Available onlinewww.ejaet.com

European Journal of Advances in Engineering and Technology, 2022, 9(4s):40-43
International Conference on Tech Trends in Science & Engineering (ICTTSE) 2022
Suryodaya College of Engineering & Technology, Suryodaya Polytechnic, Nagpur, Maharashtra, India

Research Article ISSN: 2394 - 658X

Semantic role Labelling: An Introduction with Machine Learning and artificial intelligence interface

Preeti S. Lokhande[1], Dr. Chandrakant S. Ragit[2]

Research Scholar[1], Department of Information and Language Engineering, MGAHV, Wardha
Pro-Vice Chancellor & Director[2], Department of Information and Language Engineering, MGAHV, Wardha

ABSTRACT

Semantic role labelling (SRL), sometimes referred to as thematic role labeling, case role assignment, or shallow semantic parsing, is a task in Natural Language Processing that determines the labels of words or phrases in a sentence. Agent, goal, patient/receiver, locative, object and temporal labels are the types of labels. Artificial Intelligence is the field in which semantic role labeling falls and that is why researchers try to do it automatically. It is used to determine the meaning of a sentence. This is accomplished by detecting the arguments connected with a sentence's predicate or verb, as well as how they are divided into their respective roles. In this paper we are trying to give a brief introduction to the concept of semantic role labeling, role of AI-ML in SRL, Application, challenges and state of art work done for semantic role labeling.

Key words: Semantic Role Labeling, Natural language Understanding, AI-ML, FrameNet, etc.

INTRODUCTION

Natural language processing is a branch of Artificial Intelligence concerned with programming computers to process and analyze massive volumes of natural language data used by people to communicate. A natural language text must be processed in several phases in order to be easily understood and analyzed, with each stage being based on the previous stages. Natural language processing can be categorized into two natural language understanding and natural language generation, semantic role labeling falls in natural language understanding (NLU). Semantic role labeling is a technique in natural language understanding that provides tags to phrases and words in a sentence that reflects their semantic function in the sentence, such as agent, purpose, or consequence. It's used to determine the sentence's meaning. This is accomplished by detecting the arguments connected with a sentence's predicate or verb, as well as how they are divided into their respective roles. SRL tries to determine each predicate's semantic roles, i.e., who did what to whom, when and where, and so on. SRL has been a hot topic in NLP research, and as a result, there has been a substantial amount of work done for languages like English, Chinese, and others, including efforts to create cutting-edge role labels as well as the generation of semantically annotated data such as FrameNet and PropBank. SRL also known as thematic role and it has different types of labels, the following table gives summarized information about the labels types with an example.

Table 1: Semantic Role Labels types with an examples

Thematic Role	**Definition**	**Example**
AGENT	An event volitional causer	The *Waiter* spilled the Coffee
EXPERIENCER	An event experiencer	*Seema* has a headache
FORCE	The event non-volitional causer	The *rains* broke our yards

THEME	An event affects the participant most directly	Only after Franklin broke *the bricks...*
RESULT	An event end product	The city was built *in the shape of a baseball diamond...*
CONTENT	The propositional event content or proposition	Reema asked *"You met Mary at a supermarket?"*
INSTRUMENT	An event instrument.	He poached fish, stunning them *with a vibrating device...*
BENEFICIARY	An event beneficiary	Whenever Ajay Diama makes hotel reservations *for her boss...*
SOURCE	In transfer event origin of the object	I flew in *from London.*
GOAL	In transfer event destination of an object	I drove *to Suzuki*

Semantic Role Labeling can be formulated in two ways, the first way is based on dependency-based parsing or dependency-based semantic role labeling and the second one on constituency-based parsing or span-based semantic role labeling. The dependency-based SRL finds only the headword of the argument and the span-based SRL finds a group of contiguous words that is the argument of the verb and correctly labeled them.

A sentence can be expressed in multiple ways. It may be analyzed through dependency parsing or constituent syntactically, but semantic roles can't be determined by the syntactic relationship of the sentence. While working on unstructured input text, it is challenging for SRL to understand beyond the syntax. Though the parsing is not completely useful for SRL, its performance can be impacted if the parse tree is wrong. Parsing only gives the syntactic relationship of a sentence; it is not useful to find the semantic role of a sentence; it will do so with the help of SRL. That is why there is a need for SRL along with a parser.

Semantic role labeling produces predicate-argument structures of sentences. The primary meaning of a situation described by a text is carried by predicates. Predicates are verb, verb nouns, and other types of verb forms. Arguments are sentences that specify and fill meaning slots in a situation indicated by a predicate. Arguments are sentences that define and fill in the meaning slots in a situation specified by a predicate. The all questions like who, did what, to whom, with what, where, are answered by arguments. As shown in the following example.

