

Improving Japanese semantic-role-labeling performance with transfer learning as case for limited resources of tagged corpora on aggregated language

Takuya Okamura^{*}, Koichi Takeuchi^{*}, Yasuhiro Ishihara^{*},
 Masahiro Taguchi[†], Yoshihiko Inada[‡], Masaya Iizuka[§], Tatsuhiko Abo[¶], Hitoshi Ueda[¶]
^{*}Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan
 t-okamura@s.okayama-u.ac.jp, koichi@cl.cs.okayama-u.ac.jp
[†]Graduate School of Humanities and Social Science, Okayama University, Okayama, Japan
[‡]Graduate School of Education, Okayama University, Okayama, Japan
[§]Institute for Education and Student Services, Okayama University, Okayama, Japan
[¶]Graduate School of Natural Science and Technology, Okayama University, Okayama, Japan

Abstract

In this paper we proposed the use of effective features and transfer learning to improve the accuracies of neural-network-based models for accurate semantic role labeling (SRL) of Japanese, which is an aggregated language. We first reveal that the final morphemes in each argument, which have not been discussed in previous work on English SRL are effective features in determining semantic role labels in Japanese. We then discuss the possibility of using large-scale training corpora annotated with different semantic labels from the target semantic labels by transfer learning on CNN, 3-LNN, and GRU models. The experimental results of Japanese SRL on the proposed models indicate that all of the neural-network-based models performed better with transfer learning as well as using the feature vectors of final morphemes in each argument.

1 Introduction

Several studies on semantic role labeling (SRL) have been conducted, mainly for English texts with a wide coverage from revealing syntactic and grammatical features that impact SRL decisions (Gildea and Jurafsky, 2002) to end-to-end models without syntactic inputs. This is because current rich language resources are related to annotated corpora with semantic roles such as PropBank (Kingsbury et al., 2002), FrameNet (Baker et al., 1998), and shared tasks of CoNLL 2005 (Carreras and Màrquez, 2005), 2009 (Hajič et al., 2009), and 2012 (Pradhan et al., 2012).

Instead of English SRL, most studies (Taira et al., 2008; Imamura et al., 2009; Sasano and Kurohashi, 2011; Hayashibe et al., 2011; Ouchi et al., 2015; Shibata et al., 2016; Ouchi et al., 2017; Matsubayashi and Inui, 2018) on SRL for Japanese texts, the grammar of which is quite different from English, have been focused on detecting three types of case-marker-based labels, i.e., nominative, accusative, and dative. This is because large-scale corpora annotated with these three labels have been developed (Kyoto Corpus (Kawahara et al., 2002) and NAIST Text Corpus (Iida et al., 2007)) and widely used¹. There are, however, annotated corpora with semantic tags corresponding to semantic role labels (EDR corpus², BCCWJ-PT (Takeuchi et al., 2015) and Japanese FrameNet (Ohara et al., 2011)) or semantic categories (GDA corpus³) that contain semantic relations as cause and location. Thus, by constructing a Japanese SRL system using these resources, we can discuss the effective approaches and models for a situation without standard semantic-role-labeled corpora as well as effective features for Japanese SRL.

We address the following two issues: what is an effective grammatical feature for Japanese SRL, and what is a model when there are no standard labeled corpora for Japanese SRL. We first argue that the last two morphemes of arguments have

¹In CoNLL 2009 (Hajič et al., 2009), annotated corpora with semantic roles in multiple languages including Japanese were used; however, most semantic tags in the Japanese corpus were case-marker-based relations, which are different from semantic roles annotated in PropBank.

²<http://www2.nict.go.jp/ipp/EDR/ENG/indexTop.html>

³<http://www.gsk.or.jp/catalog/gsk2009-b/>

an effect on determining SRL. This is because Japanese grammar allows scrambling (Saito, 1989); thus, case markers and functional expressions located at the last part of arguments contribute to the selection of semantic relations. We then propose three neural-network-based models, i.e., a convolutional network (CNN), 3-layer neural-network (3-LNN) and GRU, by applying transfer learning for using different semantic-role-labeled corpora. With transfer learning, using annotated corpora with different tag sets from the target semantic role labels is expected.

