A Review on Semantic Role Labeling

Rashmi R. Chouhan
Master Of Computer Application Department
Sarvajanik College of Engineering and Technology, Surat
rashmi.chouhan@scet.ac.in

Zankhana B. Vaishnav
Master Of Computer Application Department
Sarvajanik College of Engineering and Technology, Surat
zankhana.vaishnav@scet.ac.in

***Abstract*—Semantic means to draw exact meaning of the text or to extract meaning from a sentence. Semantic Role Labeling is the process that assigns a generic labels or roles to the words of the sentence that indicate their semantic role in the sentence. Semantic role labeling (SRL) identifies the syntactic constituents, predicate-argument structure in text with semantic labels. In this paper we are giving brief review about Semantic Role Labeling system developed for different languages. This review basically highlights certain parameters like techniques used, languages for which SRL is developed, dataset used, classification of algorithm and accuracy.**

***Keywords:* Natural Language Processing ,Semantic Role Labeling, Proposition bank**

I. INTRODUCTION

Semantic role labeling is the task of assigning word meaning to sentence meaning by determining semantic arguments associated with the predicate and the then determining the appropriate role for each of these arguments. For example, given a sentence like "Mr. Kalpesh sold computer to Mr. Sriyaan", Here "to sell" would be to recognize as the predicate and arguments to this predicate are "Mr. Kalpesh" is denoted as the seller (agent), "the computer" is denoted as the goods (theme), and "Mr. Sriyaan" is denoted as the recipient as shown in figure 1. The figure shows how sentence is categorized into predicate(verb) and argument, further arguments can be assigned with specific roles. The SRL plays an important role for the Natural Language Processing(NLP) research areas such as question answering, summarization, machine translation and complex information extraction. In this paper we are presenting a brief review on Semantic Role Labeler for many Indian and other languages. Section II shows the related work done in SRL. The observation made by us in this section is on technique they used to develop SRL, language on which they develop SRL, dataset, classification of algorithm (Supervised, semi supervised or unsupervised) and accuracy. Section III contains the conclusion, our observation while reviewing and future work.

Mr. Kalpesh	sold	computer to	Mr. Sriyaan
Argument (agent)	Predicate	Argument (theme)	Argument (recipient)

Fig. 1. This figure shows how sentence is categorized into predicate and arguments. Further arguments can be assigned with specific roles.

II. WORK DONE IN SEMANTIC ROLE LABELING

In [1], Semantic Role labeling for Malayalam is implemented using Memory Based Language Processing (MBLP) Machine Learning approach. This approach uses two principles - simple storage representation in memory & solving new problems by reusing solutions from previously similarly solved problems. Tokenization, POS tagging, Chunking, Clause boundary identification have been carried out to find out semantic role labels.

In [2], authors have designed span-based model. The model is based on Bidirectional Long Short-Term Memory(BiLSTMs) to induce span representations. Here the model directly scores all possible labeled spans based on span representations induced from neural networks. The spans scoring higher are selected at the time of decoding.

The proposed model in [3], has following main tasks:

- Dependency Parsing
- Feature Extraction
- Predicate and Argument Identification
- Classification using Ensemble Classifier.

The result of parser is treated as an input to extract the features like POS Tag, Head Word of Phrase, path , position, Parent Dependency Relation, Selectional Preferences (SP), Collocational Features, ChildDepRelSet of given sentence. Based on these features semantic role classification is done. The ensemble classifier includes Support Vector Machine classifier and Decision Tree. To predict the labeling of predicate and arguments in the sentence, the Ensemble-Vote-Classifier is used. This classifier is a meta-classifier that merges a set of classifiers and then classifies a new data points by taking majority vote of their predictions.

In [4], the authors have introduced a statistical semantic role labeler based on supervised machine learning approach for Hindi and Urdu languages. The system uses a Logistic Regression algorithm for identifying the predicates and Support Vector Machines to classify the arguments of a predicate into semantic labels. They have used 10 features to guide the classifiers in predicting, identifying and classifying the arguments of a verb. They predict linguistic features like predicate, head, head-POS, phrase type etc. and combinations of certain features. To improve the accuracy of the system they use the dependency parses, uses the karaka relations/dependency relations as the most discriminative feature. Later, they extracted the automatic parses by using state-of-art Hindi and Urdu parsers and used them as features in our SRL. Authors classify the semantic arguments of a predicate on the basis of dependency relations because of their close proximity with semantic labels.

