

A Dual-Layer Semantic Role Labeling System

Lun-Wei Ku

Institute of Information Science
Academia Sinica, Taiwan
lwku@iis.sinica.edu.tw

Shafqat Mumtaz Virk

Institute of Information Science
Academia Sinica, Taiwan
virk.shafqat@gmail.com

Yann-Huei Lee

Institute of Information Science
Academia Sinica, Taiwan
andycyrus@gmail.com

Abstract

We describe a well-performed semantic role labeling system that further extracts concepts (smaller semantic expressions) from unstructured natural language sentences language independently. A dual-layer semantic role labeling (SRL) system is built using Chinese Treebank and Propbank data. Contextual information is incorporated while labeling the predicate arguments to achieve better performance. Experimental results show that the proposed approach is superior to CoNLL 2009 best systems and comparable to the state of the art with the advantage that it requires no feature engineering process. Concepts are further extracted according to templates formulated by the labeled semantic roles to serve as features in other NLP tasks to provide semantically related cues and potentially help in related research problems. We also show that it is easy to generate a different language version of this system by actually building an English system which performs satisfactory.

1 Introduction

Semantic roles are utilized to find concepts automatically and assure their meaningfulness. Semantic role labeling is a research problem which finds in a given sentence the predicates and their arguments (identification), and further labels the semantic relationship between predicates and arguments, that is, their semantic roles (classification). There are several labeling sets. Researchers have widely adopted the semantic role labels defined in Propbank (Bonial *et al.*, 2010) like predicate (PRED), numbered arguments 0 to 5 (ARG0, ARG1, ARG2, ARG3, ARG4, ARG5), or modifier arguments (ARGM-X); finer labels are those defined in Sinica Treebank (Huang *et al.*, 2000) like agent, theme, target, which are labeled on each

node of the parse tree; those defined in FrameNet (Ruppenhofer *et al.*, 2006) are the finest but most expressive. Each set provides semantic information. As long as the semantic relationship between terms derives from their semantic role labels, we are able to determine whether they should be extracted from the current sentence to construct a concept.

The word *concept* usually refers to an abstract or general idea inferred or derived from specific instances. Therefore, the extraction of concepts from text is often defined as extracting terms that are in some way related to one another. These terms could be predefined by people in resources such as ontologies, or they could be typical words in texts. In this paper, we view concepts as the continuous or discontinuous meaningful units in a sentence and hence they are tightly related to semantic roles. We propose a dual-layer semantic role labeling system which provides extracted concepts according to the reported labels, and then demonstrate the functions of this system. Experimental results will show the merit of the proposed framework.

2 Related Work

Previous studies related to this work can be divided into two groups: semantic role labeling and concept extraction. Semantic role labeling (SRL) has sparked much interest in NLP (Shen and Lapata, 2007; Liu and Gildea, 2010). The first automatic SRL systems were reported by Gildea and Jurafsky in 2002 (Gildea and Jurafsky 2002); since then, their ideas have dominated the field. In their approach, they emphasize the selection of appropriate lexical and syntactical features for SRL, the use of statistical classifiers and their combinations, and ways to handle data sparseness. Many researchers have tried to build on their work by augmenting and/or altering the feature set (Xue 2004), by experimenting with various classification approaches (Pradhan *et al.* 2004; Park and Rim 2005), and by attempting different ways to handle data sparseness

(Zapirain, Agirre, and Márquez 2007). Moreover, some researchers have tried to extend it in novel ways. For example, Ding and Chang (2008) used a hierarchical feature selection strategy, while Jiang, Li, and Ng (2005) proposed exploiting argument interdependence, that is, the fact that the semantic role of one argument can depend on the semantic roles of other arguments.