We conducted experiments on the Japanese SRL accuracy of the proposed models that involved using the BCCWJ Predicate-Thesaurus (BCCWJ-PT) corpus, which is annotated with dependencies for target verbs in BCCWJ (Maekawa et al., 2014). The reasons of using BCCWJ-PT are 1) most of the semantic role labels correspond to the well-known categories of Agent, Theme, Goal, and Recipient, which are also used in other resources (e.g., VerbNet and FrameNet); and 2) the semantic role labels and their verb senses (i.e., semantic frames) are defined with example sentences in the Predicate Thesaurus (PT), which are freely available on the Internet⁴.

The experimental results indicate that the last two morphemes of arguments sufficiently boosted the SRL performance of our neural-network-based models as well as a baseline, i.e., an SVM-based model. We also reveal that transfer learning improved the accuracy of recognizing labels annotated in BCCWJ-PT using different semantic tags annotated in GDA.

2 Characteristics of Japanese SRL

There are two main characteristics of Japanese SRL, i.e., Japanese grammatical features and language resources for SRL.

2.1 Japanese Grammatical Features for SRL

In Japanese syntax, the case markers located in the final part of each phrase, play important roles in determining the semantic relations (i.e., semantic roles) to the predicate. Not only the case markers but also certain functional multi-morphemes, attached to the final part of each phrase have an

effect on determining the semantic relations to the predicate. The example sentence in Figure 1 shows

akutenkou-ni-yotte furaito-ga kyanseru-sare-ta
(no-tame-ni/de)
 bad weather-due to flight-NOM cancel-PASSIVE-PAST
 The flight was canceled due to bad weather

Fig. 1: Example of functional multi-morphemes

that the different functional morphemes *ni-yotte*, *no-tame-ni*, and *de*⁵ can designate the same type of semantic relation, i.e., “Cause” to the verb *kyanseru* (*cancel*). Also, certain functional morphemes as well as case markers can provide different semantic relations depending on their contexts.

Unfortunately, there is no standard dictionary of functional morphemes⁶ nor morphological analyzers that can detect these functional morphemes⁷. Thus, we do not currently have language resources that can help determine the possible semantic relations that functional morphemes provide⁸.

With this background, we take the last two morphemes in each argument as a simple grammatical feature to take into account the functional morphemes and case markers of arguments in Japanese, as in a previous study (Ishihara and Takeuchi, 2016).

2.2 Language Resources of Japanese SRL

Several language resources have been constructed on internal semantic relations in predicates and their arguments for Japanese. The language resource EDR provides English and Japanese parsed corpora containing about 400,000 examples, which are annotated with semantic tags and original sense tags defined in the EDR dictionary. The semantic tags not only contain semantic role labels, such as Agent and Object, but also semantic relations

⁵Most of the functional suffixes are written in Hiragana.

⁶There was a study on the collection of Japanese suffixes (Matsuyoshi et al., 2007); however, there are still no registered functional morphemes such as *to-issyoni* (*with*).

⁷For example, the Japanese morphological analyzer MeCab (<http://taku910.github.io/mecab/>) with the IPA dictionary does not extract the functional suffix *no-tame-ni*.

⁸The functional morphemes are annotated in CoNLL2009 not as features but as semantic role tags (i.e., target of disambiguations). Thus, the semantic role tags in the Japanese corpus of CoNLL 2009 might be different from those for English SRL.

⁴<http://pth.cl.cs.okayama-u.ac.jp/>

between main and subordinate clauses, e.g., Co-occurrence and Sequence; however, these similar semantic role tags are not connected to predicate concepts, such as frames, as defined in PropBank or FrameNet. Thus, we do not use EDR as the target SRL corpus.