[5], paper presents POLYGLOT: A Web-based GUI to allow users to interact with the SRL systems. This system is capable of semantically parsing sentences in 9 different languages. The core of this system is SRL models for individual languages which is trained with automatically generated Proposition Banks. The original Proposition Banks project, funded by ACE, created a corpus of text annotated with information about basic semantic propositions. Predicate-argument relations were added to the syntactic trees of the Penn Treebank. This resource is now available via LDC. The SRL models uses the parser, which implements a sequence of local logistic regression classifiers for the four steps of predicate identification, Predicate classification, argument identification & argument classification. It also implements a global reranker to rerank sets of Local predictions. It uses a standard feature set of Lexical and syntactic features.

In [6], the genetic algorithm is used to optimize syntactic features. Efficiency and accuracy of the semantic role labeling system can be improved by finding the effective features from many syntactic features. On the basis of original semantic role labeling system, the genetic algorithm is used to optimize those syntactic features. Mostly, syntactic features are selected subjectively for semantic role labeling system and in order to improve the efficiency of SRL, it is necessary to extract more effective features from those subjective ones. This system is using 21 syntactic features, but it could achieve almost the same F value as the whole extended model with 29 syntactic features.

In [7], a novel approach is used by authors named as 5W MEMM(Maximum-entropy Markov model) + RULE-BASED-POST PROCESSING, which assigns semantic roles of Bengali nouns. 5Ws include "Who", "What", "When", "Where" and "Why". The rule based post-processor is working on the output of statistical system.

In [8], a dependency tree-based semantic role labeling (SRL) system is proposed. System accomplishes predicate identification, and automatically creates dependency relation using a dependency parser. Using effective pruning algorithm system cuts off the nodes which are not related with the predicate and proposes additional features based on Hacioglu’s baseline features like Predicate, Predicate POS, Predicate voice, Sub-categorization, Path, Position, Dependency relation type, Headword. This algorithm is applied to filter out unlikely dependency relation nodes in a dependency tree by only keeping the parent/children/grand-children of the predicate, the siblings of the predicates, and the children/grandchildren of the siblings. Maximum Entropy classifier will take features as input and determine the corresponding semantic role label.

[9], paper describes about semantic role labeling task (SRL-only) of the CoNLL-2009 shared task in the closed challenge. This system consists of a pipeline of independent, local classifiers that identify the predicate sense, the arguments of the predicates, and the argument labels. A beam search is carried out to generate a pool of candidates using these local models. After that the candidates are reranked using a joint learning approach that combines the local models and proposition features.

Table I. Semantic Role Labeling Summary (Language for which SRL developed, Dataset on which authors had perform evaluation, Accuracy of the proposed system).

Sr. No	Paper-Year	Language	Dataset	Accuracy (%)
1	[1], 2018	Malayalam	Web documents collected from online malayala manorma news paper pertaining to cricket domain.	• Avg precision: 84 • Avg recall:83.67 • Avg f-score: 83.33
2	[2], 2018	Spanish	CoNLL-2005 & 2012 Datasets	CoNLL-2005 • F1: 87.4 CoNLL-2012 • F1:87
3	[3], 2017	English	Proposition bank (propbank) corpus	• Precision: 97.6 • Recall:97.4 • F-Score: 97.3
4	[4], 2016	Hindi & Urdu	Data from Urdu propbank and Hindi propbank	Hindi Language • Precision: 58 • Recall: 42 Urdu Language • Precision: 83 • Recall: 80
5	[5], 2016	Multilingual(Arabic, Chinese, French, German, Hindi, Japanese, Russian and Spanish)	**Arabic, Chinese , French, Russian , Spanish**: The UN corpus of official United Nations documents. **German**: The Europarl corpus of European parliament proceedings and the OpenSubtitles corpus of movie subtitles. **Hindi**: The Hindencorp corpus automatically gathered from web sources. **Japanese**: The Tatoeba corpus of language learning.	Arabic Language Predicate • Precision: 97 • Recall:89 • F-Score: 93 Argument • Precision: 67 • Recall:63 • F-Score: 65 Chinese Language Predicate • Precision: 97 • Recall:88 • F-Score: 92 Argument • Precision: 83 • Recall:81 • F-Score:82 French Language Predicate • Precision: 95 • Recall:92 • F-Score: 94 Argument • Precision: 86 • Recall:74 • F-Score: 80 German Language Predicate Argument