Many researchers have tried to extract concepts from texts (Gelfand *et al.*, 1998; Hovy *et al.*, 2009; Villalon and Calvo, 2009; Dinh and Tamine, 2011; Torii *et al.*, 2011). Hovy narrowed the domain of interest into concepts “below” a given seed term. Villalon and Calvo extract concepts from student essays for concept map mining, which generates a directed relational graph of the extracted concepts in an essay. For specific domains, biological or medical concepts are of greatest interest to researchers (Jonnalagadda *et al.*, 2011). Two relatively new and related approaches are the Concept parser (Rajagopal *et al.* 2013), a part of the SenticNet project (Cambria, Olsher, and Rajagopal 2014) and ConceptNet (Liu and Singh 2004). The former is a tool to decompose unrestricted natural language text to a bag of concepts, which is similar to our work. However, in the final phase a seman-

tic knowledge base is used to express a concept in all its different forms and their concept-parser does not use any semantic knowledge during decomposition. The latter is a semantic network based on the Open Mind Common Sense (OMCS) knowledge base. As it is a knowledge base, its construction process is quite different from the work described here of automatically extracting concepts from sentences.

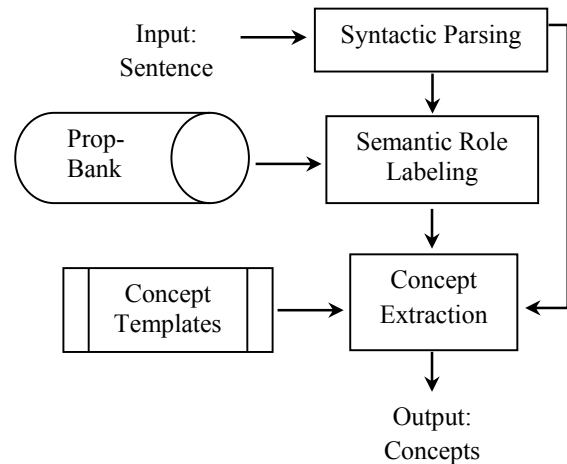


Figure 1: System Framework.

Enter any sentence:
2010年Google公司因内容审查问题与中国政府交涉，并最终关闭中国版网页搜寻服务。

Process

Parse Tree:

```
(ROOT (IP (NP (NT 2010年)) (NP (NN Google) (NN 公司)) (VP (VP (PP (P 因) (NP (NN 内容) (NN 审查) (NN 问题))) (PP (P 与) (NP (NP (NR 中国)) (NP (NN 政府)))) (VP (VV 交涉))) (PU .) (CC 并) (VP (ADVP (AD 最终)) (VP (VV 关闭) (NP (NP (NP (NR 中国)) (NP (NN 版) (NN 网页) (NN 搜寻))) (NP (NN 服务)))))) (PU 。)))
```

Predicates & Arguments:

Predicates	Arguments
交涉	ARG0: Google_公司
关闭	ARG0: Google_公司 ARG1: 中国_版_网页_搜寻_服务

Concepts:

Google_公司_交涉
Google_公司_关闭
关闭_中国_版_网页_搜寻_服务

Enter any sentence:
We bought a birthday cake from Super Market, and celebrated David's birthday.

Process

Parse Tree:

```
(ROOT (S (NP (PRP We)) (VP (VP (VBD bought) (NP (DT a) (NN birthday) (NN cake)) (PP (IN from) (NP (NNP Super) (NNP Market)))) (, ,) (CC and) (VP (VBD celebrated) (NP (NP (NNP David) (POS 's)) (NN birthday)))) (. .)))
```

Predicates & Arguments:

Predicates	Arguments
bought	ARG1: a_birthday_cake ARG2-from: Super_Market ARG0: We
celebrated	ARG1: David_'s_birthday ARG0: We

Concepts:

a_birthday_cake
We_celebrated
bought_a_birthday_cake
celebrated_David_'s_birthday
Super_Market
bought_Super_Market
David_'s_birthday
We_bought

Figure 2: System Interface (Chinese example sentence: In 2010, Google company negotiated with the China government on the issue of results censoring, and eventually shut down the web search service.)

