論 文 内 容 の 要 旨

博士論文題目

Study of Content Order-Controllable MR-to-Text Generation

（内容順序の制御が可能な MR-to-Text 生成に関する研究）

氏　　名　　東山 恵祐

（論文内容の要旨）

Content order is critical in natural language generation (NLG) for emphasizing the focus of a generated text passage.

In this thesis, we propose a novel MR (meaning representation)-to-text method that controls the order of the MR values in a generated text passage based on the given order constraints. We develop a refined MR-text dataset with additional value order annotations to train our order-controllable MR-to-text model. We also use it to train a text-to-MR model to check whether the generated text passage correctly reflects the original MR. Furthermore, we augment the dataset with synthetic MR-text pairs to mitigate the discrepancy in the number of non-empty attributes between the training and test conditions and use it to train another order-controllable MR-to-text model. Our proposed methods demonstrate better NLG performance than the baseline methods without order constraints in automatic and subjective evaluations. In particular, the augmented dataset effectively reduces the number of deletion, insertion, and substitution errors in the generated text passages.

氏　名	東山 恵祐

（論文審査結果の要旨）

This thesis sheds light on the problem of content order in natural language generation. The content order in natural language can vary in order to prioritize and emphasize something in natural language communication. This thesis work aims to control the content order in an MR (meaning representation)-to-text NLG task. The contribution of this thesis is two-fold: the development of a refined MR-text dataset and the proposal of a novel MR-to-text NLG method using Transformer-based sequence-to-sequence transduction with content order constraints. The refined dataset was built upon a popular MR-text benchmark dataset called E2E Dataset with error corrections and additional content order annotations. The proposed order-controllable MR-to-text method takes ordered MR values as its input and encodes values and their positions in a Transformer encoder. This work also leverages the MR-text dataset by data augmentation to compensate for the data imbalance in the number of non-empty attributes (contents). In the experiments, the proposed method outperformed the baseline methods without content order considerations in automatic and subjective evaluations.

This thesis work established a new direction for NLG technologies through a thorough investigation in an MR-to-text task: data refinement, development of a novel order-controllable NLG model architecture, data augmentation, and evaluation. A series of this research resulted in one peer-reviewed journal paper and one peer-reviewed international conference paper. As a result, this thesis sufficiently qualified as a Doctoral thesis of Engineering.