This strategy is helpful to improve the performance of user-oriented personalized machine translation, especially for users with few posts.
- The model can be easily generalized into the other cross-language translation tasks. When applying this model into the other cross-language translation tasks, only the training data sets including UGTC and general translation data sets need to be changed, and there is no need to adjust the architecture of the model.

4 **PROTOTYPE**

In order to exhibit how our user-oriented personalized machine translation model can be utilized, we designed and implemented a personalized machine translator using Weibo as a case of the source of UGTC. The translator provides a systematical implementation scheme of a user-oriented personalized machine translation system based on our model, including how to introduce UGTC into the system with user consent, how to train and use our model, etc. The reasons why we set Weibo as a study case are three-fold. First, Weibo is a very popular social media site in China. It has a large user base, and more than 229 million users are active in Weibo every day to obtain and share the latest news, post comments, participate in group chats, etc.⁷, which provides us sufficient posts to perform user-oriented personalized translation. Second, as a popular social media site, Weibo has the representative features that popular social media has, such as content posting, communicating and group maintaining. It is helpful to improve the reliability and generalizability of our findings. Third, similar to many popular social media platforms like Facebook and Twitter, Weibo supplies researchers and developers with some APIs to acquire user consent and access user data, based on which we can implement a personalized translation prototype and use users' posts in a reasonable manner. The main modules of the user-oriented personalized machine translator are shown in Figure 2.

Authorization management. This module is utilized for authorization management. It consists of the following three features.

⁷https://www.iimedia.cn/c1020/74841.html



Fig. 2. The architecture of the prototype.

- Account authorization. This sub-module serves the functionality for Weibo identity authentication. Weibo Open Platform ⁸ supplies an identity authentication API based on OAuth2.0, which provides us with an easy and safe way to implement the user consent functionality.
- Post authorization. After account authorization, this sub-module serves the functionality for post access. Weibo Open Platform supplies an API to acquire a specific user's posts. In our translator, we invoke this interface to implement the post authorization functionality.
- Remove authorization. This sub-module aims to help a user remove user authorization and post authorization from the translator. It is implemented based on the OAuth2.0 revoking function supplied by Weibo.

Translator. This module is utilized to receive a user's text to be translated and present the corresponding personalized translation result.

User-oriented personalized machine translation model. It is the user-oriented personalized machine translation model presented in Section 3.

When a user utilizes the prototype for translation, she/he first binds her/his Weibo account and authorizes the use of her/his posts through the account authorization and post authorization respectively. Then by utilizing these posts as a training set, a user-oriented personalized machine translation model will be trained for the user according to the architecture presented in Section 3. After the model is built, the user can input the English text to be translated, and the model will read it into the E2C Transformer and C2C Transformer to generate the personalized translation result. Finally, the result will be presented to the user. Note that in the above procedure, the user-oriented personalized machine translation model just needs to be trained once. In the follow-up use, the user can use the translator directly, and translation results can be generated fast.

Through the above descriptions, we can see our user-oriented personalized machine translator has the following two characteristics.

- Using posts with user consent. A user's Weibo posts can be accessed only if the user gives account authorization and post authorization, which ensures that the translator utilizes user data in a reasonable manner.
- Working fast as existing machine translation systems. After training, the complexity of the translation procedure is similar to that of nowadays prevalent machine translation systems. So even though our translator can perform personalized translation, its running time is similar to that of existing translation systems.

We implemented the user-oriented personalized machine translator according to the architecture depicted in Figure 2. The implementation was based on Django ⁹ which was a Python Web

⁸https://open.weibo.com/

⁹https://www.djangoproject.com/

framework supporting rapid development and clean, pragmatic design. The screenshots of the prototype are shown in Figure 3 and Figure 4. Figure 3 exhibits the home page of our prototype, and Figure 4 shows the modules for authorization management. The instructions of the web pages were written in Chinese. For understanding, we explain the Chinese instructions in English in the figures.

	authorization management	
personalized machine translator		
please input the	▶ 请输入需要翻译的内容	
submit & reset	▶ ■章 ▶ 个性化翻译结果	

Fig. 3. Screenshot of the home page of user-oriented personalized machine translator.



 (a) Submenus under (b) Authorize to log in with Weibo account. authorization management

Fig. 4. Screenshots of authorization modules: a Click "authorize" to initiate authorization; b The system will jump to the account authorization page supplied by Weibo. A user can give account authorization by scanning the QR code (the QR code was blurred for anonymity using the Weibo APP; c The system will jump to the post authorization page. The user can give post authorization by clicking "confirm".

5 EVALUATION

In this section, we present the evaluation for our user-oriented personalized machine translation model. We made both automatic evaluation and human evaluation. The former aimed to evaluate the personalization degree and accuracy of the personalized translation results in comparison with general translation results. However, the comprehensibility of translations and users' real experience could not be validated by automatic evaluation. So we further made a human evaluation wherein several genres of languages were utilized as source texts, and many participants were recruited to use the translator and give their feedback.