3 System

The proposed system includes three major components: a syntactic parser, a semantic role labeler, and a concept formulation component. The framework is shown in Figure 1. The input sentence is first transformed into a syntactic parse tree through a syntactical analysis step that almost all automatic semantic role labeling systems require (Johansson and Nugues 2008). Here the Stanford parser (Klein and Manning 2003) is utilized. Figure 2 shows the system interface. The left part is the English system and the right part is the Chinese system. After users input a sentence, the system will automatically parse, label semantic roles and report the related concepts for it.

3.1 Semantic Role Labeling

To develop a SRL system, a total of 33 features including features related to the head word related features, target word related features, grammar related features, and semantic type related features, are collected from related work (Xue, 2008; Ding and Chang, 2008; Sun and Jurafsky 2004; Gildea and Jurafsky 2002). Then the baseline maximum entropy system is developed using these features (Manning and Schutze, 1999). Two sets of data – Chinese Treebank 5.0 together with Propbank 1.0 and Chinese Treebank 6.0 with Propbank 2.0 – are separated into the training and testing sets, and are then used to build models to identify and classify semantic labels, and also to evaluate the performance, respectively. As Chinese data was selected for experiments, the hypernyms of words from E-HowNet¹, a Chinese word ontology, are utilized as the semantic type of words. When applying the whole system on data in other languages, for major languages it is not difficult to find resources to obtain hypernyms. For minor languages, it is fine to just ignore these features. According to our experience, this will yield F-Score reductions of only 1% to 2%.

We further exploit argument interdependence to enhance performance by the dual-layer framework shown in Figure 2. Suppose for any given predicate P in a sentence, the system has identified the three potential arguments A1, A2, and A3 of the predicate. Next, to predict the semantic role labels of those three arguments, a critical observation made by (Jiang, Li, and Ng

2005) is that the semantic roles of arguments may depend on each other; this phenomenon is known as argument interdependence. A common way to escape argument interdependence is to adopt sequence labeling, and use the features extracted from the arguments around the current argument together with the features of the current one to predict the label for the current argument. For example, while predicting the label of argument A2, features extracted from arguments A1 and A3 are also used. Although window sizes can be used to set the scope of this interdependence, the window-size strategy has some practical limits: the typically large feature set necessitates the use of smaller window sizes (a window size of $[-1,1]$ is common). However, small window sizes can make it impossible to capture long dependency phenomena.

To overcome the limitations of the window-size strategy, we use all the surrounding arguments' predicted labels – window size $[-\infty, \infty]$, as opposed to their features – to predict the label of the current node. This also conforms to the rule that when a role is taken by the other argument, it is less likely that the current argument is of the same role. We implement this idea using the dual-layer classification framework shown in Figure 3.

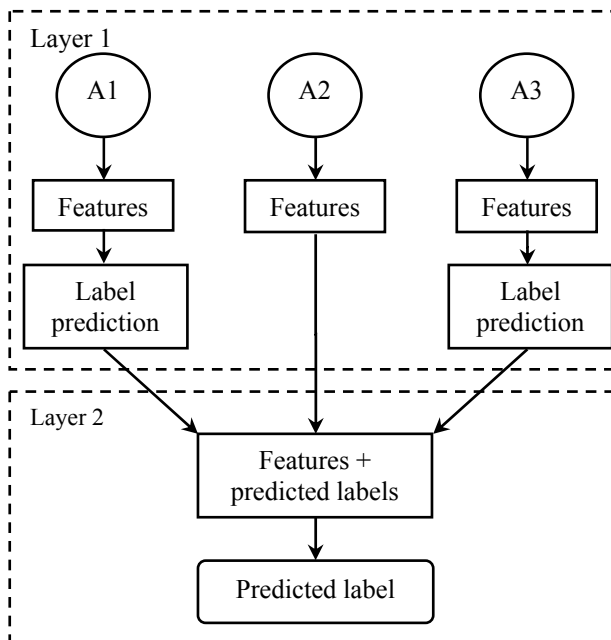


Figure 3: SRL classification framework.