"Rakesh hit Sak with a cricket bat today"
Who hit Sak with a cricket bat?
Whom did Rakesh hit with a cricket bat?
What did Rakesh hit Sak with?
When did Rakesh hit Sak with a cricket bat?

Semantic role labeling can be achieved using some lexical resources like FrameNet, Propbank, Verbnet, and Wordnet. The structure used to define the semantic meaning of a word is called the frame. The sentences in the FrameNet dataset are organized in a hierarchical sequence, with each frame relating to a concept. Propbank is a proposition bank that annotates sentences with verbal propositions and their arguments. VerbNet is a hierarchical vocabulary resource that divides English verbs into different classes based on Levin's linguistic classification. WordNet is an extensive lexical database. Cognitive synonyms (synsets) are groups of nouns, verbs, adjectives, and adverbs that each express a distinct concept.

APPLICATIONS AND CHALLENGES OF SLR

NLP applications that need semantic understanding, the role of SLR is most useful for these applications. The SLR cannot have a direct application but it helps the systems such as machine translation, information extraction, summarization, question answering, textual entailment and Word sense disambiguation, etc., to enhance their performance outcomes. In Information Extraction advantages in development, time is achieved through SRL. The information extraction that has good SRL then its outcome is also a good information extraction system. In Text Summarization, Sophisticated sentence matching is done using SRL, if improved SRL is used for Summarization it improves the performance of text summarization. In the Question Answering application, more complex semantics increase the number of questions that can be handled about three times. Whereas in Textual Entailment SRL enables complex inferences that are not allowed using surface representations. The above applications show the importance of SLR developing a good Semantic-based application.

The major challenge for developing SLR is to handle natural languages which are rich in morphology, augmentation, and ambiguous in nature. The lack of resources is also a major challenge of SLR-based systems. The SLR needs to be connected to natural language processing applications. Surface representations and syntactic analysis bear little resemblance. It is necessary to investigate the added usefulness of semantic labelling. Validating the annotation standards through application domains for well-defined corpora is also challenging. Setting the accuracy level for generated SLR is also a challenge for developing SRL based system. NL interfaces and ATIS labeling to databases are very difficult for developing an SLR-based application. Domain-specific semantic extraction systems have been successfully deployed in recent years. Moving from domain-specific to domain-independent and robust systems is a challenge. This is now achievable due to advances in machine learning techniques and the emergence of big semantic databases

STATE - OF - ART WORK OF SEMANTIC ROLE LABELING

Semantics is an area of natural language processing that deals with extracting the meaning of a sentence. Semantic role labeling takes the first step in extracting meaning from text by assigning common labels or roles to words in the text. We can assume that the meaning of this small set of labels is machine comprehensible. Most of these databases are manually labeled by a large group of people. One of the basic works in this direction was performed by Jurafsky et.al [1]. They used lexical resources such as WordNet and FrameNet. Statistical methods are applied to these semantic databases for semantic extraction. Statistical techniques are applied to these semantic databases to extract semantics. The idea is to train supervised classifiers on this corpus that can be used to automatically tag large amounts of hidden text with shallow semantic information. Semantic role labelers are typically developed using a supervised learning model. The authors Pradhan et.al used Support Vector Machines (SVM) to train Automatic Semantic Role Labeling and they uses the features as Named Entities in Constituents, Head Word POS, Verb Clustering, Partial path, Verb Sense Information, Headword of prepositional phrases, First and Last word/POS in constituent, Ordinal constituent position, Constituent Tree Distance, Constituent relative features, Temporal cue words, and Dynamic class content [2]. O. Abend et. al present an unsupervised argument identification algorithm for SRL. They used POS tag annotation and an unsupervised parser that generates an unlabeled parse tree, the algorithm is able to achieve 56% precision on the argument identification task [3]. The maximum entropy classifier used by most people to train their models [4, 5, 6]. SVM using polynomial kernels [9] or Gaussian kernels [8] are also used by some researchers to train their machine models. A sparse network of linear separators was used by Punyakanok et al. They called it the SNoW (Sparse Network of Winnows) learning architecture [10]. S. P. Ponzetto et.al used Decision Trees like C4.5 with AdaBoost optimization. They find the combination of models gives improved results [11]. The authors L. Marquez et.al and M. Surdeanu et.al used decision tree ensembles with AdaBoost optimization. They observed that by ensembling machine models Output combinations can be developed by providing different input features, different learning methods, and creating n-best solutions lists [7, 12].