5.1 Automatic Evaluation

Data Set and Model Setting. As the popularity and representative features of Weibo, we 5.1.1 also utilized a large number of Weibo posts to train and evaluate our user-oriented personalized machine translation model. As mentioned in Section 3.2, the C2E and E2C Transformers should be trained firstly by using the public Chinese-English translation data sets. We utilized the benchmark Chinese-English data sets including CWMT corpus and NIST that were recommended by stateof-the-art Chinese-English machine translation methods [5, 45]. For the Weibo posts, they were sampled randomly based on the snowball sampling rule. After about three months of sampling, we finally obtained 2,217,006 posts that belonged to 4,587 users. Note that to avoid the invasion of users' sensitive content, we just considered the posts that were published publicly by users. For the data set, we conducted pre-processing from two perspectives: post and user. From the perspective of the post, it aimed to filter out the posts that were meaningless for our model training, while from the perspective of the user, the users with less content contribution were removed from the data set. First, there are generally some noisy items like emojis and tags in social media posts. So we filtered these items from the sampled results. As we set English-to-Chinese as a research case of the user-oriented personalized machine translation task, there was no need to consider the bilingual posts (containing both Chinese words and another language) and the posts that were written in the non-Chinese language. Moreover, the reposts could hardly reflect a user's linguistic style, and the posts with short length (lower than ten words according to previous research [1]) were less useful for the model training. So we removed such posts from the data set. After the above pre-processing, 1,135,247 posts corresponding to 3,446 users were left. Second, among these users, some users had low activity degrees, and they could not supply sufficient content for user-oriented personalized machine translation model training. To filter out such users, we followed previous research [6, 29], and removed the users whose posts were less than 50 from the data set. Finally, 1,128,327 posts that belonged to 3,256 users were retained as the data set for further analysis. Such a large-scale data set was very difficult to analyze. So we randomly selected 1,000 users' posts (352,516 posts) from the post set and set them as our data set for automatic evaluation. We also utilized an incremental strategy to test the reliability of evaluating using these 1,000 users. The data set (1,000 users) was expanded through five rounds, wherein 100 other users were added to the data set in each round. During the expansion, we observed little change in the evaluation results, which indicated a convergence of the results based on the 1,000 users. So we thought 1,000 users were sufficient for our evaluation and utilized them as data set in the end. For each of the 1,000 users, the posts were randomly split into a training set, a validation set and a test set according to a proportion of 0.8:0.1:0.1.

For the C2E and E2C Transformers, we utilized the Transformer base (8 attention heads, 512 dimensional hidden state, and 2048 dimensional feed-forward state) [43] and set the input length and output length as 300, word embedding dimension as 768, batch size as 4,096 tokens [57], dropput as 0.3, learning rate as 0.0015, epoch number as 50, optimizer as Adam (β_1 =0.9, and β_2 =0.98) [22], and the size for beam search as 4. For the C2C Transformer, the epoch size was set as 25 (10 in model pre-training and 15 in model tuning), and the other hyper-parameters were the same as those of the C2E and E2C Transformers.

5.1.2 Preliminary Analysis. The core of our work is to transfer the linguistic style of general transition results to the style of one's social media UGTC. The basic requirement of that is the general transition result and the post that correspond to the same semantics should have different styles. So we first analyzed how general translation results and the corresponding Weibo posts differed in linguistic style. In our model, the module for personalized linguistic style generation (C2C Transformer) is obtained by being pre-trained on all users' posts and tuned on a user's own posts. So

we conducted analysis from both the perspective of all users (platform-based analysis) and the perspective of a specific user (user-based analysis). The platform-based analysis aimed to explore the style difference between general translation results and the corresponding Weibo posts by regarding all users as a whole, while the user-based analysis focused on the fine-grained difference, i.e., the style difference between general translation results and a specific user's posts. The model for user-based analysis was also utilized to evaluate the personalization degree of the translation results in Section 5.1.3.

Platform-based analysis:

- Data preparation. This analysis aimed to explore the linguistic style difference between general translation results and corresponding Weibo posts by regarding all users as a whole. In the analysis, the semantic difference needed to be controlled in order to observe the linguistic style difference. So it required many pairs of parallel samples, where each pair consisted of a Weibo post and a general translation result that were characterized with the same semantics. To obtain such a data set, we input the 352,516 posts (Weibo corpus) into the C2E Transformer and E2C Transformer to obtain the corresponding general translation results (general translation corpus). This processing aims to make the Weibo posts and general translation results to correspond to the same semantics. So the Weibo corpus and general translation corpus were set as the data set for platform-based analysis.
- Model. We built a classification model to test whether general translation results and Weibo posts differed in linguistic style. The model was based on a well-known deep learning method BiLSTM (Bidirectional Long Short-Term-Memory) [37]. We chose this method for the following two reasons. First, BiLSTM has the capability to capture long-range dependency from sequential data and prevent gradients from vanishing [2]. Second, it encodes each token from two directions, which can consider past and future contextual information and improve the performance of text modeling. As these advantages, it has been a prevalent method for social media UGTC modeling and utilized in diverse kinds of tasks like post classification [39], sentiment analysis [2] and event detection [37]. For the model training, we tuned parameters on the validation dataset and optimized them via a small grid search. We set the vocabulary size as 20,000 and word embedding dimension as 128. Due to the large size of the data set, we set the mini-batch size as 1024, epoch size as 5, loss function as binary cross-entropy and optimizer as Adam.
- **Result.** After training and predicting, we obtained a higher classification accuracy (84.59%) on the test set. It suggests that most general translation results and the corresponding Weibo posts can be classified correctly.

User-based analysis:

- Data preparation. In order to evaluate the style difference between general translation results and a specific user's posts, we need to construct a corpus from the perspective of users. Similar to the data set preparation of platform-based analysis, for a given user, her/his Weibo posts were set as a corpus (user post corpus), and then the posts were input into the C2E Transformer and E2C Transformer to obtain the corresponding general translation results (general translation corpus). The user post corpus and general translation corpus were set as the data set for user-based model training and evaluation.
- Model. Similar to the platform-based analysis, we built a classification model for each user based on BiLSTM. Compared with the platform-based model, a user-based model's training samples were fewer (just the current user's posts). To improve the model's accuracy, we

adopted the pre-training and tuning strategy. For a given user, her/his corresponding userbased model was built by tuning the platform-based model using her/his post corpus and general translation corpus.

• **Result.** The CDF (Cumulative Distribution Function) of classification accuracy on the test set among the 1,000 users is shown in Figure 5. From the results, we can see most users obtain higher accuracy scores. For example, more than 95% users' accuracy values are higher than 0.8, and more than 81% users' accuracy values are higher than 0.85. The result suggests that most Weibo users' posts are significantly different from the corresponding general translation results.



