Chapter 75
Semantic Role Labeling Approach for Evaluation of Text Coherence

Mohamed H. Haggag
Helwan University, Egypt

ABSTRACT

Detection of semantic roles associated with linguistic elements is important to the textual classification of communicative context into specific identities. In this paper, a new model for semantically identifying sentences is presented through contextual patterns. The proposed contextual pattern originated its structure from a labeling process of the semantic roles provided by constituents of a sentence within a semantic frame. Semantic roles of the pattern elements are properly identified through word sense disambiguation and accordingly the entire patterns sense is evaluated. Such semantic identification of text sentences is a generic semantic role labeling approach that could support many computational linguistic applications. A utilization of the proposed semantic labeling approach is introduced in the paper through a novel algorithm for text coherence evaluation. Coherence evaluation is provided by a matching task to individual semantic patterns and their relations to each other as well as patterns organization within the text segments. Results proved good capability of the modelling of contextual pattern, addressing semantic roles, to accurately evaluate text coherence. It has been shown that both contextual patterns labeling and coherence evaluation algorithm proposed here are generic, topic free and semantically arbitrated by the global concept within context.

1. INTRODUCTION

Semantic interpretation is a centric need for different Natural Language Processing applications like Information Extraction, Question Answering, Text Summarization, and Emotion Extraction from Text, Machine Translation and Question Generation. The primary task of Semantic Role Labeling (SRL) is to indicate the semantic relations belonging to a predicate and its associated participants and properties. Such relations are pre-specified in a list of all possible semantic roles for that predicate.

Semantic role labeling is the process, in natural language processing, of detecting the semantic arguments associated with a predicate or verb

DOI: 10.4018/978-1-4666-6042-7.ch075

of a sentence. This detection identifies the arguments classification into specific roles. Semantic arguments include Agent, Patient, Instrument, Locative, Temporal, Manner and Cause Aspects. A semantic role is a relationship having a syntactic predicate, which is often the verb of a sentence (Wu, 2010). For example 'he likes the deer', the predicate or is 'likes'. 'He' and 'deer' have the semantic role labels agent and subject (theme). The crucial fact about semantic roles is that regardless of their syntactic structure, the underlying predicates remain the same. Recognizing and labeling semantic arguments is a key task for answering "Who," "When," "What," "Where," "Why," etc.

Researchers have been proposed to either identify semantic roles or to build semantic classifier. Most of the existing models for automatic semantic role labeling are based on a full syntactic parse of the sentence. This parsing defines argument boundaries and accordingly, relevant information for classifiers' training is needed to disambiguate role labels. Of the earliest models, (Hirst 1988) presented a foundation to deal with the semantic complexities of text, which is suitable basis for both lexical and structural disambiguation. MindNet identified words labelled with semantic relations and structures acquired from semantic rules of natural language (Richardson, Dolan & Vanderwende, 1998).

In 1990s, statistical machine learning approaches have been increasingly developed in the computational linguistics domain. Models was able to learn from complex linguistic knowledge like learning sub-categorization frames (Brent & Berwick, 1991) and classifying verbs according to argument structure properties (Merlo & Stevenson, 2001). Gildea et al. (Gildea & Jurafsky, 2002) presented a system to make predicate-argument structure that reads from parse tree. This structure identified semantic roles through predicate, agent, theme and manner.

Hierarchical SRL proposed by Moschitti, Giuglea, Coppola and Basili (2005) generalized the classical two-level approach (boundary detection and classification) and showed more efficiency and accuracy. A semantic frame, classically pictured in Fillmore (1975), Fillmore (1977), Fillmore (1978), Fillmore (1982) and Fillmore (1985) is a conceptual representation describing an event, relation, or object and the participants in it. FrameNet shows how frame-semantic annotations are created as frame-specific semantic roles generalizations that are not possible with more traditional linguistic approaches. Such conceptual roles are considered frame elements. Frame semantics relates words into semantic frames, where semantic and syntactic properties of predicating words are characterized (Johnson & Fillmore, 2000). A framework for semantic role annotations based on the FrameNet paradigm is introduced by Shen and Lapata (2007) Semantic role assignment is used to extract question answering through an optimization problem in a bipartite graph.

Disambiguation of the word senses is considered by Moreda, Fernández, Palomar, and Suárez (2004), who used supervised learning method for semantic role labeling based on maximum entropy conditional probability. The sense of the target verbs is used to determine the arguments of theses verbs. Linguistic knowledge, represented in features, is acquired from an annotated corpus. This method is based on three basic steps. Word Sense Disambiguation (WSD), heuristics and semantic role disambiguation. The process to obtain the semantic role needs the sense of the target verb. Following to this, several heuristics are applied to get the arguments of the sentence. Finally, the semantic roles that fill these arguments are obtained.

Form the language prospective, most of the reported SRL systems are for English. Most of the data resources exist for English as well. There are some approaches for other languages such as German (Erk, Frank, Kowalski, Pado & Pinkal, 2006) and Chinese (Sun, & Jurafsky, 2004; Chen, Sun & Jurafsky, 2004). In Diab, Moschitti and Pighin (2008), SRL is proposed for morphological features of the Arabic language based on support