In layer 1 the baseline system is used to predict the labels for identified nodes. Then in layer 2, these predicted labels of all surrounding arguments (in this example, A1 and A3) together with other features of the current node (A2) are used

¹ <http://ckip.iis.sinica.edu.tw/CKIP/conceptnet.htm>

to predict the label of the current node. Note as this approach is under no window size limitation, the labels of all arguments under the same predicate are taken into account. Experimental results show that this strategy works better than the window-size strategy. Table 1 shows the system accuracies for the single- and dual-layer frameworks. The predicted dual-layer framework utilized the SRL labels predicted in layer 1, while the gold dual-layer framework used as features the gold SRL labels of the surrounding arguments.

System	Accuracy
Ding and Chang, 2008 (state of the art)	94.68
Single-layer framework	94.60
Dual-layer framework (predicted)	94.86
Dual-layer framework (gold)	95.40

Table 1. Accuracy of SRL classification phase.

To further evaluate the performance of the proposed system and offer comparisons, we applied it on Chinese Treebank 6.0 with Propbank 2.0 in the same way as in the CoNLL 2009 SRL-only task data according to the information provided by the CoNLL organizers. Table 2 shows the results of the proposed system. Table 3 further shows the performance of the best systems in CoNLL 2009.

	Identification	Classification	SRL
Precision	94.38	90.22	86.89
Recall	96.24		80.11
F-Score	95.30		83.36
Accuracy	97.92		96.25

Table 2. SRL results on Propbank 2.0.

System name	Type	Score
Nugus (Björkelund <i>et al.</i> , 2009)	Closed challenge, SRL-only	78.50 (F-Score)
Meza-Ruiz (Meza-Ruiz and Riedel, 2009)	Closed challenge, SRL-only	82.66 (Precision)
Täckström (Täckström, 2009)	Closed challenge, SRL-only	79.31 (Recall)
Che (Che <i>et al.</i> , 2009)	Open challenge, Joint Task	76.42 (F-Score)

Table 3. CoNLL 2009 SRL performance².

The CoNLL 2009 task builds dependency-based SRL systems, while the proposed system works on the constituent-based parsing trees. Also the settings of the proposed system are not

all the same as the CoNLL 2009 SRL systems. In CoNLL 2009, as noted in Table 5, participants can participate in open or closed challenges, and can choose whether they want to attempt both syntactic and semantic labeling tasks (joint task) or only to attempt the SRL task. The setting of the proposed system is open challenge, SRL-only, while researchers working on the Chinese data selected only two other different settings: closed challenge, SRL only and open challenge, joint task. However, Table 5 shows that the proposed system outperforms the CoNLL 2009 best systems in terms of precision (86.89 vs. 82.66), recall (80.11 vs. 79.31), and f-score (83.36 vs. 78.50). Moreover, lately, dependency-based SRL has shown advantages over constituent-based SRL (Johansson and Nugues, 2008); thus we expect to show better results if working on dependency-based parsed data. Therefore, we believe the proposed system is comparable or even superior to other systems.

3.2 Concept-Formulations

Once the sentence has been annotated semantically, the concepts are formulated by concept templates designed according to Propbank SRL labels. Propbank provides semantic role labels of two types. One type is numbered arguments Arg0, Arg1, and so on until Arg5; the other type is modifiers with function tags, which give additional information about when, where, or how the event occurred. Tables 4 and 5 list the descriptions of the Propbank arguments utilized for the concept template generation. Table 6 then lists the generated concept templates.

As shown in Table 6, the predicate and its arguments are placed in various orders to build a list of concepts according to their semantic roles. These role combinations serve as templates which can capture a complete and important piece of information described in one sentence to form a concept. Additionally, the arguments (i.e., the subjects and objects of the predicate) in themselves can represent useful concepts, and for this reason, the arguments alone are also included in extracted concepts. For comparison, in Table 7 the extracted concepts are listed with those from the SenticNet concept parser.

