JAIST Repository
https://dspace.jaist.ac.jp/

| | |
|---|---|
| Title | エンティティ認識型機械翻訳を改善するための多面的強化学習アプローチ |
| Author(s) | TRIEU, HOANG AN |
| Citation | |
| Issue Date | 2026-06 |
| Type | Thesis or Dissertation |
| Text version | ETD |
| URL | https://hdl.handle.net/10119/20625 |
| Rights | |
| Description | Supervisor: NGUYEN, Minh Le |

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

**Abstract**

Language barriers have long hindered effective knowledge sharing and the advancement of society. In recent years, the rapid development of machine translation, one of the most complex tasks in natural language processing, has contributed significantly to overcoming this obstacle. The impressive performance of modern translation systems demonstrates their capability in handling general-domain text with high accuracy. However, when it comes to terminology-sensitive domains such as medicine or law, these systems often fail to accurately translate domain-specific terms. This limitation poses a serious issue: even a single mistranslated term can alter the meaning of the original text and lead to severe misinterpretations, especially in high-stakes contexts. Motivated by this challenge, the goal of this thesis is to develop a machine translation system capable of reliably and correctly translating terminologies in specialized domains.

The first challenge lies in ensuring that the system maintains strong general-domain translation performance after domain adaptation. Fine-tuning a model for a specific field may cause it to lose previously learned knowledge, resulting in degraded performance on general text. To address this, we propose a multiple-aspect reinforcement learning approach that evaluates and optimizes the model using multiple complementary criteria. This technique allows us to control knowledge retention inside the model, preventing catastrophic forgetting while simultaneously enhancing its ability to correctly translate specialized terminology. We further introduce clear definitions of the proposed aspects and describe how they can be incorporated into the training process to improve overall system performance.

The second challenge pertains to the limited availability of high-quality data for assessing and improving terminology translation accuracy. Creating datasets that precisely identify and evaluate domain-specific terminology usage is costly and resource-intensive. To overcome this, we propose a data generation framework for producing a high-quality silver-standard dataset to support training and fine-tuning. Additionally, we outline criteria for constructing a smaller but reliable gold test dataset used for evaluation and testing.

In summary, this thesis aims to enhance machine translation systems by improving their ability to translate specialized terminology while preserving strong performance in general-domain translation. Through a combination of reinforcement learning and

efficient data generation methods, our approach provides a promising direction for advancing terminology-aware machine translation.

**Keywords**: Machine Translation, Large Language Models, Entity-Aware Machine Translation, Reinforcement Learning, Dataset Construction.