Unlike the EDR corpus, the JFN (Ohara et al., 2003) was constructed in the same structure of English FrameNet, however, JFN is a closed corpus, and several essential information such as total amount of the annotated sentences and number of lexicons are not revealed. Thus, it does not seem to be easy to use the JFN corpus. The GDA corpus has 37,000 sentences that are annotated with 100 semantic tags that contain not only semantic role tags but discourse-related, syntactic, and grammatical tags. This can be regarded as a sufficient amount; however, the semantic role tags are not related to frames for predicates nor are lexical frames provided.

Instead of the above semantic role-related corpora in Japanese, BCCWJ-PT contains 64 semantic role labels that are based on the analysis of semantic relations between predicates and arguments for about 5,000 sentences. The semantic role labels are also connected to frames defined in the PT, which is freely published on the Web. This framework of tags and frames is the same as PropBank and FrameNet. We therefore use BCCWJ-PT as the target corpus of Japanese SRL.

3 Task Definition

To focus on the impact of how features and learning models can recognize semantic roles in Japanese texts, we apply the simple SRL task, i.e., SRL systems identify the semantic role labels for a pair of an argument and head verb that are correctly extracted in the preprocessing step.

Figure 2 shows an example of a partial sentence annotated with semantic roles in BCCWJ-PT.

The BCCWJ-PT corpus has not only semantic role labels but also annotated tags of morphemes, the target predicate, and its arguments; thus, we convert the tagged texts into the target data, which are separated into target semantic role labels and their features. Figure 3 shows an example of the target data converted from the annotated data in Figure 2.

In the example features in Figure 3, basic forms as well as surface forms of the verbs are stored for normalization as the predicate features. The basic forms and/or the surface forms are selected as the predicate features depending on the models of neural networks.

4 Proposed Models and Approaches

Previous work has shown that neural-network-based approaches models are powerful for English SRL (He et al., 2017) as well as Japanese case-marker disambiguation and anaphora detection (Ouchi et al., 2015; Shibata et al., 2016; Ouchi et al., 2017; Matsubayashi and Inui, 2018); however, neural-network-based models have various tuning parameters such as network structure, hyper parameters, and several methods of avoiding over fitting. Therefore, it is not easy to determine the cause-and-effect relations between models and the final results. Thus, we used simple neural-network architectures to focus on clarifying what input features are effective, which learning models are powerful, and how different labeled corpora can be effective for the target label set.

As described above, several corpora annotated with different sets of semantic role labels for Japanese have been constructed, but the amount of labeled examples for each annotated corpus is limited. Thus, we applied transfer learning, which uses corpora annotated with different semantic labels from the target labels. Because of the flexibility of the neural-network structures, transfer learning can be easily implemented by changing the network structure.

In the following sections, we give details of the three proposed models, their input features, and the methods of transfer learning.

4.1 3-LNN

Our 3-LNN model is applied to determine semantic role labels by varying the following different input feature vectors to determine the effectiveness of the input feature vector.

- (1) Bag-of-Words (BOW) for morphemes in an argument and its head verb. Each argument is separated into morphemes using the Japanese morphological analyzer MeCab with UniDic dictionary.

Japanese sentence: [Theme^{guranturismo-4-o}] [Complement_ACC^{shinpin-de}] ka-ou-to-omotte-iru
 gran turismo-4-ACC brand new-with buy-want-think-ing
 English sentence: I would like to buy a brand new Gran Turismo 4.

Fig. 2: Example annotated sentence of semantic role labels in BCCWJ-PT

Semantic role label	Features		
	Argument	Predicates	
		Basic form	Surface form
Theme	<i>guranturismo-4-o</i> (gran turismo 4-ACC)	<i>kau</i> (buy)	<i>ka-ou</i> (buy-want)
Complement_ACC	<i>shinpin-de</i> (brand new-with)	<i>kau</i> (buy)	<i>ka-ou</i> (buy-want)

Fig. 3: Example of target data for Japanese SRL

- (2) Feature vector (1) with two skip-gram vectors: the content head morpheme and its verb in an argument. The skip-gram vectors are extracted from word embedding dataset from NINJAL Web Japanese Corpus (nwjc2vec) (Asahara, 2018) that are made from 1 billion corpus of Japanese using fasttext⁹. These vectors are expected to determine the tendency between semantic role tags and the combination of a content morpheme and verb.
- (3) Feature vector (2) with a BOW of the last two morphemes in an argument. This added feature is expected to capture functional suffixes in arguments, as described in Section 2.1.

