International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 04, April 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Semantic Role Labeling: The leading task in today's computational linguistics for the computational identification and labeling of arguments

Mr.Amit P.Patil, Ms.Chhaya S. Patil

Assistant Professor[1, 2], RCPET's Institute of Management Research and Development, Shirpur (MH)
patilamit03@gmail.com,chhaya.imrd@gmail.com

Abstract:

Semantic role labeling, sometimes referred to as shallow semantic parsing, is a task in Natural Language Processing that determines the labels of words or phrases (groups of words) in a sentence. Semantic Role Labeling (SRL) is the process of identifying essential event structures in a sentence, such as "who" did "what" to "whom," "when," and "where." Nowadays, the computational identification and labeling of arguments in text is a one of the important tasks in computational linguistics. Agent, receiver/patient, temporal, goal, locative or objects are the types of agents. Semantic role labeling belongs to one of the domains of Artificial Intelligence. This paper gives detailed overview of the literature in the field of semantic role labeling followed by the various techniques used to tackle shallow semantic parsing in the past, the development of different datasets built for semantic analysis and semantic role labeling task for Indian languages.

Keywords*: - SRL, NLP, Computational linguistics, Semantic Labeling*

Introduction:

One of the important but challenging tasks in NLP is semantic role labelling, which is also known as shallow semantic parsing. It determined "Who ",did "What", to "Whom", "How", "When" and "Where". SRL aim is to identify events in sentences, their participants and properties of the events. SRL determines the semantic roles of each predicate that present in the given sentence. For example,"Ram killed Kartik with a bat",here SRL should recognize 'killed' (represents the phrase "to kill") as a predicate. Then 'Pawan' as killer (agent), 'Kartik' as the recipient/receiver and 'a bat' as the theme/object. Knowledge of semantic relationships is very useful in many downstream applications of NLP such as information extraction, machine translation, text summarization, worsens disambiguation and text entailment.

Using two pipeline states SRL is done traditionally, first state is predicate identification and second one is argument role labelling, nowadays end to end SRL is achieved by integrating both steps in a single model[1]. Predicate detection, predicate disambiguation, argument identification, and argument classification are the four subtasks in which SRL can be split up. Argument annotations have two formulations or styles such as constituents that is phrase or span and dependencies. The CoNELL 2008 proposed semantic dependency parsing which annotates the heads of arguments instead of phrasal arguments [2].

Semantic Role Labelling in various theories of linguistics, also referred to as thematic roles, theta role, case frames, theta-grids, deep case, participant role etc.[3]. A predicate and various other dependents expressed by an event. These other dependents can be usefully classified into a small set of semantically content full classes and these classes are helpful for explaining lots of things, this claim by theory of semantic roles .Semantic role labelling identifies events in sentences, properties and participants of the events. Practically it finds arguments that accept a semantic role in a relation to a predicate. The following are the common semantic roles:

- **Agent**: the event initiator or doer.
- **Patient**: the event affected entity or undergoes the action
- **Experiencer**: An event that feels or perceives.
- **Stimulus**: the thing that is felt or perceived
- **Goal**: In event destination of transfer object
- **Recipient**: may or may not be distinguished from Goal
- **Benefactive**: may be grouped with Recipient.
- **Source**: origin of the object in the event
- **Instrument**: An event instrument.
- **Theme**: undergoing a change of state or location of an object in the event or of which location is predicated

Techniques/Approaches of semantic role labelling:

In this section we are explaining various techniques used for semantic role labelling.

Traditionally, SRL systems were rule based, which depended on rules derived from grammar. In the mid1990s the automatic semantic role labelling approaches, which are the statistical techniques, became popular due to the FrameNet and PropBank that was used for training the model. Using the feature sets, the classifier was trained over corpora of sentences. The feature set consist of predicate, head word and its POS, constituent phrase type, active/passive voice, constituent position and predicate-constituent path, etc.

SRL was traditionally a supervised task, but adequate annotated resources for training are a limitation to this approach. In early 2010s semantic roles and frames were research focused; also research on transferring an SRL model to low-resource languages was started.

In the supervised model, one novel approach trains using question-answer pairs. Using this technique, a non-expert can accurately generate a number of diverse pairs from a given sentence. Due to this there is no need of compiling a pre-defined inventory of frames or semantic roles.

