Inducing a Semantically Rich Nested Event Model

Nyuk Hiong Siaw[(⊠)], Bali Ranaivo-Malançon,
Narayanan Kulathuramaiyer, and Jane Labadin

Faculty of Computer Science and Information Technology,
University Malaysia Sarawak, Kuching, Malaysia
ftsm2006@yahoo.com,
{mbranaivo,nara,ljane}@fit.unimas.my

Abstract. Research has revealed that getting data with named entities (NEs) labels are laboured intensive and costly. This paper is proposing two approaches to enable NE classes to be added to the semantic role label (SRL) predicate-argument structure of Nested Event Model. The first approach associates SRL to Named Entity Recognition (NER), which is named as SRL-NER, to tag the appropriate entity class to the simple argument of the model. The second approach associates SRL to NER by fine-tuning entities in complex argument structures with Automatic Content Extraction (ACE) structure. This approach is called SRL-ACE-NER. Stanford NER tool is used as the benchmark for evaluation. The result shows that the proposed approaches are able to recognize more PERSON entities. However, the approaches are not able to recognize LOCATION/PLACE as efficiently as the benchmark. It is also observed that the benchmark tool is sometimes not able to tag as comprehensively as the proposed approaches. This paper has successfully demonstrated the potential of using a semantically enriched Nested Event Model as an alternative for NER technique. SRL-ACE-NER has achieved an average precision of 92 % in recognising PERSON, LOCATION/PLACE, TIME, and ORGANIZATION.

Keywords: Named entity recognition · Nested event model · Semantic · Semantic role label · Predicate argument structure

1 Introduction

Named Entity Recognition (NER) is the task of identifying the names of entities from unstructured text and grouped them into predefined semantic categories [1]. Semantic categories provide same types of entities into classes such as PEOPLE, PLACE, ORGANIZATION, TEMPORAL and NUMERICAL EXPRESSIONS. NER is a very important part of an Information Extraction (IE) system. Handcrafted rules and learning based methods are the two main approaches for NER [2]. However, these two approaches have some drawbacks. The handcrafted rules method relies fully on rules created manually. As for the learning based methods, training labelled data is needed which can be created manually or semi-automatically. The manual process of creating extraction rules and training data is claimed to be costly and time consuming. Another issue related to