For example, feature vector (1) for the first data in Figure 3 is a BOW vector of *granturismo*, *4*, *o*, and *kau*. In feature vector (2), the skip-gram vectors of the content morpheme-head, i.e., *4* and its verb *kau*, are added to feature vector (1). In feature vector (3), a BOW vector of the two morphemes *4* and *o* is added to feature vector (2).

Regarding the network structure and tuning methods, ReLU (Nair and Hinton, 2010) is applied to the non-linear function of the intermediate units, and softmax is used as the function of the units at the final layer.

To obtain better accuracy for test data, we applied a dropout method for the intermediate units with 50%; and used Adam as an optimization

⁹<https://github.com/facebookresearch/fastText.git>

method with the recommended setting¹⁰.

4.2 CNN

Instead of using BOW, we take certain sequential characteristics in verb-argument features with a simple CNN. Our CNN is a simple structure that consists of an input layer, convolution layer, pooling layer, and output layer (Figure 4). The input of the CNN is a sequence of feature morphemes in verb-argument order. The nwjc2vec was used to convert the input morphemes to the skip-gram vectors with 200 dimensions. We define the three types of filters that take into account three, four, and five successive morphemes. The number of filters for each type is 128.

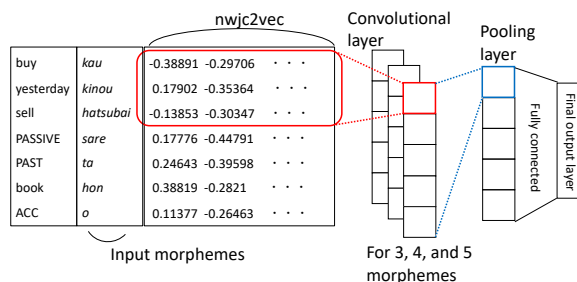


Fig. 4: Semantic role labeling with CNN model

After the convolution layer, a pooling layer with max pooling is applied, then a fully connected layer is applied to the connection to the final output layer. In addition to the simple network, we also use a

¹⁰The parameters of Adam were set to $\alpha = 0.001$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$.

variation model in which the input features defined in the 3-LNN model are added into the previous final layer, as mentioned in Section 6.

4.3 GRU

To focus on the sequential feature of arguments and verbs, we use our GRU model, which is a type of recurrent neural network.

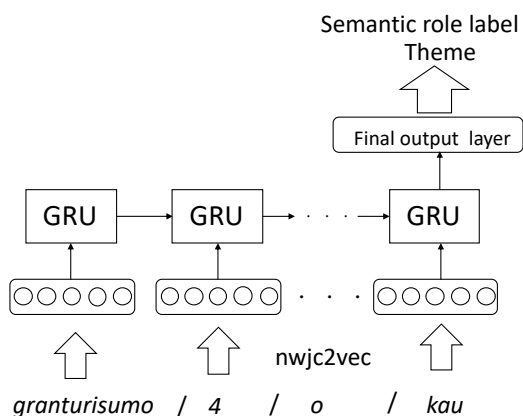


Fig. 5: Semantic role labeling with GRU model

Figure 5 shows the structure of how we apply our GRU model for SRL. Input feature morphemes are converted to dense vectors with *nwjc2vec*, the final GRU state is applied to a dense layer, then the final output is obtained.

To observe how the order and inflection of input morphemes can contribute to the accuracy of identifying semantic role labels, four types of input features are defined in Table 1. The slash in the examples denotes a delimiter of morphemes. The GRU model takes a verb with the final position in *v1*, but a verb or verbs at the first position in the other features. Feature vectors *v3* and *v4* have a surface form of the verb.

4.4 Transfer Learning

As described in Section 2.2, the amount of the target corpus, BCCWJ-PT, is limited; thus, we apply a method of using other annotated corpora whose semantic labels are even different from that of the target corpus, which is called transfer learning.

