Utilizing Longer Context than Speech Bubbles in Automated Manga Translation

Hiroto Kaino[†], Soichiro Sugihara[†], Tomoyuki Kajiwara[†], Takashi Ninomiya[†], Joshua Tanner[‡], Shonosuke Ishiwatari[‡]
[†]Graduate School of Science and Engineering, Ehime University, Japan
[‡]Mantra Inc, Japan
{kaino@ai., sugihara@ai., kajiwara@, ninomiya@}cs.ehime-u.ac.jp
{josh@, ishiwatari@}mantra.co.jp

Abstract

This paper focuses on improving the performance of machine translation for manga (Japanese-style comics). In manga machine translation, text consists of a sequence of speech bubbles and each speech bubble is translated individually. However, each speech bubble itself does not contain sufficient information for translation. Therefore, previous work has proposed methods to use contextual information, such as the previous speech bubble, speech bubbles within the same scene, and corresponding scene images. In this research, we propose two new approaches to capture broader contextual information. Our first approach involves scene-based translation that considers the previous scene. The second approach considers broader context information, including details about the work, author, and manga genre. Through our experiments, we confirm that each of our methods improves translation quality, with the combination of both methods achieving the highest quality. Additionally, detailed analysis reveals the effect of zero-anaphora resolution in translation, such as supplying missing subjects not mentioned within a scene, highlighting the usefulness of longer contextual information in manga machine translation.

Keywords: Machine Translation, Style Transfer, Manga

1. Introduction

In recent years, Japanese-style comics, known as manga, have become popular worldwide. However, unofficially translated pirated copies of manga are circulating overseas in large numbers, causing serious damage to manga publishers. According to a survey conducted by Shogakukan in 2021,[1] approximately five times more pirated copies of Japanese manga are circulating overseas than official versions, and some pirated copies have been viewed more than 100 million times. This issue is largely attributed to the lack of immediately available official translations. Therefore, there are high expectations for machine translation (MT) of manga (Hinami et al., 2021) to increase the speed of translation and publication.

As mentioned in previous work on manga MT (Hinami et al., 2021), translating speech bubbles is a difficult problem. Speech bubbles generally consist of a balloon shape containing the utterances of characters, and are used as a unit for translation, i.e., manga text is split into a sequence of speech bubbles and each speech bubble is translated individually. However, sentences that span over speech bubbles frequently present a problem. For example, in the left scene of Figure 1, the line "「あさがお」はお父さんの夢だからな ('Asagao' is my dream.)" is split into two speech bubbles,

Figure 1: An example of scenes where contextual information is necessary. ©Mitsuki Kuchitaka

"「あさがお」は" and "お父さんの夢だからな." Therefore, the last speech bubble in Figure 1 is difficult to translate properly by itself. Figure 1 also shows an example where contextual knowledge about named entities specific to the work is necessary, e.g., "「あさがお」('Asagao')" in Figure 1 is generally a name of flower, but in this case, it is a name of a rocket.

To address this problem, Hinami et al. (2021) proposed 2+2 translation, which uses the previous speech bubble as context, and scene-based translation, which translates all text in each scene (comic panel) together, confirming an improvement

[1] https://prtimes.jp/main/html/rd/p/000000004.000059295.html

LREC-COLING 2024, pages 17337–17342
20-25 May, 2024.

