

Semantic Role Labeling of English Tweets

Dwijen Rudrapal¹, Amitava Das²

¹ National Institute of Technology, Agartala,
India

² Indian Institute of Information Technology, Sri City, Chittoor,
India

{dwijen.rudrapal, amitava.santu}@gmail.com

Abstract. Semantic role labeling (SRL) is a task of defining the conceptual role to the arguments of predicate in a sentence. This is an important task for a wide range of tweet related applications associated with semantic information extraction. SRL is a challenging task due to the difficulties regarding general semantic roles for all predicates. It is more challenging for Social Media Text (SMT) where the nature of text is more casual. This paper presents an automatic SRL system for English tweets based on Sequential Minimal Optimization (SMO) algorithm. Proposed system is evaluated through experiments and reports comparable performance with the prior state-of-the-art SRL system.

Keywords. Social media text, tweet stream, semantic role labeling, tweet summarization.

1 Introduction

Twitter is a popular online social platform where instant responses of users help in discrimination of information to a large online community. However, all the contents posted in Twitter is not credible or informative about an event. Often a tweet published without proofreading. Thus, tweet includes wrong spelled words, user created acronym and grammatical mistakes which raise challenges towards semantic information extraction. Semantic role labeling (SRL) [4] is the best possible way to extract semantic information by identifying the abstract role to the arguments of predicate. These roles represent general semantic properties of the arguments in the sentence [26].

Different Natural Language Processing (NLP) applications like question answering, text summarization, information extraction systems etc are strongly governed by semantic relation between events and its constituents. The task of SRL is more challenging for English tweets than formal text. The reasons are: first, diverse nature of Twitter text creates several obstacles on the way of processing tweets. Users often misspell words either deliberately or accidentally by expanding (onk) and abbreviating words (prob, c), using lexical/numeric substitutions (b4, r8), nonstandard acronyms such as (lol, smh). User also use Hashtags, user tags, Twitter specific terminologies frequently. This writing style poses problems for standard NLP techniques to pre-process tweets. For example:

Tweet: *As #Iraqi , I think we should hv r8 2 vote in the #USelections since #US rule our country & they appoint the puppet Iraqi governemt.*

Second, classical SRL system defines arguments which are in cohesive structure with strong dependencies. But, for tweets this assumption is less viable due to ill-formed grammar. For example:

Tweet: *Much a do about nothing " FBI : Please Ignore All the Email Fuss . We Found Nothing New After All . "*

Third, the role of arguments to a predicate is restricted to a sentence. Thus, SRL task follows two important constraints [13], like, one predicate-argument pair role label in a sentence and each of the arguments can occurs at most once to the predicate.

But a tweet is a short document rather a sentence having multiple sentences. Multiple sentences have multiple predicates which pose difficulties towards semantic role extraction. For example, following English tweet comprises with multiple sentences even though proper punctuation is used.

Tweet: *Good Luck DSA (Disunited States of America) Tomorrow will be historical I just hope not hysterical ... #USelections*

In this paper, our main contribution is an automatic SRL system for English tweets based on SMO classifier. Our approach provides a potential solution to the problem of a tweet having multiple sentences and multiple predicate-argument pairs. We transformed a tweet into possible sentences [20] and identify semantic roles to the predicate in sentences.

The rest of our paper is organized as follows. Section 2 describes related work followed by Section 3 describing corpus preparation and annotation process. Section 4 describes proposed system and section 5 describes experiment setup along with result analysis. Finally, section 6 concludes the work with future scope.

2 Related Work

This section briefly reviews the promising research works related to semantic role labeling of English tweets as well as traditional text.

Gildea and Jurafsky [5] first introduced the task of semantic role labeling for English formal text. They proposed syntactic constituent tree based features for automatic labeling of predicate-argument relationships. Since the introduction, the research works on SRL for formal text has become a well defined problem and explored rich, expressive features. These features involve Predicate-Argument Structures [21, 23], dependency relations among arguments [16], verb-direct-object relations [15], relations of all arguments of the same predicate [22], dependency parser [7]. The work by Yang et al. [25] focuses on the features like several arguments and multiple predicates in a sentence for SRL task.

SRL system by [8] introduced a tensor-based approach to semantic role labeling. The approach investigate meaningful interactions between predicate and its role and compress each feature representation in to a lower dimensional space. A four-way low-rank tensor maintains associated parameters and optimized for the SRL task. The work [27] proposed a deep bi-directional recurrent network for SRL task which works on only original plain text rather than any syntactic information. Recent SRL research [2, 3, 9, 19] also focuses on neural models which automatically learn feature without using syntactic information of text.