Fig. 5. The CDF of user's classification accuracy in terms of linguistic style.

Above all, we can see no matter from the perspective of regarding users as a whole or from the perspective of an individual user, most Weibo posts and corresponding general translation results can be classified correctly. In the classification tasks, other factors like translation errors might be utilized as features. However, they appear less frequently according to the good performance of BLEU and perplexity scores (see Section 5.1.3) and human evaluation. So we thought for most Weibo posts and the corresponding general translation results that are classified correctly, it is very likely that the posts and corresponding translation results are characterized by the different linguistic styles, which lays the foundation for building user-oriented personalized machine translator based on Weibo posts.

5.1.3 Automatic Evaluation Metrics and Results.

Metrics. We evaluated the personalized translation model by the common-used metrics for personalized language model evaluation and machine translation model evaluation. These metrics include:

• Personalization degree [30]. It measures how closely the generated translation results correspond to a particular user's linguistic style. For each test sample of a given user, we constructed a test list containing ten translation results, wherein one was the corresponding translation result (gold result) generated by our user-oriented personalized machine translation model, one was the general translation result, and the other eight results were randomly selected from the other users' translation results (four pairs of personalized translation result and general translation result). We input the ten results into the user-based classification model to see if the gold result could be predicted as the user's post. We expected that the gold result could get the highest likelihood of belonging to the user among the ten translation results, i.e., rank number one among the test list. To obtain a cumulative measurement among all test samples, we followed previous work [30] and set User Matching Accuracy (UMA) and Mean Reciprocal Rank (MRR) as metrics. UMA means the proportion where the gold result is ranked highest, and MRR is defined as:

$$MRR = \frac{1}{|T|} \sum_{t=1}^{|T|} \frac{1}{rank_t},$$
(1)

where *T* means the test set, and $rank_t$ means the rank of the gold result among the test list for the *t*th test sample.

- BLEU [60]. This metric reflects the accuracy of translation. It counts how many n-grams (n=1,2) in the generated translation result overlap with that in the reference (the original post in our work).
- Perplexity [23, 60]. It measures how a model fits the test data. In traditional language systems, a lower perplexity score indicates a model's higher certainty to fit the test data.

Result. From the perspective of personalization degree, we obtained UMA=0.45, and MRR=0.21 for personalized translation results, while UMA=0.15, and MRR=0.001 for general translation results. It suggests that the linguistic styles of translation results generated by our user-oriented personalized machine translation model are more consistent with each user's style preference. From the perspective of BLEU, the value (27.33 of personalized translation results is slightly higher than that (23.27 of general translation results. These two results jointly indicate that our model's translation results are more consistent with the Weibo posts. We exhibit some examples of the translation results in Table 1 (E1 - E5, and the differences between each general translation result and the corresponding personalized translation result are underlined. First, we can see the vocabularies (e.g. "大家", "别", "跟", "阿姨", "通常" and syntactic structures (e.g. "按时催你" of the personalized translation results are more consistent with the Weibo posts and more natural than the corresponding vocabularies ("每个人", "不要", "和", "大妈", "一般" and syntactic structures ("督促你按时" of the general translation results. Second, there are some complicated vocabularies (e.g. "督促" in the general translation results, while the personalized translation results utilize the users' common-used vocabularies ("催" to express the same semantics. However, from the perspective of perplexity, the value (150.61 of personalized translation results is slightly higher than that (144.45 of general translation results, which suggests that our model's certainty to fit the test data is slightly lower than the general machine translation model. This finding is consistent with the result of personalization degree. Compared with the general translation model, our model generates diversified translations carrying rich persona-related features, which hurts the model's certainty to fit the test data.

Moreover, we also observed that the C2E Transformer played an important role in our model, i.e., the quality of a C2E result had a significant effect on the quality of translation results especially the general translation result. For example, for the examples E6, E7 and E8 in Table 1, there are some translation errors or inaccuracies in the C2E results, e.g. "介样" (an Internet slang with the meaning of "这样" in Chinese and "such" in English, "从未来过" and "完毕" are incorrectly translated as "a sample", "in the future" and "after" respectively, which results in the corresponding incorrect or inaccurate translations in the general translation results and personalized translation results ("样本", "未来" and "之后"/"后". However, we found the personalized translation results were less likely to be affected compared with general translation results. For example, in E9, E10 and E11, the C2E Transformer incorrectly or inaccurately translates "肿莫" (an Internet slang with the meaning of "怎么" in Chinese and "what's wrong" in English, "国脖" (an Internet slang with the meaning of "微博" in Chinese and "Weibo" in English) and "觉得" into "swollen", "scarf" and



Fig. 6. Evolution of performance in terms of number of posts.

"think" respectively, and the general translation results thus suffer from the corresponding incorrect or inaccurate vocabularies "肿", "围巾" and "看", while the personalized translation results obtain the accurate translations "肿么", "围脖" and "觉得". In our model, personalized translation results are obtained by tuning the linguistic styles of general translation results. Based on these experimental results, we can see such a tuning procedure can not only make the translation results more personalized but also revise some incorrect and inaccurate vocabularies of general translation results.