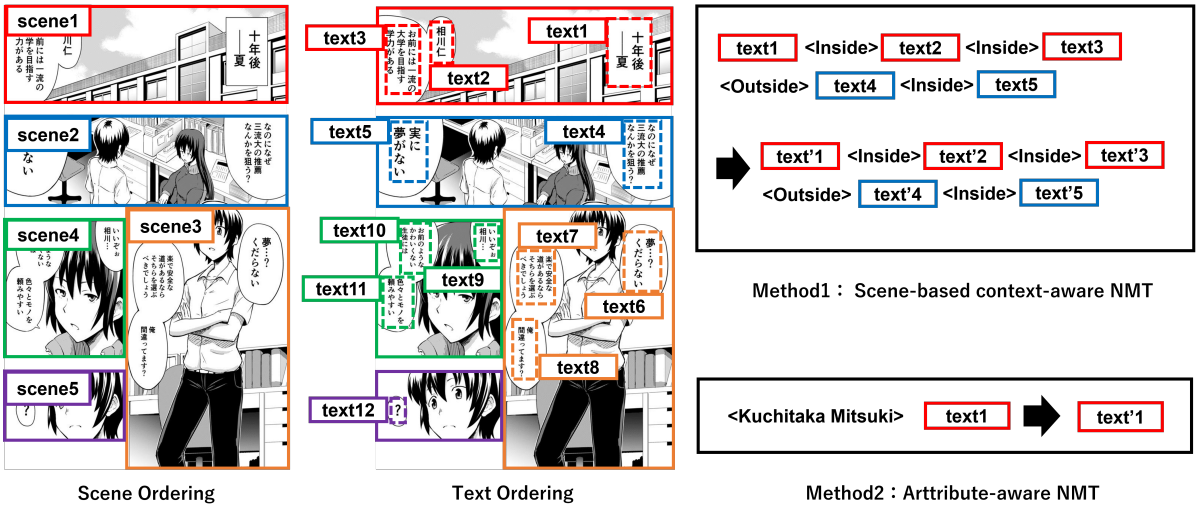

Figure 2: Proposed methods. N' represents the translation of a source sentence. ©Mitsuki Kuchitaka

in translation performance compared to the baseline, which translates each speech bubble independently. However, the contextual information used in existing work is confined to the scene containing the target speech bubble, and context information across scenes is not taken into account.

In this work, we propose two methods for manga MT with the aim of using more contextual information from longer contexts: one that takes into account the previous scenes and another that considers bibliographic information such as the title of the work and the name of the author. In experiments on Japanese to English manga translation, our methods achieved improved translation performance. Furthermore, by combining both methods, we achieved a 4.24 point improvement in BLEU (Papineni et al., 2002) compared to the baseline, thus achieving SOTA performance.

2. Methods

This section explains our proposed methods for manga MT. Figure 2 shows an overview of them.

2.1. Scene-based context-aware NMT

In Figure 2, the speech bubble at the beginning of the second scene is part of a dialogue that continues from the previous scene. Therefore, it is difficult to translate this speech bubble properly using 2+2 translation and scene-based translation, because the subject of the sentence in the speech bubble is missing. However, by referring to the previous scene, the subject can be identified as "お前 = 相川仁 (you = Jin Aikawa)".

To address this problem, we propose a scene-based translation method that utilizes the previous scene. Based on the approach of the concatenation-based context-aware machine translation (Tiedemann and Scherrer, 2017), our method concatenates the utterances in the scene to be translated with the utterances in the previous scene, with special tokens inserted between the scenes. Since there can be multiple speech bubbles within a scene, two types of special tokens are used to distinguish between changes in scene and speech bubble. In the example shown in Figure 2, "相川仁 <Inside> お前には･･･学力がある <Outside> なのになぜ･･･狙う？ <Inside> 実に夢がない (Jin Aikawa, <Inside> you have the academic ability ･･･ <Outside> why aim for ･･･ ? <Inside> don't you have aspirations?)" is input to the MT model.

2.2. Attribute-aware NMT

It is possible that there is a bias in the expressions used in different manga genres (sports, romantic comedy, mystery...), or that each author or work has its own characteristics of expression. To capture such characteristics, we propose a MT method that incorporates the bibliographic attributes of manga by applying a style control approach (Sennrich et al., 2016a; Niu et al., 2017; Niu and Carpuat, 2020; Agrawal and Carpuat, 2019; Tani et al., 2022; Yamagishi et al., 2017; Johnson et al., 2017). For example, in the manga shown in Figure 2, a special token is attached to the beginning of the input text to indicate the author, as in "<朽鷹みつき> 俺間違ってます? (<Kutitaka Mitsuki> am i wrong?)."