Since the mid-2010s Neural network approaches to SRL are the state-of-the-art. Roth et.al used a dependency path between the predicate and its argument. An input to LSTM represented with Words and relations along the path. Binary features encoded by another input layer, these two inputs combined by hidden layer using RELU function. Classification was done by classifier layer[4]

Deep BiLSTM with highway connections and recurrent dropout was used by He et.al. This technique took the word predicate as an input, BIO tag notation used for predicated tag using softmax function and Glove is used for input embedding[5]. In other neural network techniques, a combination of CNN and BiLSTM was used to learn character encoding for input.

Self-attention has been used for SRL since 2018. Self-attention applied to train a model to do parsing, jointly predict POS tags and predicates, assign semantic roles, and attend to syntactic parse parents. Syntactic relations attended by one of the self –attention layers. BERT was used by shi et.al without using syntactic features for SRL and still got best results [6]

Related Work:

In this section, state-of-the-art on semantic role labelling is overviewed. An automatic semantic role labelling system was developed by Daniel Gildea and Daniel Jurafsky. They used the FrameNet corpus and they were the first who used a two-way approach to SRL that is the identification of argument constituents and labelling the identified arguments [7]

Gildea et.al built the system using Propbank dataset; they evaluated the system on gold standard and to depict automatic SRL. Using an automatic parser they first parse the sentence and do the role labelling.[8]

The C5 decision tree classifier was used for developing the SRL system; they also found the new features like named entity, content word and its POS tag [9].

Support vector machine (SVM) used by Pradhan et al. to develop the SRL system along with introducing new features like, named entities, POS tag of head word , head word of prepositional phrase and using only partial path, etc. [10]

Xue et.al.gave a crucial contribution to filtering out constituents which are definitely not going to be arguments with respect to a predicate. They used pre-processing as a filtration step [11].

Maximum entropy classifiers [12, 13, 14],SVM using polynomial kernels [15] or Gaussian kernels [16] have been applied to train their models. Sparse Network of Winnows) learning architecture is used by Punyakanok et al., it is s a sparse network of linear separators. AdaBoost optimization is used with Decision Trees like C4.5 [17], decision tree ensembles [18, 19].

Researchers observed that a combination of or ensembling of machine learning models gave the outperformance. The combination for model output was achieved by varying the inputs and features set, having different learning methods and creating n-best solutions lists. A stacking approach by learning a chunk-based SRL system was developed by Pradhan et al. [20].

Hacioglu [21] introduced a dependency parse for developing SRL systems. He used English propbank, to label only the syntactic heads. Dependency based SRL Officially announced by [22, 23]. They not only focused on English but also on other languages like Chinese, Spanish, Catalan, etc X. Carreras released a shared task for SRL systems development. In the shared task SRL system, all sentences were converted from constituency parse to dependency parse. Nominal predicates were introduced to Propbank from the Penn Treebank data [24].

DaniilLarionov et.al developed the first full pipeline SRL system for Russian Text. They used language models ELMo and BERT which gave outperform compared to shallow algorithms like word2vec and FastTex.[25]

Zuchao Li et.al.was introduced three SRL frameworks, namely sequence-based, tree-based, and graph-based, which are accompanied by syntax pruning and syntax feature syntactic information. The CoNLL-2005, -2009, and -2012 datasets were used to conduct the experiments.[26]

Resources for Semantic Role Labelling:

Framenet:

A word's semantic meaning is defined by a frame, which is a structure that defines it. It's a reusable concept with intuitively recognisable frame pieces. FrameNet presently comprises approximately 170,000 sentences that have been manually annotated. In this dataset each frame refers to a concept, and the sentences are placed in a hierarchical sequence. Higher-level frames correspond to a more general concept, whereas lower-level frames refer to more specific conceptions.

Propbank:

It is a proposition bank, in which sentences are annotated with verbal propositions and their arguments, and it is proposed by Martha Palmer et. al. In this dataset all the verbs in the corpus are annotated and all arguments to a verb must be syntactic constituents.

Verbnet:

It is a hierarchical lexical resource, based on the verbal classification of Levin it organizes English verbs into different classes. Mappings to other resources such as FrameNet and WordNet contained in Verbnet. The hierarchical VerbNets verb classes define a set of possible thematic roles and syntactic realisations.

Wordnet:

It is a large lexical dataset. Initially it developed for English Language later as work is advanced in other natural languages, it is also available for other foreign and Indian Languages. WordNet is a set of cognitive synonyms or synsets consisting of Nouns, verbs, adjectives and adverbs. Synsets are interlinked through conceptual-semantic and lexical relations

Indian Languages Semantic Role Labelling:

Indian languages have a problem of low resource scarcity. Due to this issue, the work for automatic semantic role labelling was only done for Hindi and Urdu Language using Hindi PropBank and Urdu PropBank. Anwar et al. developed this automatic SRL using the features like syntactic categories, dependency labels, head word's POS tag, Named entities and head word of the chunk, etc. Aishwary Gupta et.al developed a automatic semantic role labelling system for Hindi and Urdu Languages using Hindi and Urdu PropBank by introducing features like path from chunk to predicate, word embedding's to address the data sparsity issue and post-positionals of the chunk. Their system is divided into two subtasks such as Argument Identification and Argument Classification.