The above experimental results indicate that our model can generate more personalized translation results by learning users' linguistic style preferences from their Weibo posts. By referring to previous research [6, 29], we required that each user's Weibo posts should be more than a threshold - 50 in these experiments. We further analyzed how the model's performance changed with different threshold settings. We made 5 groups of experiments by varying the baseline of the number of posts from 10 to 50 (10, 20, 30, 40, 50). For example, when setting the baseline as 10, 10 posts were randomly selected from each user's Weibo posts as the training corpus. The results are exhibited in Figure 6. From the results, we can see UMA, MRR and BLEU exhibit minor changes with the increase of the number of posts. However, perplexity experiences a prominent decline when the number of posts increases from 10 to 30, while reaches a convergence in the following intervals. These results jointly indicate that our personalized machine translation model does not need a large number of training posts for each user. The baseline is suggested to be configured as 30 or higher as UMA, MRR, BLEU and perplexity all reach convergence on the setting. As illustrated in Section 3.2, we train the personalized machine translation model (C2C Transformer) by pre-training it on all users' posts and tuning it on each user's posts. Pre-training and tuning has been proved to be an effective strategy to process prediction tasks with insufficient labeled data, which results in that our model's training does not depend on a large number of posts.

5.2 Human Evaluation

5.2.1 Setting. We made human evaluation by pairwise comparison [24, 30]. The main settings are described below.

Participants. We recruited participants by posting recruitment to WeChat. The participants should meet the following requirements. First, each participant should have a certain activity degree in Weibo (had published no less than 50 posts) so that the user-oriented personalized machine translation model could have a corpus to learn her/his linguistic style. Second, to evaluate the translation results, the participants should have basic proficiency in English. We required that they had passed the Band Four of College English Test. Finally, we recruited 31 participants (14 male users

Example #	Post	C2E result	General translation	Personalized trans-
			result	lation result
E1	"大家的晚间生活	Everyone's evening life	<u>每个人</u> 的晚上生	<u>大家</u> 晚上的生活
	太健康了"	is too healthy	活都太健康了	太健康了
E2	"如果你不相信爱	If you don't believe in	如果你不相信爱	如果你不相信爱
	情,就别奢望爱情。	love, don't expect it.	情, 就 <u>不要</u> 期待爱	情,就 <u>别</u> 期待爱情,
_	晚安!"	Good night!	情。晚安!	晚安!
E3	"你是需要一个按	Do you need someone	你需要一个 <u>督促</u>	你需要一个 <u>按时</u>
	时催你睡 <u>见</u> 的人,	who urges you to sleep	<u>你按时</u> 睡见的人,	<u>催你</u> 睡见的人, 也 本 西 会 应 你
	业定一个愿息简	who is willing to stay	业定一个愿息简	定一个愿息喧你 教友的人?
		up late with you?	小然仅的八	
E4	"跟一个怀柔本地	I chatted with a local	和一位怀柔本地	跟一个怀柔本地
	的阿姨聊了两句,	aunt in Huairou and	<u>大妈</u> 聊天后,发现	<u>阿姨</u> 聊完后,发现
	发现这边已经是	found that there is al-	这边已经是东北	这边已经东北口
	东北口音了"	ready a northeast ac-	口音了	音了
7-		cent here		
E5	有 唯 知 追 原 因	Who knows why?	准知道 <u>为什么</u> ?这	谁知追 <u>原因</u> ?这时 6
	呢:这个时候一叔 郑右答安的!"	answers at this time!	个 时 恢 <u>週 吊</u> 郁 有	恢 <u>一成</u> 仰月谷禾!
F6	"扫扫盲 百束玄介	To eliminate illiteracy	百禾· 打百 百来早—个	」 「「「」」 「」 「」 「」 「」 「」 「」 「」 「」 「」 「」 「」
LU	适应。"	it turned out to be a	11日, 小 小 疋 1 样本。	11日,
		sample.	<u></u>	
E7	"人们说我精明事	People say I'm shrewd,	人们说我精明, 但	人说我精明, 可是
	故,可我确 <u>从未来</u>	but I've never been in	我从来没有活过	我从来没有活过
	<u>过</u> 人间"	the future	<u>未来</u>	<u>未来</u>
E8	"酒会 <u>完毕</u> ,还是挺	$\underline{\text{After}}$ the reception, it	酒会 <u>之后</u> 还是挺	酒会 <u>后</u> 还是蛮热
	热闹的"	was quite lively	热闹的	闹的
E9	"我亲爱的小鬼,乃	My dear little devil, it's	我亲爱的孩子, <u>肿</u>	我亲爱的小孩,你
D40	<u> 肥臭</u> ∫…″ " A エロ L わ ま チ	swollen	$\int \dots$	<u> 胛么</u> ∫
E10	"今大早上起米看	What I saw most when	找今大早上起米	今早起米看 <u>围脖</u> , 毛列星名的前星
	<u> 固</u> 府	I got up and looked at	定力」 有 <u><u></u><u></u><u></u>同</u> <u></u> 。有 到 <u></u> 是 夕 <u></u> <u></u> <u></u> <u></u> <u></u> <u></u> <u></u>	有到取多的机定 吃晚很夕人生呢
	机定叶吮似多八 生眠还有计具系	was that many people	到取夕时定叶吮 很名人生眠 还右	叶 ····································
	布斯的新闻"	suffered from insomnia	成少八八帆, <u></u> 无有斯的新闻	<u></u> 宜发布了
		last night and the news		
		about Jobs		
E11	"我怎么 <u>觉得</u> 联通	How do I t <u>hink</u> Unicom	我怎么看联通光	我怎么 <u>觉得</u> 联通
	光纤比电信光纤	optical fiber is more	纤比电信光纤更	光纤比电信光纤
	还给力啊!"	powerful than telecom	厉害?	还厉害?
		optical fiber?		

Table 1. Examples of translation results.

and 17 female users, Weibo post size ranged from 50 to 5,302, Weibo tenure ranged from 1 to 11 year(sthat met these requirements. Similar to the automatic evaluation, we also utilized an incremental strategy to test the reliability of evaluating by these 31 users. Other five participants were recruited to participate in the evaluation one by one, while the results experienced little change, which indicated a convergence trend. So we thought the results based on the 31 participants were representative and convincing.