In this work, we use five manga attributes: series, author, publisher, magazine, and genre. This attribute information was collected from Wikipedia's Infobox and MangaPedia[2], an online encyclopedia for manga.

[2] https://mangapedia.com/

	BLEU	COMET
Sentence-NMT	18.93	0.761
2+2 NMT (Hinami et al., 2021)	21.14	0.779
Scene-NMT (Hinami et al., 2021)	21.02	0.782
Scene-aware NMT (Ours)	**21.46**	**0.790**

Table 1: Translation performance of scene-aware manga translation.

	BLEU	COMET
Sentence-NMT	18.93	0.761
+ series	20.87	**0.771**
+ author	20.46	0.767
+ publisher	18.91	0.760
+ magazine	19.49	0.762
+ genre	20.58	0.766
+ all attribute information	**21.08**	0.766
Scene-aware NMT	21.46	0.790
+ series	22.51	0.787
+ author	23.02	0.788
+ publisher	22.14	0.790
+ magazine	22.65	**0.794**
+ genre	22.61	0.789
+ all attribute information	**23.17**	0.792

Table 2: Translation performance of attribute-aware manga translation.

3. Experiments

3.1. Setup

We run experiments to evaluate manga translation from Japanese to English using Manga Corpus (Hinami et al., 2021). We randomly extract 50 series from Manga Corpus, and 10 pages from the latest volume of each series to construct the validation data. We also construct the evaluation data in the same way. As a result, we obtain two datasets of 50 series, 500 pages, 1,000 scenes, and 2,000 speech bubbles as the validation and evaluation data. Other pages were used as training data.

We use KyTea[3] (Neubig et al., 2011) for Japanese word segmentation, Moses[4](Koehn et al., 2007) for English word segmentation, and perform subword segmentation using Byte Pair Encoding[5] (Sennrich et al., 2016b) for both languages.

We trained a Transformer (big) model (Vaswani et al., 2017) using fairseq[6] (Ott et al., 2019). We used Adam (Kingma and Ba, 2015) for optimization. Our batch size was set to 1,024, the label smoothing to 0.1, the dropout rate to 0.3, the learning rate to 1e-7, and warmup step count was 4,000. Training was stopped when the cross-entropy loss for the validation data stopped improving with a patience of five epochs.

To automatically evaluate translation quality, BLEU (Papineni et al., 2002) and COMET (Rei et al., 2020) were calculated using SacreBLEU[7] (Post, 2018) and XCOMET[8] (Gurreiro et al., 2023), respectively. BLEU and COMET were evaluated scene by scene and the average of three models trained using different random seeds is reported.

For the baseline, we use Sentence-NMT, which translates each speech bubble individually. For the existing methods, we use 2+2 NMT (Hinami et al., 2021) and Scene-based NMT (Scene-NMT) (Hinami et al., 2021). We then compare our scene-based context-aware translation method (Scene-aware NMT) explained in Section 2.1 with these methods, and apply our attribute-aware translation method (Attribute-aware NMT) explained in Section 2.2 to the baseline and the Scene-aware NMT to evaluate the effectiveness of our methods.

3.2. Result

Table 1 shows the results of evaluation for the baseline model, the existing methods and our proposed scene-aware method. Compared to the baseline model, the existing and proposed methods each improved BLEU by more than 2 points, showing that contextual information is useful in manga MT. Among them, our method achieved the best performance, demonstrating the importance of using more contextual information from longer contexts.

Table 2 shows the results of MT that incorporates the manga attribute information. In most cases, the translation performance was improved using the attribute information for both the baseline without context and the proposed model with context. However, we did not observe any improvement in BLEU in case of adding publisher information. This may be due to the fact that it is difficult to characterize the translation style from the publisher alone, as a variety of manga is published by a single publisher across many genres and authors. The other attribute information was effective, and improved BLEU. Furthermore, using all fives types of attribute information together achieved the best performance for both the baseline and our method.