Most recent work [6] proposed a unified neural model for SRL that utilizes contextual, syntactical and lexical semantic features. The model extract features by using bidirectional long short-term memory (LSTM)-based recurrent neural networks. An integer linear programming (ILP) procedure is applied to meet structural constraints in SRL task. The work introduced in [24] generalized huge set of features used for SRL task. The work embed lexical and syntactical information into features vector and clustered into similar semantic role using k-means clustering algorithm.

SRL task for informal text like tweets also equally attracted by the NLP researchers in last decade due to the popularity of social media communication. But the SRL task for social media text poses a number of challenges due to the nature of text. The system proposed in [10] first introduced the SRL task for domain specific English tweets, precisely, two categories of English news tweets. The categories are: news excerpt tweets, having a link to a news article and news tweets, having news related information only. Based on the conventional SRL features, authors propose a Conditional Random Field (CRF) learning framework for SRL system.

Another tweet SRL system [11] grouped similar tweets by K-means clustering process and train CRF classifier to improve the earlier system. Proposed approach used dependency parse tree related features in addition of the features used in the work [10]. The approach proposed in the work [14] developed an SVM classifier based SRL system to identify emotions in tweets.

The approach annotated tweet dataset for Experiencer, State, and Stimulus to identify different roles of emotions.

3 Data Collection and Annotation

3.1 Corpus Preparation and Annotation

We developed a corpus of English tweets collected on the event 2016 US presidential election during the period from August to December 2016 using Twitter4j¹. We used commonly known hashtags pertaining to the election, words like Donald Trump, Trump, Hillary, Hillary Clinton, because they are names of the two presidential candidates. After discarding re-tweets, non-English tweets, very short and long tweets having large percentage of misspelled words, we prepare a corpus of 21,000 English tweets.

For our current research work, we randomly select 1200 tweets covering both the stages of the election like pre-election and immediate post-election period. A tweet is not restricted to one sentence only. Rather, often a tweet includes multiple sentences having multiple predicates-argument pair which makes SRL task difficult. To defeat this challenging issue, we recognize possible sentences in a tweet by identifying the boundary of sentences as proposed in the work [20]. The sentence identification system successfully split 86% tweets into possible sentences. In few cases, system could not correctly transformed tweets into sentences. For example,

Tweet: *Good Luck DSA (Disunited States of America) Tomorrow will be historical I just hope not hysterical ... #USelections*

Sentence 01: *Good Luck DSA (Disunited States of America)*

Sentence 02: *Tomorrow will be historical*

Sentence 03: *I just hope not hysterical*

Sentence 04: *#USelections*

These exceptional tweets are manually transformed into possible sentences. The final dataset includes 18,207 nos. of tokens and 1,920 nos. of sentences for 1200 tweets.

¹<http://twitter4j.org/en/>

We agree to general semantic roles as defined in the Shared Tasks of CoNLL-2004 and CoNLL-2005² for manual SRL annotation task. The arguments roles to the Predicate (V) are Agent (A0), Patient (A1), Indirect object (A2), Attribute (A3), Modal verb (AM-MOD) and Negation (AM-NEG). We developed one web based SRL annotation tool for our research work. Using the tool, a user can select any tweet from the list to display sentences along with SRL tags to be tagged. Any number of consecutive tokens from the sentence can be selected to form a constituent and labeled by corresponding SRL tag.

We involved two human annotators who are native English speaker to annotate all the tweets in the dataset. After complete annotation task, we measured tag-wise inter-annotator agreement (IAA) Cohen's Kappa co-efficient [1] to assess the annotation task and prepare gold standard SRL tagged data. Detail statistics of IAA is shown in table 1. We also report distribution of various semantic roles in the table. The report shows that a major portion of the tokens do not carry any semantic role. During annotation process both the annotators faced some ambiguities which is discussed in section 3.2. After through discussion among the annotators 200 numbers of tweets are dropped due to in-consistent annotation and finally 1000 tweets are kept in the corpus forming the gold-standard dataset.

Table 1. Inter annotator agreement statistics

SRL Tags	Frequency (%)	IAA
V	7.70	0.91
A0	10.26	0.92
A1	28.66	0.90
A2	1.69	0.67
A3	7.02	0.65
AM-MOD	0.86	0.89
AM-NEG	0.54	0.83

3.2 Annotation Challenges

Manual annotation of semantic roles for English tweets raises certain ambiguities.

²<http://www.lsi.upc.es/srlconll/>

Most of these ambiguities are concerned with the main predicate identification and argument role boundary identification. In this section we briefly explain these ambiguities with proper examples.