Translation corpus. We utilized the Corpus of Contemporary American English (COCA)¹⁰ as the source texts. The reasons are two-fold. First, COCA contains 8 genres including spoken, fiction, popular magazines, newspapers, academic texts, TV and Movies subtitles, blogs, and other web pages. It supplies us with diverse kinds of English genres to evaluate our user-oriented personalized machine translator, which can promote the generalizability and reliability of our evaluation. Second, the corpus contains more than one billion words that were collected from 1990 to 2019. Such a large data set provides sufficient texts for our evaluation.

Evaluation procedure. For each participant, we utilized 50 samples (source texts) for evaluation, wherein 40 samples were randomly selected from COCA, and 10 samples were input by herself/himself during evaluation. In the evaluation, the translation results corresponding to the 50 samples were presented to the participant through 50 rounds, i.e., each round presented one sample's translation results. The translation results contained the personalized translation result that was generated by our model as well as the general translation result. The participant did not know which one was the personalized result and which one was the general one, and positions where the two translation results were presented on the screen were randomized. The participant needed to give feedback regarding the following metrics and questions.

- Comprehensibility: the degree to which the translation result can be comprehended easily [59, 60]. The major goal of building the user-oriented personalized machine translator is to make the translation results easily comprehended by each user. So we first considered comprehensibility as a metric in human evaluation.
- Linguistic style consistency: whether the translation result is consistent with the user's linguistic style [60]. The core of our model is to generate a linguistic style that can be consistent with each user's preference. So we also considered linguistic style consistency as a metric to get a user's feedback about whether the translation result was consistent with her/his linguistic style.

The two metrics were judged on a 5-point scale (1,2,3,4,5), wherein 1 means the worst of a metric, and 5 means the highest. Besides these two metrics, we also presented two other questions to the participant in each round: preferred translation result, and the reasons. These two questions aimed to explore the strengths and weaknesses of the personalized translation results generated by our model compared with general translation results. The final design of the pairwise comparison page is shown in Figure 7. This page is an extension of the page shown in Figure 3.

5.2.2 Result. First, we compared personalized translation results and general translation results in terms of the two metrics - comprehensibility and linguistic style consistency. For personalized translation results, the mean values of comprehensibility and linguistic style consistency are 4.11 and 3.26 respectively, and for general translation results, the mean values of comprehensibility and linguistic style consistency are 2.96 and 2.17 respectively. We also made a *t*-test for each of the two metrics to see if the difference of the metric between personalized translation results and general translation results was significant and obtained significant results (p<0.01) in both the two tests. The above results jointly suggest that the personalized translation results have better comprehensibility and linguistic style consistency versus general translation results.

Second, we analyzed users' preferred translation results. Among all responses, personalized translation results, general translation results and neutral options account for 63.33%, 29.29% and 7.38% respectively, which indicates our model's superiority versus general machine translation systems. We also analyzed the difference in terms of the eight COCA kinds. We found that the proportion value of personalized translation results was slightly lower on two kinds - spoken and

¹⁰https://www.english-corpora.org/coca/



Fig. 7. Screenshot of the web page for human evaluation.

web pages, higher on the academic kind, and moderate on the other kinds. Such a result suggests that for professional languages that are essentially difficult to comprehend, people's personalized demand is stronger. This finding highlights the more meaningful application scenarios of our user-oriented personalized machine translation model.

Third, we analyzed the reasons provided by users for choosing or not choosing personalized translation results in evaluation to uncover the strengths and weaknesses respectively of personalized translation results.

Strengths of personalized translation results. The strengths of personalized translation results reflected in users' responses are summarized below.

- Better comprehensibility. Many responses described that personalized translation results were easier to comprehend than general translation results, e.g. "the translation result A (personalized translation result) is conceived in simple language", "the translation result B (personalized translation result) contains less professional words, and it is easier to comprehend", and "the translation result B (general translation result) is more formal than the translation result A (personalized translation result), so the translation result A is easier to comprehend".
- More natural. Another important strength of personalized translation results is they are more natural than general translation results. For example, some users responded that "the translation result A (personalized translation result) is characterized by a higher degree of spoken language", "the translation result B (personalized translation result) fits snugly into daily expression", and "there is translationese in the translation result A (general translation result), and the translation result B (personalized translation result) is more natural".
- More prominent linguistic style. The responses also indicate the prominent linguistic style existing in personalized translation results, which is consistent with our automatic evaluation result. Such responses include "the translation result A (personalized translation result) is more like my translation", "the linguistic style of translation result B (personalized translation result) is more evident, and it seems like Weibo's style", "the linguistic style of translation result A (personalized translation result A (personalized translation result) is more like translation result) is more like Chinese language", etc.
- More lively. Some people responded that personalized translation results were more interesting, emotional or vivid than general translation results. For example, they described "*although the translation result A (general translation result) is more accurate, the translation result B (personalized translation result) is pretty lovely*", "*the translation result A (personalized*

translation result) has strong tone and emotion", and "the translation result B (personalized translation result) is more vivid".

Weaknesses of personalized translation results. We analyzed the weaknesses of personalized translation results reflected in users' responses. The results suggested that lack of fluency and lack of accuracy were the major problems of some personalized translation results. These two problems occurred mainly in the academic, newspaper and magazine corpus while less commonly in the other kinds like spoken and fiction.

- Lack of fluency. Lack of fluency is the most prominent weakness mentioned by users. In some test cases, users responded that "the translation result B (general translation result) is more fluent", "the translation result B (personalized translation result) lacks of coherence", "the translation result A (personalized translation result) contains grammatical issues", etc.
- Lack of accuracy. Another weakness of personalized translation results suggested by users' responses is lack of accuracy, including "compared to the translation result A (personalized translation result), the translation result B (general translation result) reveals the real meaning of the source text", "the translation result A (personalized translation result) is too concise to express the accurate meaning", etc.