² <http://ufal.mff.cuni.cz/conll2009-st/results/results.php>

Numbered Argument	Description
Arg0	agent, causer, experiencer
Arg1	theme, patient
Arg2	instrument, benefactive, attribute
Arg3	starting point, benefactive, attribute
Arg4	ending point
Arg5	Direction

Table 4. Propbank numbered arguments.

Modifier	Desc	Modifier	Desc
ArgM-LOC	Location	ArgM-COM	Comitative
ArgM-TMP	Time	ArgM-DIR	Direction
ArgM-GOL	Goal	ArgM-EXT	Extent
ArgM-MNR	Manner	ArgM-NEG	Negation
ArgM-CAU	Cause	ArgM-PRP	Purpose

Table 5. Propbank modifier arguments.

#	Concept Template
1	ARG0 Pred
2	Pred ARG1
3	Pred ARG1 ARG2
4	Pred ARG1 ARG2 ARG3
5	Pred ARG1 ARG2 ARG3 ARG4
6	Pred ARG1 ARG2 ARG3 ARG4 ARG5
7	Pred with ARGM-COM
8	Pred in ARGM-LOC
9	Pred in order to ARGM-PRP
10	Pred in the direction ARGM-DIR
11	Pred because ARGM-CAU
12	Pred when ARGM-TMP
13	Pred ARGM-GOL
14	Pred by ARGM-EXT
15	Pred ARGM-MNR
16	Pred ARGM-NEG
17	ARGX's
18	ARGM's

Table 6. Concept templates.

Proposed System
a_birthday_cake, bought_Super_Market, bought_a_birthday_cake, Super_Market, celebrated_David's_birthday, We_bought, David's_birthday, We_celebrated
SenticNet Concept Parser
birthday_cake, birthday_from_market, buy_birthday_cake, birthday_cake, birthday_david, buy_from_market, super_market, celebrate_david

Table 7. Concepts generated by the proposed system and the SenticNet Concept Parser.

4 Conclusion

We have presented a system to decompose a sentence into a set of concepts through the proposed well-performed semantic role labeling system (<http://doraemon.iis.sinica.edu.tw/srl-concept/>), which differs from previous related attempts. We demonstrated that this dual-layer semantic role labeling framework that exploits argument inter-dependence performs slightly better than the state of the art, and that it is relatively simple as no feature selection or engineering processes are required. We easily generated another English system under the same framework, which showcased the language independency of the system. In addition, it reached an F-Score 0.84, which was considered satisfactory. In the future, we plan to investigate how to further represent and utilize these extracted concepts efficiently in more NLP tasks which call for deep language understanding.

Acknowledgement

Research of this paper was partially supported by National Science Council, Taiwan, under the contract NSC101-2628-E-224-001-MY3.

References

- Björkelund, A., Hafdel, L., & Nugues, P. 2009. Multilingual semantic role labeling. In Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, 43-48.
- Bonial, C.; Babko-Malaya, O.; Choi, J. D.; Hwang, J.; and Palmer, M. 2010. Propbank annotation guidelines. Center for Computational Language and Education Research Institute of Cognitive Science University of Colorado at Boulder.
- Cambria, E.; Olsher, D.; and Rajagopal, D. 2014. Senticnet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. In Proceedings of AAAI, 1515-1521.
- Che, W., Li, Z., Li, Y., Guo, Y., Qin, B., & Liu, T. 2009. Multilingual dependency-based syntactic and semantic parsing. In Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, 49-54.
- Dinh, D., & Tamine, L. 2011. Biomedical concept extraction based on combining the content-based and word order similarities. In Proceedings of the 2011 ACM Symposium on Applied Computing, 1159-1163. ACM.
- Gelfand, B., Wulfekuler, M., & Punch, W. F. 1998. Automated concept extraction from plain text.

