

論文の内容の要旨

論文題目 Automatic Semantic Role Labeling using Multiple Generalization Criteria of Semantic Roles
 (多様な意味役割の汎化指標を利用した自動意味役割付与)

氏 名 松林 優一郎

This thesis focuses on the problem of generalizing semantic roles in automatic Semantic Role Labeling (SRL), and increases the accuracy and robustness of SRL, proposing several new generalization criteria for FrameNet and PropBank corpora, and a classification model which enables us to combine these multiple criteria. The study also clarifies the characteristics of roles which can be captured by each criterion, through the comparative experiments of the criteria. In recent years, a number of studies have applied machine-learning approaches to SRL due to the availability of semantic corpora such as FrameNet and PropBank. However, these corpora define specific semantic roles for each semantic structure (or frame) of the predicates, and this causes severe problems for machine-learning approaches since the corpora contain a number of infrequent roles, which hinder efficient learning. Previous studies have addressed this problem by replacing frame-specific semantic roles with common labels among the frames and sharing the instances of the roles having common characteristics. However, this is a method which employs equivalence classes constructed from one linguistic perspective as classification labels, and there are two problems in the method. First, computational linguistic resources having some linguistic theory as their basis assign to semantic roles the information which characterizes the roles from multiple aspects. In other words, this indicates that the characteristics of semantic roles are difficult to explain from only one perspective. Therefore, a model which predicts semantic roles by generalizing them from multiple viewpoints could improve the performance of role classification in the case of SRL. Secondly, it is also an important aspect to evaluate whether a computational model which employs the information from a linguistic theory can really contribute to SRL for real-world texts, for further improvement of linguistic theories and NLP technologies. However, previous studies could not strictly compare the different types of generalized labels since they redefined a SRL problem as a different task where a model directly predicts generalized labels when they generalize the semantic roles. Therefore, in this thesis, we perform the following three. First, we construct a model which can achieve a simultaneous use of multiple types of generalized labels by using a log-linear model taking into account features of prediction labels. Secondly, we propose several new types of generalized labels derived from different generalization criteria both for FrameNet and PropBank, using characteristics of semantic roles annotated in the corpora based on some linguistic theory. The effectiveness of each criterion and incorporation of them into a model are also evaluated in the experiments. Thirdly, we clarify the characteristics of semantic roles which are captured by each criterion, based on comparative experiments and analyses. Experimental results confirmed that combining multiple generalization criteria capturing syntactic/semantic properties of roles from different perspectives improves the performance of SRL. The combined model for each corpus achieved 19.16% error reduction in total accuracy and 7.42% in macro F1 average on FrameNet, and reduced 24.07% of errors in total accuracy and 26.39% of errors in the evaluation of unseen verbs on PropBank.