Above all, the main findings of automatic evaluation and human evaluation are summarized as follows. From the perspective of personalization degree, automatic evaluation and human evaluation all validate our model's good capability to generate personalized translation results. From the perspective of perplexity, our model can generate diversified translations carrying rich linguistic features, which makes the translations more natural and lively. These benefits enhance the comprehensibility of translation results. From the perspectives of fluency and accuracy, the results suggest that performing linguistic style personalization can sometimes hurt the translation fluency and accuracy for some professional kinds of texts (mainly academic, newspaper and magazine), while the other kinds like spoken, fiction and web pages are affected less. However, personalized translation results are more preferred by users in human evaluation, which highlights our model's overall superiority versus existing general translation systems.

6 **DISCUSSION**

6.1 Implications

In this section, we discuss the implications of our work on the research and design of machine translation systems.

Most of nowadays machine translation models and systems focus on promoting objective indicators like translation accuracy, fluency and speed, while few researchers and designers optimize machine translation from the view of users, and users' preferences and perspectives to translation are ignored. First, as suggested by our work, our user-oriented personalized machine translation model can generate more personalized, natural and lively translation results and enhance the comprehensibility of translation results, which makes its translations more preferred by users versus general translations. So the future design of machine translation models and systems should consider linguistic style as an important benchmark. Moreover, in reality, linguistic style is just one of people's diverse traits. There might be other user traits that can affect users' experience in translation system use. For example, people use a machine translator for different purposes like informal communication and formal expression. By considering these different purposes, it is helpful for machine translation systems to generate appropriate translations and gain users' confidence. So in the future study, many other user traits that can affect users' experience in translation system use should be systematically investigated and incorporated into machine translation systems to further improve user experience, which reflects the thoughts of human-centered computing and humanlike language systems [12]. Second, in our work, we find there needs a trade-off between some object indicators and the linguistic style consistency in the design of machine translation models and systems. Exhibiting too many personal linguistic style features can hurt translation fluency, accuracy and perplexity in our evaluation. Machine translation models are generally trained based on a formal corpus or fine-tuned corpus, and the most common-used corpus is news¹¹, which enhances fluency, accuracy and perplexity of translation results. When incorporating users' linguistic styles into translation models, it can bring some personal linguistic features like idiomatic expressions, abbreviations and rare words into translation results, which will affect the object indicators like translation fluency, a ccuracy and perplexity. Moreover, there can be inaccurate or incorrect expressions in some people's languages, which also has a negative effect on the fluency, accuracy, perplexity and other object indicators of personalized translation results. Thus the future design of machine translation models and systems should not consider object indicators and user-related indicators separately, and there needs a balance (e.g. using multi-objective optimization methods [10] between some object indicators and some user-related indicators to ensure the systems can obtain satisfactory performance on both object indicators and user experience.

6.2 Limitations

As an exploratory study focusing on the user-oriented personalized machine translation, this research suffers from some limitations in data preparation and model building and evaluation.

From the perspective of data preparation, we aim to generate personalized machine translations for each user using her/his UGTC in social media. Machine translations are users' consumed content, while UGTC is their produced content. Although previous studies [24, 25, 58, 60] have explored the feasibility of generating consumed content using produced content, building a consumption model based on produced content might have some gaps. Even so, produced content is more feasible than consumed content for the user-oriented personalized machine translation task. First, consumed content like browsing history in a machine translation system or other systems is more difficult to obtain as people don't disclose the content publicly, which limits the feasibility of sampling sufficient consumed content to train a user-oriented personalized machine translation model. Second, people consume a large amount of content every day, and a person's preferred linguistic style is embedded in some of her/his consumed content. So it is difficult to distinguish a user's preferred linguistic style from the others. These challenges limit the feasibility of building a user-oriented personalized machine translation model based on users' consumed content. Furthermore, as the difficulty of getting persona data, we follow the back-translation strategy and propose an approach for training corpus preparation by utilizing the C2E Transformer, which results in that the quality of the personalized translation results is dependent on the C2E module, and the problems of C2E results can be transmitted to the following general machine translations and personalized machine translations. On one hand, as suggested in our evaluation, personalized translation results are less likely to be affected compared with general translation results. On another hand, we try to alleviate this problem by utilizing state-of-the-art Chinese-to-English translation techniques to improve the performance of the C2E module. Many new promising machine translation approaches are continually emerging, and the C2E module in our model can be easily upgraded to these approaches without adjusting the model's architecture, which can further improve the performance of the C2E module and alleviate this drawback in the future.

From the perspective of model building and evaluation, a person might have multiple linguistic styles in different contexts, and there may be many corpora that can reflect his/her styles. This

¹¹https://chinesenlp.xyz/#/docs/machine_translation

work just uses the UGTC of social media context as a corpus to learn each person's linguistic style. However, our model has good generalizability. If another corpus is utilized to learn people's linguistic styles, just the training set should be changed, while the architecture of our model does not need to be adjusted. For the model's implementation, as the difficulty of data sampling, only one kind of translation task (English-to-Chinese) is presented in this paper, which might limit the generalizability of our model's performance. On the one hand, as mentioned above, the architecture of our model can be applied to all kinds of translation tasks, and the performance on the other cross-language translation tasks can be improved by just tuning the parameters. On the other hand, in future research, we will utilize the model to handle other cross-language translation tasks to further validate its generalizability and reliability. For the model's evaluation, BLEU is set as one significant metric in our automatic evaluation, while previous research [4] has suggested that the improved BLEU score cannot always reflect the improvement of translation quality since it allows a tremendous amount of variation. However, BLEU is still a solid metric to track broad and incremental changes to a machine translation system. We also draw our conclusions based on automatic and human evaluations, making our findings more convincing.