The basic procedure of transfer learning is as follows: 1) a neural network is trained using an annotated corpus of different semantic tags from the target corpus; 2) the units at the final layer of

the neural network are replaced with new units for the target semantic tags; and 3) all the weights in the neural network are trained using the target corpus¹¹. Transfer learning is applied to the three proposed neural-network-based models.

4.5 Baseline Model

We apply an SVM-based model that has a linear kernel to the SRL task as a baseline model. For multi-class categorization, we use the one-versus-rest method. The input feature is the BOW, which is the same as case (1) used in the 3-LNN model.

5 Experiments on Japanese SRL

Accuracy

5.1 Experimental Setup

The BCCWJ-PT corpus was the target corpus containing 5069 sentences, 64 semantic role labels for 548 verb types¹². We extracted 10,390 instances of the target data whose format is shown in Figure 3, i.e., combinations of an argument and its verb. The target data were divided into the three parts: training (65%), development (5%), and test (30%). The GDA corpus containing 100 semantic tags, which are different from those in BCCWJ-PT, was selected as the corpus for transfer learning. We extracted 82,892 instances whose format is the same as that in Figure 3. The instances of GDA were divided into development (85%), training (5%), and test (10%) data. Table 2 shows the details of the data sets used in the experiments.

The top five most frequent semantic role labels contained in the BCCWJ-PT data are listed in Table 3. The semantic role labels of Theme and Agent were most frequent, which indicates the same tendency shown in PropBank (Palmer et al., 2005), which is an English semantic-role-labeled corpus.

We applied the proposed models to the training data for training parameters and applied them to the test data for evaluating their SRL accuracy.

The SRL accuracy of the SVM model was evaluated with 5-fold cross validation of all target data. This indicates that the SVM model is advantageous

¹¹In preliminary experiments, we found that training only the weights to the final layer does not perform well compared with training all the weights.

¹²The version of BCCWJ-PT we used is 0.9.

TABLE 1. Variation of input feature vectors

	description	example
v1	place argument followed by basic form of verb	<i>guranturismo / 4 / o / kau</i>
v2	place basic form of verb followed by argument	<i>kau / guranturismo / 4 / o</i>
v3	place surface form of verb followed by argument	<i>ka-ou / guranturismo / 4 / o</i>
v4	place basic and surface form of verbs followed by argument	<i>ka-ou / kau / guranturismo / 4 / o</i>

TABLE 2. Size of each data set used in experiments

	BCCWJ-PT	GDA
training	6,753 (65%)	70,458 (85%)
development	520 (5%)	4,144 (5%)
test	3,117 (30%)	8,290 (10%)

TABLE 3. Top five most frequent labels in BCCWJ-PT data

Semantic role label	frequency
Theme	3,100
Agent	1,200
Manner	549
Modifier	443
Adverb	439

compared to our neural-network-based models, because the amount of training data is larger than that of our neural-network-based models.

The SRL accuracies of our models and the SVM model were evaluated based on the instances of the test data using the following accuracy formula:

$$accuracy = \frac{\text{\#instances correctly estimated labels}}{\text{\#total instances}} \quad (1)$$

In the SVM model, evaluation on test data in one hold was conducted with the above accuracy formula. The final accuracy of the SVM model was then an average evaluated based on all the test data in 5-fold cross validation.

The mini-batch size of all three proposed neural-network-based models was set to 100, and the number of training iterations, i.e., epochs, was determined using the accuracy of the development data. To avoid overfitting to the training data, we used 20 epochs on the BCCWJ-PT data because almost all the proposed models converged in 20 epochs.

6 Experimental Results

Table 4 lists that experimental results of the SRL accuracies of the proposed and baseline models using only the BCCWJ-PT data.

TABLE 4. Results of SRL by using BCCWJ-PT data.