1. **Identification of main predicate :** When a sentence includes multiple predicates, selection of main predicate makes ambiguous constituents for various semantic roles. For example,

Tweet: *FBI chief James Comey clears Hillary Clinton of wrong handling emails*
<https://t.co/rLIiKg0UEp/U>

was tagged by the two annotators as

- **Annotator 1:** [FBI chief James Comey/**A0**] [clears/**V**] [Hillary Clinton of wrong handling emails/**A1**]
<https://t.co/rLIiKg0UEp/U>
- **Annotator 2:** FBI chief James Comey clears [Hillary Clinton/**A0**] of wrong [handling/**V**] [emails/**A1**]
<https://t.co/rLIiKg0UEp/U>

In the above example, annotator 1 identified the role of the arguments depending on the predicate “clears” whereas annotator 2 identified the roles based on the predicate “handling”. To overcome from this kind of confusions, annotators were agreed to select the first occurred predicate as the main predicate in sentence and assigned concerned roles. For above example, annotation by annotator 1 is treated as correct semantic role labeled sentence.

2. **Identification of arguments boundaries:** The annotation task faced sufficient disagreement while identifying the boundary of constituents for argument roles. For example,

Tweet: *#DONALDTRUMP cries as FBI clears #HillaryClinton in emails saga on eve of #USElections*
<https://t.co/RZns8LGdt2>

Annotated as:

- **Annotator 1:** [#DONALDTRUMP/**A0**] [cries/**V**] [as FBI clears #HillaryClinton/**A1**] [in emails saga/**A2**]

[on eve of #USElections/**A3**]
<https://t.co/RZns8LGdt2>

- **Annotator 2:** [#DONALDTRUMP/**A0**] [cries/**V**] [as FBI clears #HillaryClinton/**A1**] [in emails saga on eve of #USElections/**A2**]
<https://t.co/RZns8LGdt2>

In the above example, annotator 1 identifies “in emails saga” as A2 and “on eve of #USElections” as A3, while annotator 2 identifies “in emails saga on eve of #USElections” as A2. Similar type of ambiguities is faced by the annotators while tagging other arguments like A1 and A3.

4 Proposed Approach

Our task is to recognize each semantic role represents by the constituents of a sentence in tweets. We followed word-by-word semantic parsing approach as proposed in [15]. In the following subsections, we described the proposed SRL system in details.

4.1 Feature Selection

We draw feature set for our SRL task from the work proposed in [5, 15, 24]. In this subsection, we elaborate features selection procedure.

- **Current word:** The current content word of the sentence in focus.
- **Part of Speech (POS) of the current word:** The POS category of the current word.
- **Lemma of the current word:** Root form of the current word.
- **Predicate:** The main verb in focus for which the argument role to be assigned.
- **Lemma of predicate:** Root word of the main verb.
- **Phrase type:** The syntactic category (NP, VB, etc.) of the constituent to be classified.

- **Phrase position:** The linear position of the current phrase based on IOB representation (as B-NP or I-NP or O)
- **Linear position:** The linear position of the current word with respect to the predicate (as “before” or “after”).
- **Path:** The path feature value for a word is uni-directional (from the word in focus to the predicate) and measured based on the flat tree concept explained in the work [24, 15]. The path is a chain of phrase chunk labels terminated with the POS tags of the word in focus and predicate. Consecutive chunks with identical labels are collapsed into one. We extract the path feature of each word using chunking parser for tweets [17, 18] instead of a full syntactic parser. For example, figure 1 shows the calculation of the path from the word “**F.B.I.**” to the predicate “**gets**”. The result path feature value is:

NNP → NP → PP → NP → VP → VBZ

Word	POS	Phrase
Clinton	NNP	B-NP
gets	VBZ	B-VP
clean	JJ	B-NP
chit	NN	I-NP
by	IN	B-PP
the	DT	B-NP
F.B.I.	NNP	I-NP
U.S.A	NNP	I-NP

Fig. 1. Illustration of path feature value extraction

- **Head word:** Head word is the most essential part to the meaning of the phrase. Based on the chunking parser output, we retrieve the head word of each constituent.

Features like predicate, part of speech of the current word, phrase type and phrase position extracted using Ritter’s twitter_nlp tool ³ proposed in the work [17, 18]. The feature like current word

³https://github.com/aritter/twitter_nlp

and word’s linear position extracted directly from the sentence. Lemmatization of current word and predicate is done using Stanford Lemmatizer [12].