7 CONCLUSION

In this paper, we constructed a user-oriented personalized machine translation model based on social media UGTC and several state-of-the-art deep learning techniques. To show the model's usage, we designed and implemented a user-oriented personalized machine translator supported by the Weibo Open Platform. The translator provides researchers and developers with a systematical implementation scheme of a user-oriented personalized machine translation system based on our model. Automatic evaluation and human evaluation suggest that our model can generate more personalized, natural and lively translation results and enhance the comprehensibility of translation results, which makes it more preferred by users versus general translation systems. For the limitations of the current model, we will continue to study the model in the future by incorporating more user traits into it and validating its performance in other cross-language translation tasks.

8 ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (NSFC) under the Grants nos. 61902075, 61932007, and 62172106.

REFERENCES

- [1] Kholoud Khalil Aldous, Jisun An, and Bernard J Jansen. 2019. View, like, comment, post: Analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 47–57.
- [2] Georgios Balikas, Simon Moura, and Massih-Reza Amini. 2017. Multitask learning for fine-grained twitter sentiment analysis. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Rretrieval. 1005–1008.
- [3] Deana Brown and Rebecca E Grinter. 2016. Designing for transient use: A human-in-the-loop translation platform for refugees. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 321–330.
- [4] Chris Callison-Burch, Miles Osborne, and Philipp Koehn. 2006. Re-evaluating the role of BLEU in machine translation research. In Proceedings of the 11th Conference of the European Chapter of the Association for Computional Linguistics. 249–256.
- [5] Guanhua Chen, Yun Chen, and Victor OK Li. 2021. Lexically constrained neural machine translation with explicit alignment guidance. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35. 12630–12638.
- [6] Xuetong Chen, Martin D Sykora, Thomas W Jackson, and Suzanne Elayan. 2018. What about mood swings: Identifying depression on Twitter with temporal measures of emotions. In *Companion Proceedings of the The Web Conference 2018*. 1653–1660.

- [7] Yu-Hsiu Chen, Pin-Yu Chen, Hong-Han Shuai, and Wen-Chih Peng. 2020. TemPEST: Soft template-based personalized EDM subject generation through collaborative summarization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 7538–7545.
- [8] Na Cheng, Rajarathnam Chandramouli, and KP Subbalakshmi. 2011. Author gender identification from text. Digital Investigation 8, 1 (2011), 78–88.
- [9] Ning Dai, Jianze Liang, Xipeng Qiu, and Xuan-Jing Huang. 2019. Style Transformer: Unpaired text style transfer without disentangled latent representation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 5997–6007.
- [10] Kalyanmoy Deb. 2014. Multi-objective optimization. In Search Methodologies. Springer, 403-449.
- [11] Sergey Edunov, Myle Ott, Michael Auli, and David Grangier. 2018. Understanding back-translation at scale. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 489–500.
- [12] Lucie Flek. 2020. Returning the N to NLP: Towards contextually personalized classification models. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 7828–7838.
- [13] Ge Gao, Bin Xu, Dan Cosley, and Susan R Fussell. 2014. How beliefs about the presence of machine translation impact multilingual collaborations. In Proceedings of the 17th ACM conference on Computer Supported Cooperative Work & Social Computing. 1549–1560.
- [14] Ge Gao, Bin Xu, David C Hau, Zheng Yao, Dan Cosley, and Susan R Fussell. 2015. Two is better than one: improving multilingual collaboration by giving two machine translation outputs. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. 852–863.
- [15] Spence Green, Jeffrey Heer, and Christopher D Manning. 2013. The efficacy of human post-editing for language translation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 439–448.
- [16] Kotaro Hara and Shamsi T Iqbal. 2015. Effect of machine translation in interlingual conversation: Lessons from a formative study. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 3473– 3482.
- [17] Wei He, Zhongjun He, Hua Wu, and Haifeng Wang. 2016. Improved neural machine translation with SMT features. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 30.
- [18] Nico Herbig, Santanu Pal, Josef van Genabith, and Antonio Krüger. 2019. Multi-modal approaches for post-editing machine translation. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 1–11.
- [19] Chang Hu, Benjamin B Bederson, Philip Resnik, and Yakov Kronrod. 2011. Monotrans2: A new human computation system to support monolingual translation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing* Systems. 1133–1136.
- [20] Harsh Jhamtani, Varun Gangal, Eduard Hovy, and Eric Nyberg. 2017. Shakespearizing modern language using copyenriched sequence-to-sequence models. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing 6 (2017), 10.
- [21] Osama Khalid and Padmini Srinivasan. 2020. Style matters! Investigating linguistic style in online communities. In Proceedings of the International AAAI Conference on Web and Social Media, Vol. 14. 360–369.
- [22] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
- [23] Satwik Kottur, Xiaoyu Wang, and Vitor R Carvalho. 2017. Exploring personalized neural conversational models. In Proceedings of the 26th International Joint Conference on Artificial Intelligence. 3728–3734.
- [24] Jiwei Li, Michel Galley, Chris Brockett, Georgios P Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155* (2016).
- [25] Junjie Li, Haoran Li, and Chengqing Zong. 2019. Towards personalized review summarization via user-aware sequence network. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33. 6690–6697.
- [26] Pan Li and Alexander Tuzhilin. 2019. Towards controllable and personalized review generation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 3228–3236.
- [27] Hajin Lim, Dan Cosley, and Susan R Fussell. 2018. Beyond translation: Design and evaluation of an emotional and contextual knowledge interface for foreign language social media posts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [28] Hajin Lim and Susan R Fussell. 2017. Making sense of foreign language posts in social media. Proceedings of the ACM on Human-Computer Interaction 1, CSCW (2017), 1–16.
- [29] Tianyi Lin, Wentao Tian, Qiaozhu Mei, and Hong Cheng. 2014. The dual-sparse topic model: Mining focused topics and focused terms in short text. In Proceedings of the 23rd International Conference on World Wide Web. 539–550.
- [30] Bodhisattwa Prasad Majumder, Shuyang Li, Jianmo Ni, and Julian McAuley. 2019. Generating personalized recipes from historical user preferences. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 5978–5984.

