

知能情報メディア主専攻 201611350 江畑 拓哉

指導教員 Claus Aranha (コンピュータサイエンス専攻)

櫻井鉄也 (コンピュータサイエンス専攻)

北川高嗣

今倉暁

二村保徳

保國恵一

提出日 2019 年 2 月 1 日

## Abstract

One of the fields in NLP (Natural Language Processing) is Dialogue systems. Recently, Dialogue systems are often built using many systems, each solving a different small problem for the main dialogue system. These systems use various methods such as rule-based, machine learning etc. In this report, I aimed at constructing a Dialogue System focused in the Japanese language. First, I designed an abstract structure for the dialogue system, and next, I studied some of the required methods for this structure, using mainly deep learning techniques. In addition, I investigated the method for collecting and preprocessing Japanese data, which is indispensable for constructing this system.

In other words, the subjects of this research were the following five themes. 1st, the dealing methods for Japanese data. I observed Japanese utterance data and conversation data from free data sets such as the corpus published by NTT(Nippon Telegraph and Telephone Corporation) and Twitter data and consider these natures, as well as some preprocessing methods for them. 2nd, the classification task for imbalanced data. The meaning of “imbalance” indicates that the difference of the abstraction degree of each class and the proportion of each class in data. 3rd, the conversation model using machine translation model. I tried some machine translation models for conversation model which responses 1-by-1 such as a general greeting. 4th, the sentence style translation. I tried several models which I chose to learn translation of sentence style. In Japanese, the change at the sentence end (this research deals with the main agenda of style translation) affects people’s detection of persona too much. So, I believe the style translation will lead the user of this Dialogue System to feel more comfortable by choosing the best style for its system. 5th, the error (in which sentences generated by some deep learning models) detection methods. The main reason of the low motivation for choosing deep learning models to build dialogue systems is the distrust of the ability of deep learning systems to build natural phrases, in comparing to rule-based models. In this theme, to connect some error handling, I experimented the detection method of unnatural sentence generated by my model which made in the 3rd theme.