ROLE OF AI-ML IN SLR

Nowadays automation plays a vital role in natural language processing applications. This automation is achieved through AI-ML algorithms. The AI-ML learning techniques are supervised, unsupervised and semi-supervised. It is observed from state-of-art work done on SLR that supervised algorithms give a reasonably good performance as compared to other traditional techniques. However, its reliance on labeled training data, which is difficult and expensive to acquire, is a significant barrier to the broad use of semantic role labeling across languages and text genres. Individual predicates also have a scarcity of labeled data. A better option is to employ semi-supervised algorithms to annotate a large number of unlabeled instances that are comparable to the training instances using a limited number of manually labeled training instances. Unlike manually labeled data, which is costly to produce, unlabeled data is frequently available in vast amounts. The latter method seeks to boost a supervised Semantic Role's performance.

CONCLUSION

Semantic role labeling is an exciting field of study. Semantic role labeling is relatively new, but it has received a great deal of interest from researchers. Current State of the Art systems achieved about 80% per-argument F-measure performance is respectable but still, there is a lot of room for improvement. SRL Provides robust broad-coverage semantic representation and it is easy to integrate with applications like Question Answering, Information Extraction, Textual Entailment, and Summarization, etc. Good results in these tasks are achieved by SRL by using available large corpora with annotated data such as FrameNet, PropBank.

Future directions for SRL improvement are: Increase robustness to a syntactic parser error, find ways to collect additional knowledge, Use unlabeled data, Share information across verbs, improve the statistical models, other features, and other dependencies, and improve search/inference procedures. Try to find the answer to the question," Can applications create more data for SRL automatically?"

REFERENCES

[1]. Gildea Daniel and Jurafsky Daniel. Automatic Labeling of Semantic Roles. In Computational Linguistics, 2002

[2]. S. Pradhan, W. Ward, K. Hacioglu, J. Martin, and D. Jurafsky. Shallow semantic parsing using support vector machines. In Proceedings of HLT/NAACL, pages 233– 240, 2004

[3]. O. Abend, R. Reichart, and A. Rappoport. Unsupervised argument identification for semantic role labeling. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1, pages 28–36. Association for Computational Linguistics, 2009

[4]. K. Toutanova, A. Haghighi, and C. D. Manning. Joint learning improves semantic role labeling. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, pages 589–596. Association for Computational Linguistics, 2005

[5]. S.-t. Yi and M. Palmer. The integration of syntactic parsing and semantic role labeling. In Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005), pages 237–240, 2005.

[6]. A. Bharati, S. Venkatapathy, and P. Reddy. Inferring semantic roles using sub-categorization frames and maximum entropy model. In Proceedings of the Ninth Conference on Computational Natural Language Learning, pages 165–168. Association for Computational Linguistics, 2005

[7]. L. Marquez, P. Comas, J. Gim ` enez, and N. Catala. Semantic role labeling as sequential tagging. In ´ Proceedings of the Ninth Conference on Computational Natural Language Learning, pages 193–196. Association for Computational Linguistics, 2005.

[8]. N. E. Ozgencil and N. McCracken. Semantic role labeling using libsvm. In Proceedings of the Ninth Conference on Computational Natural Language Learning, pages 205–208. Association for Computational Linguistics, 2005.

[9]. S. Pradhan, K. Hacioglu, V. Krugler, W. Ward, J. H. Martin, and D. Jurafsky. Support vector learning for semantic argument classification. Machine Learning, 60(1):11–39, 2005.

[10]. P. Koomen, V. Punyakanok, D. Roth, and W.-t. Yih. Generalized inference with multiple semantic role labeling systems. In Proceedings of the Ninth Conference on Computational Natural Language Learning, pages 181–184. Association for Computational Linguistics, 2005

[11]. S. P. Ponzetto and M. Strube. Semantic role labeling using lexical statistical information. In Proceedings of the Ninth Conference on Computational Natural Language Learning, pages 213–216. Association for Computational Linguistics, 2005

[12]. M. Surdeanu and J. Turmo. Semantic role labeling using complete syntactic analysis. In Proceedings of the Ninth Conference on Computational Natural Language Learning, pages 221–224. Association for Computational Linguistics, 2005

[13]. https://nlp.stanford.edu/kristina/papers/SRL-Tutorial-post-HLT-NAACL-06.pdf

[14]. https://en.wikipedia.org/wiki/Semantic_role_labeling

[15]. https://devopedia.org/semantic-role-labelling#qst-ans-0