- [31] Paul Michel and Graham Neubig. 2018. Extreme adaptation for personalized neural machine translation. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 312–318.
- [32] Shachar Mirkin, Scott Nowson, Caroline Brun, and Julien Perez. 2015. Motivating personality-aware machine translation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 1102–1108.
- [33] Daisuke Oba, Naoki Yoshinaga, Shoetsu Sato, Satoshi Akasaki, and Masashi Toyoda. 2019. Modeling personal biases in language use by inducing personalized word embeddings. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2102–2108.
- [34] Thierry Poibeau. 2017. Machine Translation. MIT Press.
- [35] Ella Rabinovich, Raj Nath Patel, Shachar Mirkin, Lucia Specia, and Shuly Wintner. 2017. Personalized machine translation: Preserving original author traits. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. 1074–1084.
- [36] Sudha Rao and Joel Tetreault. 2018. Dear sir or madam, may I introduce the GYAFC dataset: Corpus, benchmarks and metrics for formality style transfer. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). 129–140.
- [37] Ramit Sawhney, Harshit Joshi, Saumya Gandhi, and Rajiv Ratn Shah. 2021. Towards ordinal suicide ideation detection on social media. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining. 22–30.
- [38] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving neural machine translation models with monolingual data. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 86–96.
- [39] Abhishek Singh, Eduardo Blanco, and Wei Jin. 2019. Incorporating emoji descriptions improves tweet classification. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2096–2101.
- [40] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In Proceedings of the 2014 Annual Conference on Neural Information Processing Systems. 3104–3112.
- [41] Alexandru Tatar, Marcelo Dias De Amorim, Serge Fdida, and Panayotis Antoniadis. 2014. A survey on predicting the popularity of web content. *Journal of Internet Services and Applications* 5, 1 (2014), 8.
- [42] Hans Van Halteren, Harald Baayen, Fiona Tweedie, Marco Haverkort, and Anneke Neijt. 2005. New machine learning methods demonstrate the existence of a human stylome. *Journal of Quantitative Linguistics* 12, 1 (2005), 65–77.
- [43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems. 6000–6010.
- [44] Hao-Chuan Wang, Susan Fussell, and Dan Cosley. 2013. Machine translation vs. common language: Effects on idea exchange in cross-lingual groups. In Proceedings of the 2013 Conference on Computer Supported Cooperative Work. 935–944.
- [45] Qiang Wang, Bei Li, Tong Xiao, Jingbo Zhu, Changliang Li, Derek F Wong, and Lidia S Chao. 2019. Learning deep Transformer models for machine translation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 1810–1822.
- [46] Wen Wang, Wei Zhang, Jun Wang, Junchi Yan, and Hongyuan Zha. 2018. Learning sequential correlation for user generated textual content popularity prediction. In Proceedings of the 27th International Joint Conference on Artificial Intelligence. 1625–1631.
- [47] Xing Wang, Zhengdong Lu, Zhaopeng Tu, Hang Li, Deyi Xiong, and Min Zhang. 2017. Neural machine translation advised by statistical machine translation. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 31.
- [48] Yue Wang, Cuong Hoang, and Marcello Federico. 2021. Towards modeling the style of translators in neural machine translation. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 1193–1199.
- [49] Yunli Wang, Yu Wu, Lili Mou, Zhoujun Li, and Wenhan Chao. 2020. Formality style transfer with shared latent space. In Proceedings of the 28th International Conference on Computational Linguistics. 2236–2249.
- [50] Kyle Wiggers. 2020. How Google is using emerging AI techniques to improve language translation quality. https://venturebeat.com/2020/06/03/how-googleis-using-emerging-ai-techniques-to-improve-languagetranslation-quality/
- [51] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144 (2016).
- [52] Bin Xu, Ge Gao, Susan R Fussell, and Dan Cosley. 2014. Improving machine translation by showing two outputs. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 3743–3746.

- [53] Naomi Yamashita, Rieko Inaba, Hideaki Kuzuoka, and Toru Ishida. 2009. Difficulties in establishing common ground in multiparty groups using machine translation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 679–688.
- [54] Naomi Yamashita and Toru Ishida. 2006. Effects of machine translation on collaborative work. In Proceedings of the 2006 Conference on Computer Supported Cooperative Work. 515–524.
- [55] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 1480–1489.
- [56] Wenhuan Zeng, Abulikemu Abuduweili, Lei Li, and Pengcheng Yang. 2019. Automatic generation of personalized comment based on user profile. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. 229–235.
- [57] Tianfu Zhang, Heyan Huang, Chong Feng, and Longbing Cao. 2021. Self-supervised bilingual syntactic alignment for neural machine translation. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35. 14454–14462.
- [58] Wei-Nan Zhang, Qingfu Zhu, Yifa Wang, Yanyan Zhao, and Ting Liu. 2019. Neural personalized response generation as domain adaptation. World Wide Web 22, 4 (2019), 1427–1446.
- [59] Yinhe Zheng, Guanyi Chen, Minlie Huang, Song Liu, and Xuan Zhu. 2019. Personalized dialogue generation with diversified traits. arXiv preprint arXiv:1901.09672 (2019).
- [60] Yinhe Zheng, Rongsheng Zhang, Minlie Huang, and Xiaoxi Mao. 2020. A pre-training based personalized dialogue generation model with persona-sparse data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 9693–9700.
- [61] Jianling Zhong, Weiwei Guo, Huiji Gao, and Bo Long. 2020. Personalized query suggestions. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 1645–1648.