Conclusion:

In semantic analysis the highly popular task is semantic role labelling and using different methods for different languages, it has been investigated over the years. It is an important step towards finding the meaning of the sentence. It gives a lower level of abstraction instead of full syntactic parsing of sentences. In this paper we try to give a basic introduction to the SRL system, its application, and resources available for developing the SRL

system. We also explain the review of the state-of-art work done and current status of the SLR system in the context of Indian languages.

References:

1. Qingrong Xia et.al.," Syntax-aware Neural Semantic Role Labeling", Association for the Advancement of Artificial Intelligence (www.aaai.org,2019).
2. Zuchao Li et.al.,"Syntax Role for Neural Semantic Role Labeling", Association for Computational Linguistics 2021.
3. Shu-Ling Huang et.al.," Semantic Roles and Semantic Role Labeling", Technical Report · January 2015
4. Roth, Michael, and MirellaLapata. 2015. "Context-aware Frame-Semantic Role Labeling." Transactions of the Association for Computational Linguistics, vol. 3, pp. 449-460. Accessed 2019-12-28
5. He, Luheng, Kenton Lee, Omer Levy, and Luke Zettlemoyer. 2018b. "Jointly Predicting Predicates and Arguments in Neural Semantic Role Labeling." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 364-369, July. Accessed 2019-12-28
6. https://devopedia.org/semantic-role-labelling
7. D. Gildea and D. Jurafsky. Automatic labeling of semantic roles. Computational linguistics, 28(3):245–288,2002
8. D. Gildea and M. Palmer. The necessity of parsing for predicate argument recognition. In Proceedings of the 40th annual meeting on association for computational linguistics, pages 239–246. Association for Computational Linguistics, 2002
9. M. Surdeanu, S. Harabagiu, J. Williams, and P. Aarseth. Using predicate-argument structures for information extraction. In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics,2003
10. S. P. Ponzetto and M. Strube. Exploiting semantic role labeling, wordnet and wikipedia for coreferenceresolution.In Proceedings of the Human Language Technology Conference of the NAACL, Main Conference,2006.
11. N. Xue and M. Palmer. Calibrating features for semantic role labeling. In EMNLP, pages 88–94, 2004
12. 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
13. 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.
14. 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
15. S. Pradhan, K. Hacioglu, V. Krugler, W. Ward, J. H. Martin, and D.Jurafsky. Support vector learning for semantic argument classification MachinLearning, 60(1):11–39, 2005.

16. 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.
17. 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.
18. L. M arquez, P. Comas, J. Gimenez, and N. Catala. Semantic role labeling as sequential tagging. In Proceed-ings of the Ninth Conference on Computational Natural Language Learning, pages 193–196. Association for Computational Linguistics, 2005.
19. 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.
20. 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
21. N. FitzGerald, O. T ackstrom, K. Ganchev, and D. Das. Semantic role labelling with neural network factors.In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 960–970, 2015
22. M. Surdeanu, R. Johansson, A. Meyers, L. M arquez, and J. Nivre. The conll-2008 shared task on joint parsing of syntactic and semantic dependencies. In Proceedings of the Twelfth Conference on Computational Natural Language Learning, pages 159–177. Association for Computational Linguistics, 2008
23. J. Hajic, M. Ciaramita, R. Johansson, D. Kawahara, M. A. Mart L. M arquez, A. Meyers, J. Nivre, S. Pad et al The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages.InProceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task ages 1–18. Association for Computational Linguistics, 2009
24. X. Carreras and L. M arquez. Introduction to the conll-2005 shared task: Semantic role labeling. In Pro-ceedings of the ninth conference on computational natural language learning , pages 152–164. Association for Computational Linguistics, 2005
25. DaniilLarionov et.al, “Semantic Role Labeling with Pretrained Language Models for Known and Unknown Predicates",Proceedings of Recent Advances in Natural Language Processing, pages 619–628, Varna, Bulgaria, Sep 2–4, 2019
26. ZuchaoLi,”Syntax Role for Neural Semantic Role Labeling”, Association for Computational Linguistics ,2021