Model	Feature	Accuracy
SVM	BOW	0.508
	BOW + skip	0.562
	BOW + skip + two	0.598
3-LNN	BOW	0.538
	BOW + skip	0.610
	BOW + skip + two	0.650
GRU	v1	0.599
	v2	0.633
	v3	0.619
	v4	0.631
CNN	conv	0.641
	conv + BOW + skip + two	0.665

All three neural-network-based models outperformed the SVM model regarding SLR accuracy. When we look at the effectiveness of the features in the SVM and 3-LNN models, the skip-gram vectors significantly improved the accuracies of the both models. Adding a BOW of the last two morphemes in each argument further improved their accuracies.

For the GRU results, **v2** showed the best performance among the other feature sets. This indicates that the verb should come first in the input sequence by comparing to **v1**; and the base form of the verbs must be more effective than the inflected form compared to the results of **v3** and **v4**. The SRL accuracy of the GRU model, however, was inferior to that of the 3-LNN model. This indicates that the GRU model currently does not seem to fully use contextual information.

The best SRL accuracy of all the models was that of our CNN model. Compared to the 3-LNN model,

the convolution and pooling structure contributes to improve 0.015 points. The CNN model with only using the base feature vector **conv** did not perform as well as the 3-LNN model with the **BOW + skip + two** feature vector. This indicates that the manually designed feature vectors, i.e., **skip + two** with **BOW**, work well to obtain the characteristics of the semantic role labels.

TABLE 5. Results of transfer learning: SRL accuracy of proposed models with BCCWJ-PT for GDA data in pre-training.

Training epochs	Accuracy with BCCWJ-PT			
	[with GDA data]	3-LNN	GRU	CNN
0		0.650	0.633	0.665
10		0.670	0.638	0.669
30		0.669	0.641	0.664
50		0.665	0.641	0.659
100		0.652	0.644	0.654
300		0.655	0.628	0.629
500		0.615	0.615	0.619

Table 5 shows the results of incorporating transfer learning, i.e., using the GDA data for training the initial values of the weights. All three proposed neural-network-based models improved their accuracies with transfer learning, but the increase in the accuracies differed depending on the model. The 3-LNN model had the most improvement with transfer learning and showed the best accuracy among all the models. The CNN model improved in accuracy, but was not as effective as the 3-LNN model. The accuracy of the GRU model also increased 0.011 points within the maximum score, but the best accuracy of the GRU model was lower than those of the other models.

According to the effects of the training epochs in the GDA data, too many training epochs for the GDA data will decrease the SLR accuracy with BCCWJ-PT for all three models. This indicates that neural-network-based models would have caused overfitting to the GDA data if the models were trained with too many iterations. Thus, we need to stop the training in GDA data with a small number of iterations. Table 5 shows that the best training epoch for the 3-LNN and CNN models is only ten iterations, which would be the best for obtaining the initial weights towards learning the final BCCWJ-PT data.

7 Discussions

The results of transfer learning in Table 5 indicate that transfer learning contributed to the improvement in the SRL accuracies of our neural-network-based models, but the best accuracy score of 0.67 is not so different from 0.665 with the CNN model without transfer learning even though the GDA data are about ten times larger than the BCCWJ-PT data. The role of transfer learning is to obtain better initial weights in neural-network-based models than randomized initial weights. In transfer learning, all the units in the final layer for GDA tags are discarded; however, some of the tags are almost the same as the semantic role tags defined in BCCWJ-PT, such as Agent, Theme, and Goal. Therefore, we must consider how we can use the similar semantic tags in transfer learning.

As described in Section 2.1, the last two morphemes in each argument are defined to capture functional suffixes that have an effect on determining its semantic role of the argument. The experimental results listed in Table 4 reveal that the multi-word functional suffix i.e., **two** feature vector, improves the accuracy of the GRU model as well as those of the SVM and 3-LNN models. The three feature vectors **v2**, **v3**, **v4** outperformed **v1** in terms of SRL accuracy. Since **v1** is only the case in which a verb comes at the end, the other case markers come last. In the GRU model, the final layer of the GRU loaded at the final positions of a time sequence determines the label. Therefore, the last two morphemes located at the final positions of a time sequence are naturally taken into account in **v2**, **v3**, and **v4**. The accuracies of all the models show that the last morphemes have a positive effect on determining the semantic role labels in Japanese.