4.2 Classifier Selection

Prior SRL techniques for traditional text are mostly based on Support Vector Machine (SVM) [15] and Maximum Entropy classifier [23, 24]. However, Conditional Random Field (CRF) classifier [10, 11] and SVM classifier [14] are mostly used in SRL task for tweets. In our research work, we exploited three classifiers like BayesNet, Logistic regression and Sequential Minimal Optimization (SMO) classifier.

5 Experiment and Result Analysis

Our proposed approach is evaluated in two fold experiments. In one fold, we evaluated our approach on English tweets after transforming into sentences. We experimented with BayesNet, Logistic regression and SMO classifier in WEKA 3.8 machine learning tool ⁴. A detail result of ten-fold cross-validation for each experiment is shown in table 2. Result shows that SMO classifier performs better for all the semantic roles and shows the best average F-measure of 59.76. The result also reports that identification of semantic role for argument A2 and argument A3 is more challenging than identifying argument A0 and A1.

Table 2. Performance evaluation of proposed system

SRL tags	F1 Score		
	BayesNet	Logistic Regrsn.	SMO
A0	61.20	56.10	66.72
A1	65.60	64.36	71.10
A2	29.90	28.00	44.34
A3	51.10	42.20	54.20
AM-MOD	58.63	53.54	62.76
AM-NEG	44.92	35.96	59.42
Avg.	51.89	46.69	59.76

In another fold, we evaluated our system performance on same English tweets without

⁴<http://www.cs.waikato.ac.nz/ml/weka/>

Table 3. Performance comparison of sentence-focused and tweet-focused SRL approach

Approach	A0	A1	A2	A3	AM-MOD	AM-NEG
Sentence-Focused	66.72	71.10	44.34	54.20	62.76	59.42
Tweet-Focused	54.70	59.40	36.10	38.00	62.50	50.00
Deviation	12.02	11.70	08.24	16.20	0.26	09.42

transforming into sentences. Performance result (F1 score) of each SRL tag is reported in table 3. Comparative analysis of the performances reveals that, sentence-focused SRL approach for tweets does outperform over tweet-focused SRL approach. The deviation of F1-score for each semantic role is reported in the table 3. We could not compare our system performance with existing SRL approaches on tweets, due to un-availability of SRL system and dataset.

Detail analysis of experiment results reveals that, F1 score of Argument role A2 and A3 recognition are lower than the other semantic roles. Proposed approach recognizes A2 and A3 semantic roles with F1 score of 44.34 and 54.20 respectively. This performance justified that the identification of these roles for English tweets are more challenging than other semantic roles. The performance of AM-MOD and AM-NEG role recognition for English tweets are also challenging. The reason may be due to the fact that writing negation combined with modal verb or pronoun is a common practice in social media. This writing style creates ambiguities to identify the specific role of AM-MOD and AM-NEG. For example,

Tweet: *I haven't forgiven #Bush voters yet . Lets see what happens tomorrow #USElections #elections2016*

6 Error Analysis

In-depth analysis of the system performance reveals that, most of the errors made by our system are due to the sparse nature of tweet text. Some of the errors are due the fact that,

1. Presence of conjunction in sentence increases the existence of multiple predicates and multiple predicate-argument pair. For example,

Tweet: *The #FBI clears #Hillary a day before the #USElections and suddenly #Trump's allegations of a **rigged** system don't sound so moronic any more.*

2. The use of modal verbs with negation or pronoun is very ambiguous in tweets. The writing style makes difficult to identify the role of modal verb or negation. For example:

Tweet: *I haven't forgiven #Bush voters yet .*

Tweet: *That's what happenend 4 years ago . Lets see what happens tomorrow #USElections #elections2016.*

7 Conclusion

The task of SRL for English tweets is challenging, since tweets are often too short, informal and do not provide sufficient semantic information. In this research work, we proposed an automatic SRL system based on SMO classifier. We conducted experiments on a SRL annotated English tweet dataset and showed that the proposed system can achieve an absolute F1 score of 59.76% which is a comparable performance over earlier SRL research on tweets.

Twitter specific more features may be incorporated into the proposed SRL task for more stable and accurate system as future scope of the work. Another future scope of this work is to make the system applicable for code-mixed social media text.

References

1. **Cohen, J. (1960).** A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37.