[3] http://www.phontron.com/kytea/
[4] https://github.com/moses-smt/mosesdecoder
[5] https://github.com/glample/fastBPE
[6] https://github.com/facebookresearch/fairseq
[7] https://github.com/mjpost/sacrebleu
[8] https://github.com/Unbabel/COMET

	Input	Output
Sentence-NMT	実に夢がない	I don't have a dream.
2+2 NMT	なのになぜ··· 狙う? <SEP> 実に夢がない	I don't have any dreams.
Scene-NMT	なのになぜ··· 狙う? <SEP> 実に夢がない	I don't have any dreams.
Scene-aware NMT	相川仁 <Inside> お前には··· 学力がある <Outside> なのになぜ··· 狙う? <Inside> 実に夢がない	You really don't have a dream.
Reference	実に夢がない	Don't you have aspirations?

Table 3: Example of translation for "実に夢がない(Don't you have aspirations?)" in Figure 2.

	BLEU	COMET
Scene-NMT	17.94	0.751
+ 1 previous scene	**18.55**	**0.757**
+ 2 previous scenes	18.04	0.749
+ 3 previous scenes	16.72	0.728
+ 4 previous scenes	18.11	0.746
+ 5 previous scenes	17.52	0.739

Table 4: Translation results when the previous scenes are considered as context.

	BLEU	COMET
Scene-aware NMT	21.46	0.790
Scene-aware NMT (change scene ordering)	21.39	0.788
Scene-aware NMT (change the speech bubble ordering within a scene)	22.15	0.791
Next scene-aware NMT	21.71	0.791

Table 5: Translation results with randomly changing reading order of scenes and speech bubbles and with consideration of the next scene.

4. Analysis

Table 3 shows examples of translations improved by the proposed method. In the second scene of Figure 2, the speech bubble "実に夢がない (Don't you have aspirations?)" does not have its subject in this scene, making accurate translation difficult with existing methods. Referring to the previous scene, we see that the subject of "お前には一流の大学を目指す学力がある (you have the academic ability to enter top universities.)" is "お前 (you)." The baseline model and existing models that do not consider the previous scene incorrectly output "I" as the subject, but our model can refer to the previous scene and output "you" properly. This example shows that considering longer contexts can improve manga MT.

We also investigate the extent to which it is effective to consider previous scenes in manga MT.

	BLEU	COMET
Two types of special tokens	**21.46**	**0.790**
One type of special token	21.44	0.785
w/o sentence boundaries	19.42	0.696

Table 6: Analysis of the effectiveness of tags to identify context types in our Scene-aware NMT.

Table 4 shows the translation performance when even more distant scenes are considered as context. The translation performance was the highest when one previous scene (+1 previous scene) is considered. However, when considering two or more previous scenes, the translation performance deteriorated. These results indicate that considering only one previous scene was the most effective.

Table 5 shows the translation performance when the scene ordering is randomly changed, when the speech bubble ordering is randomly changed within a scene, and when the next scene is used as context. Translation performance decreased when the order of scenes was changed, and improved when the speech bubbles ordering within a scene was changed. The performance was also better when the next scene was taken into account.

Table 6 shows the results of our analysis of the effectiveness of tags to identify context types in our Scene-aware NMT. Our context identification based on two types of special tokens (<Outside> and <Inside>) achieves higher translation performance for both BLEU and COMET than methods that do not explicitly represent sentence boundaries or represent sentence boundaries based on one type of special token (<SEP>).