- In AAAI 1998 Workshop on Text Categorization, 13-17.
- Gildea, D., and Jurafsky, D. 2002. Automatic labeling of semantic roles. *Comput. Linguist.* 28(3):245–288.
- Hovy, E., Kozareva, Z., & Riloff, E. 2009. Toward completeness in concept extraction and classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, 948-957.
- Huang, C. R., Chen, F. Y., Chen, K. J., Gao, Z. M., & Chen, K. Y. (2000, October). Sinica Treebank: design criteria, annotation guidelines, and on-line interface. In *Proceedings of the second workshop on Chinese language processing: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 12*, 29-37.
- Jiang, Z. P.; Li, J.; and Ng, H. T. 2005. Semantic argument classification exploiting argument interdependence. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence, IJCAI'05*, 1067–1072.
- Johansson, R., & Nugues, P. 2008. The effect of syntactic representation on semantic role labeling. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, 393-400.
- R. Johansson and P. Nugues. 2008. Dependency-based semantic role labeling of PropBank. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*.
- Jonnalagadda, S., Cohen, T., Wu, S., & Gonzalez, G. 2012. Enhancing clinical concept extraction with distributional semantics. *Journal of biomedical informatics*, 45(1), 129-140.
- Klein, D., and Manning, C. D. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1, ACL '03*, 423–430.
- D. Liu and D. Gildea. 2010. Semantic role features for machine translation. In *Proceedings of the 23rd International Conference on Computational Linguistics*.
- Liu, H., and Singh, P. 2004. Conceptnet: A practical commonsense reasoning toolkit. *BT TECHNOLOGY JOURNAL* 22:211–226.
- Manning, Christopher D. and Schütze, Hinrich. 1999. *Foundations of statistical natural language processing*, Cambridge, Mass.: MIT Press.
- Meza-Ruiz, I., & Riedel, S. 2009. Multilingual semantic role labelling with markov logic. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, 85-90.
- Park, K.-M., and Rim, H.-C. 2005. Maximum entropy based semantic role labeling. In *Proceedings of the Ninth Conference on Computational Natural Language Learning, CONLL '05*, 209–212.
- Pradhan, S.; Ward, W.; Hacioglu, K.; and Martin, J. H. 2004. Shallow semantic parsing using support vector machines. In *Proceedings of the Conference on the Human Language Technologies and North American Association for Computational Linguistics (HLT-NAACL 2004)*, 233–240.
- Rajagopal, D.; Cambria, E.; Olsher, D.; and Kwok, K. 2013. A graph-based approach to commonsense concept extraction and semantic similarity detection. In *Proceedings of the 22Nd International Conference on World Wide Web Companion, WWW '13 Companion*, 565–570.
- Ruppenhofer, J., Ellsworth, M., Petrucci, M. R., Johnson, C. R., & Scheffczyk, J. (2006). *FrameNet II: Extended theory and practice*.
- D. Shen and M. Lapata. 2007. Using semantic roles to improve question answering. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and on Computational Natural Language Learning*.
- Täckström, O. 2009. Multilingual semantic parsing with a pipeline of linear classifiers. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, 103-108.
- Torii, M., Waghlikar, K., & Liu, H. 2011. Using machine learning for concept extraction on clinical documents from multiple data sources. *Journal of the American Medical Informatics Association, amiajnl-2011*.
- Villalon, J., & Calvo, R. A. 2009. Concept extraction from student essays, towards concept map mining. In *Proceedings of Ninth IEEE International Conference on Advanced Learning Technologies*, 221-225.
- Xue, N. 2004. Calibrating features for semantic role labeling. In *Proceedings of EMNLP 2004*, 88–94.
- Xue, N. 2008. Labeling chinese predicates with semantic roles. *Comput. Linguist.* 34(2):225–255.
- Zapirain, B.; Agirre, E.; and Màrquez, L. 2007. Ubcupc: Sequential srl using selectional preferences: An approach with maximum entropy markov models. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval '07*, 354–357.