2. **FitzGerald, N., Täckström, O., Ganchev, K., & Das, D. (2015).** Semantic role labeling with neural network factors. *EMNLP*, pp. 960–970.
3. **Foland, W. & Martin, J. H. (2015).** Dependency-based semantic role labeling using convolutional neural networks. * *SEM@ NAACL-HLT*, pp. 279–288.
4. **Gildea, D. & Jurafsky, D. (2002).** Automatic labeling of semantic roles. *Computational linguistics*, Vol. 28, No. 3, pp. 245–288.
5. **Gildea, D. & Jurafsky, D. (2002).** Automatic labeling of semantic roles. *Comput. Linguist.*, Vol. 28, No. 3, pp. 245–288.
6. **Guo, J., Che, W., Wang, H., Liu, T., & Xu, J. (2016).** A unified architecture for semantic role labeling and relation classification. *COLING*, pp. 1264–1274.
7. **Johansson, R. & Nugues, P. (2008).** Dependency-based semantic role labeling of propbank. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 69–78.
8. **Lei, T., Zhang, Y., Moschitti, A., Barzilay, R., et al. (2015).** High-order low-rank tensors for semantic role labeling. *In proceedings of the North American Chapter of The Association for Computational Linguistics–Human Language Technologies (NAACL-HLT 2015)*, Association for Computational Linguistics, pp. 1150–1160.
9. **Li, T. & Chang, B. (2015).** Semantic role labeling using recursive neural network. *In Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*. Springer, pp. 66–76.
10. **Liu, X., Li, K., Han, B., Zhou, M., Jiang, L., Xiong, Z., & Huang, C. (2010).** Semantic role labeling for news tweets. *Proceedings of the 23rd International Conference on Computational Linguistics*, Association for Computational Linguistics, pp. 698–706.
11. **Liu, X., Li, K., Zhou, M., & Xiong, Z. (2011).** Collective semantic role labeling for tweets with clustering. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, AAAI Press, pp. 1832–1837.
12. **Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014).** The Stanford CoreNLP natural language processing toolkit. *Association for Computational Linguistics (ACL) System Demonstrations*, pp. 55–60.
13. **Meza-Ruiz, I. & Riedel, S. (2009).** Jointly identifying predicates, arguments and senses using markov logic. *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, pp. 155–163.
14. **Mohammad, S., Zhu, X., & Martin, J. D. (2014).** Semantic role labeling of emotions in tweets. *WASSA@ Association for Computational Linguistics*, pp. 32–41.
15. **Pradhan, S., Hacioglu, K., Krugler, V., Ward, W., Martin, J. H., & Jurafsky, D. (2005).** Support vector learning for semantic argument classification. *Machine Learning*, Vol. 60, No. 1, pp. 11–39.
16. **Punyakanok, V., Roth, D., Yih, W.-t., & Zimak, D. (2004).** Semantic role labeling via integer linear programming inference. *Proceedings of the 20th international conference on Computational Linguistics*, Association for Computational Linguistics, pp. 1346.
17. **Ritter, A., Clark, S., Mausam, & Etzioni, O. (2011).** Named entity recognition in tweets: An experimental study. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 1524–1534.
18. **Ritter, A., Mausam, Etzioni, O., & Clark, S. (2012).** Open domain event extraction from twitter. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 1104–1112.
19. **Roth, M. & Lapata, M. (2015).** Context-aware frame-semantic role labeling. *Transactions of the Association for Computational Linguistics*, Vol. 3, pp. 449–460.
20. **Rudrapal, D., Jamatia, A., Chakma, K., Das, A., & Gambäck, B. (2015).** Sentence boundary detection for social media text. *Proceedings of the 12th International Conference on Natural Language Processing*, NLP Association of India, Trivandrum, India, pp. 254–260.
21. **Surdeanu, M., Harabagiu, S., Williams, J., & Aarseth, P. (2003).** Using predicate-argument structures for information extraction. *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, Association for Computational Linguistics, pp. 8–15.
22. **Toutanova, K., Haghighi, A., & Manning, C. D. (2008).** A global joint model for semantic role

labeling. *Computational Linguistics*, Vol. 34, No. 2, pp. 161–191.

23. **Xue, N. & Palmer, M. (2004)**. Calibrating features for semantic role labeling. *EMNLP*, pp. 88–94.
24. **Yang, H. & Zong, C. (2016)**. Learning generalized features for semantic role labeling. *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 15, No. 4, pp. 28.
25. **Yang, H., Zong, C., et al. (2014)**. Multi-predicate semantic role labeling. *EMNLP*, pp. 363–373.
26. **Zhang, Y., Jiang, M., Wang, J., & Xu, H. (2016)**. Semantic role labeling of clinical text: Comparing syntactic parsers and features. *AMIA Annual Symposium Proceedings*, volume 2016, American Medical Informatics Association, pp. 1283.
27. **Zhou, J. & Xu, W. (2015)**. End-to-end learning of semantic role labeling using recurrent neural networks. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, pp. 1127–1137.

Article received on 10/01/2018; accepted on 05/03/2018.
Corresponding author is Dwijen Rudrapal.