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

				• Precision: 96 • Recall:92 • F-Score: 94 / • Precision: 91 • Recall:73 • F-Score: 81 Russian Language Predicate: • Precision: 96 • Recall:94 • F-Score: 95 Argument: • Precision: 79 • Recall:65 • F-Score: 72 Spanish Language Predicate: • Precision: 96 • Recall:93 • F-Score: 95 Argument: • Precision: 75 • Recall:72 • F-Score: 74 Hindi Language Predicate: • Precision: 91 • Recall:68 • F-Score: 78 Argument: • Precision: 58 • Recall:54 • F-Score: 56
6	[6], 2012	Chinese	Chinese PropBank (CPB)	• Precision: 87.12 • Recall:87.60 • F-Score: 87.36
7	[7], 2010	Bengali	The corpus from the ICON 2009 Dependency Parsing shared task.	• Avg F-score:68.10
8	[8], 2009	English	WSJ corpus supplied by CoNLL2008 shared task, the data is from PropBank and Nombank, including train, dev and test set.	1) Gold • Precision: 84.46 • Recall:84.84 • F-Score: 86.63 2) MaltParser • Precision: 77.11 • Recall: 73.12 • F-Score: 75.06 3) MSTParser • Precision: 83.49 • Recall: 80.50 • F-Score: 81.95
9	[9], 2009	Catalan,Spanish,Chinese ,German,Japanese,Czech	CoNLL- 2009 shared task for every language.	• Avg F1: 80.31

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

		& English		

ISSN: 2278-4632
Vol-10 Issue-7 No. 9 July 2020

III. Conclusions

We presented a review on Semantic Role Labelers for different languages. The review on SRL covers Indian languages like Malayalam, Hindi, Urdu, Bengali and other languages like English, Spanish, Arabic, Chinese, French, German, Japanese, Russian, Chinese, Catalan, Czech. We described about the algorithms and the techniques used by the authors. We have observed that the techniques used mainly include Supervised Learning. This requires a tremendous amount of annotated data which is not available easily for low resource south Asian languages like Hindi. Our further work will be focused on the development of a comprehensive system for SRL for Hindi Language.

References

[1] D. A. J. A. G. SUNITHA C, "SEMANTIC ROLE LABELING OF MALAYALAM WEB DOCUMENTS IN CRICKET DOMAIN," *Journal of Theoretical and Applied Information Technology,* p. 10, 2018.

[2] H. S. Y. M. Hiroki Ouchi, "A Span Selection Model for Semantic Role Labeling," 4 October 2018. [Online]. Available: https://arxiv.org/abs/1810.02245. [Accessed 17 January 2020].

[3] M. M. K. Ms. Neethu P H, "Semantic Role Labeling Using Ensemble Classifier," in *2017 International Conference on Intelligent Computing,Instrumentation and Control Technologies (ICICICT)*, Chemperi, 2017.

[4] D. S. Maaz Anwar, "Towards Building Semantic Role Labeler for Indian Languages," in *Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016.

[5] Y. L. Alan Akbik, "POLYGLOT: Multilingual Semantic Role Labeling with Unified Labels," in *54th Annual Meeting of the Association for Computational Linguistics—System Demonstrations*, Berlin, 2016.

[6] S. Y. L. C. Mao Ning, "Chinese Semantic Role Labeling Based On Genetic Algorithm," in *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 2012.

[7] A. G. a. S. B. Amitava Das, "Semantic Role Labeling for Bengali Noun using 5Ws," in *6th International Conference on Natural Language Processing and Knowledge Engineering(NLPKE-2010)*, Beijing, 2010.

[8] G. Z. P. Q. X. L. Hanxiao Shi, "Semantic Role Labeling based on dependency Tree with multi-features," in *2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing*, Washington, 2009.

[9] L. H. P. N. Anders Bj¨orkelund, "Multilingual semantic role labeling," in *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, Stroudsburg, 2009.

[10] M. S. S. a. O. E. Janara Christensen, "Semantic Role Labeling for Open Information Extraction," in *FAM-LbR '10: Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms and Methodology for Learning by Reading*, 2010.

[11] S. H. J. W. P. A. Mihai Surdeanu, "Using Predicate-Argument Structures for Information Extraction," in *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, 2003.

[12] D. G. Ding Liu, "Semantic Role Features for Machine Translation," in *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, Beijing, 2010.