Table 7 shows an example of how translation results change when manga attributes are taken into account. We examine the translation results of the proposed method when different genres were specified for the input sentence "最初から翔ばないほうがいい(why fly in the first place.)." When the genre of <Baseball> was specified the expression "jumped" was changed to "hit", as in "to hit a ball". When the <Robot> genre was specified, the expression was changed to "take flight". In addition, when

Input	Genre	Output
最初から翔ばないほうがいい (why fly in the first place.)	<Shonen manga> <Baseball> <Robot> <Seinen manga>	You shouldn't have jumped from the start. You shouldn't have hit from the start. You shouldn't have taken flight in the first place. It's better if you didn't fly from the start.

Table 7: Example of how a change in genre changes the translation result.

a genre such as <Seinen manga> was specified, the structure of the sentence changed significantly, rather than word by word. This example shows how considering the genre of manga can change translation results.

5. Conclusion

In this work, we present methods for utilizing longer context to improve the performance of manga MT. Specifically, we propose two methods: (1) scene-based context-aware translation and (2) attribute-aware translation. Experiments on manga MT from Japanese to English showed that each of the two proposed methods improved translation performance, and the combination of the two methods achieved the best performance. These results demonstrate the effectiveness of using more context information than just single scenes.

In future work, we would like to further improve translation quality by utilizing multimodal MT and multilingual MT models that have been pre-trained on a large corpus.

Acknowledgements

These research results were obtained from the commissioned research (No.22501) by National Institute of Information and Communications Technology (NICT), Japan.

Bibliographical References

Sweta Agrawal and Marine Carpuat. 2019. Controlling Text Complexity in Neural Machine Translation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 1549–1564.

Nuno M. Gurreiro, Ricardo Rei, Daan van Stigt, Luisa Coheur, Pierre Colombo, and André F. T. Martins. 2023. xCOMET: Transparent Machine Translation Evaluation through Fine-grained Error Detection. *arXiv:2310.10482*.

Ryota Hinami, Shonosuke Ishiwatari, Kazuhiko Yasuda, and Yusuke Matsui. 2021. Towards Fully Automated Manga Translation. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence*, pages 12998–13008.

Melvin Johnson, Mike Schuster, Maxim Krikun Quoc V. Le, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, , and Jeffrey Dean. 2017. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Diederik P. Kingma and Jimmy Lei Ba. 2015. AdamA Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference on Learning Representations*.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180.

Graham Neubig, Yosuke Nakata, and Shinsuke Mori. 2011. Pointwise Prediction for Robust, Adaptable Japanese Morphological Analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 529–533.

Xing Niu and Marine Carpuat. 2020. Modeling Coherence for Discourse Neural Machine Translation. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, pages 2374–3468.

Xing Niu, Marianna Martindale, and Marine Carpuat. 2017. A Study of Style in Machine Translation: Controlling the Formality of Machine Translation Output. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2814–2819.

Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. fairseq: A Fast, Extensible Toolkit for Sequence Modeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 48–53.

Kishore Papineni, Salim Rukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318.

Matt Post. 2018. A Call for Clarity in Reporting BLEU Scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191.

Ricardo Rei, Craig Stewart, and Alon Lavie Ana C Farinha. 2020. COMET: A Neural Framework for MT Evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 2685–2702.

Rico Sennrich, Barry Haddow, and Alexandra Brich. 2016a. Controlling Politeness in Neural Machine Translation via Side Constraints. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 35–40.

Rico Sennrich, Barry Haddow, and Alexandra Brich. 2016b. Neural Machine Translation of Rare Words with Subword Units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 1715–1725.

Kazuki Tani, Ryoya Yuasa, Kazuki Takikawa, Akihiro Tamura, and Tomoyuki Kajiwara. 2022. A Benchmark Dataset for Multi-Level Complexity-Controllable Machine Translation. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6744–6752.

Jörg Tiedemann and Yves Scherrer. 2017. Neural Machine Translation with Extended Context. In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 82–92.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 5998–6008.

Hayahide Yamagishi, Shin Kanouchi, Takayuki Sato, and Mamoru Komachi. 2017. Improving Japanese-to-English Neural Machine Translation by Voice Prediction. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing*, pages 277–282.