8 Related Work

Several SRL studies have been conducted mainly in English because of existing high-quality language resources such as FrameNet and PropBank as well as shared tasks of semantic role labels such as in CoNLL-2005 (Carreras and Màrquez, 2005), 2009 (Hajič et al., 2009) and 2012 (Pradhan et al., 2012).

In the early stage of SRL investigation, statistical modeling and effective features for SRL have been studied. Gildea and Jurafsky (2002) revealed that several syntactic features, such as parse tree path,

phrase type, and voice, can improve the accuracy of a statistical learning model. More detailed features were studied by Surdeanu et al. (2003) and Xue and Palmer (2004). Toutanova et al. (2008) showed effective combinations of statistical joint models with rich features.

Syntactic features are powerful; however, parse errors will decrease the accuracies of SRL systems. Thus Zhou and Xu (2015) proposed an end-to-end SRL model using bi-directional long short-term memory (LSTM) without any syntactic features. Roth and Lapata (2016) used dependency information on LSTM. The dependency path is convenient, but He et al. (2017) revealed that higher accuracies on neural-network-based SRL models could be obtained if correct parsed information is available.

Most studies on Japanese SRL (Taira et al., 2008; Imamura et al., 2009; Sasano and Kurohashi, 2011; Hayashibe et al., 2011; Ouchi et al., 2015; Shibata et al., 2016; Ouchi et al., 2017; Matsubayashi and Inui, 2018) have been focused on recognizing three types of case-marker-based semantic roles with anaphora resolutions.

Taira et al. (2008) have shown that the detailed noun categories of nominals in arguments improve the accuracy of statistical models for recognizing the three case-markers. Imamura et al. (2009) proposed effective grammatical features such as dependency path, phrase positions, and several detailed characteristics. Sasano and Kurohashi 2011 and Hnagyo et al. 2013 proposed models to use large-scale case frames to provide selectional preference between a head noun in an argument and its predicate.

Ouchi et al. (2015), Shibata et al. (2016), Ouchi et al. (2017) and Matsubayashi and Inui (2018) proposed neural-network-based models. These studies are focused on anaphora resolution, i.e., detecting arguments for a predicate without dependency relations and recognizing their semantic roles. Thus, these studies discussed how to incorporate the effectiveness of multiple predicates.

For SRL in Japanese, Ishihara and Takeuchi (2015) revealed that the last morphemes in an argument are effective on a linear-chain CRFs for determining 64 semantic roles for BCCWJ-PT.

Thus, the effective grammatical features as well

as approaches on neural-network-based models for Japanese SRL are required to be studied.

9 Conclusion

We propose three neural-network-based models and described the effective features and methods for Japanese SRL. We revealed that the last two morphemes in an argument, the dense morpheme vector concatenated with a head noun morpheme and its predicate, and bag-of-morphemes in an argument, are effective for our 3-LNN and CNN models. We conducted experiments on Japanese SRL with BCCWJ-PT containing 64 semantic roles, which was different from most previous studies, which focused on 3 semantic roles.

We applied transfer learning using GDA, which has different semantic role tags from BCCWJ-PT. After pre-training the weights using the GDA data, the neural network models were trained on the BCCWJ-PT data. The experimental results indicate that transfer learning improved the accuracies of all three proposed neural network models compared to the cases without transfer learning.

We are planning more detailed analyses of the combinations of features and neural network models on BCCWJ-PT.

References

- Masayuki Asahara. 2018. NWJC2Vec: Word embedding dataset from ‘NINJAL Web Japanese Corpus’. *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication*, 24(1):7–22.
- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics*, pages 86–90.
- Xavier Carreras and Lluís Màrquez. 2005. Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling. In *Proceedings of the CoNLL-2005 Shared Task*.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28(3):1–45.
- Jan Hajič, Massimiliano Ciaramita, Richard Johanson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu,

- Nianwen Xue, and Yi Zhang. 2009. The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL): Shared Task*, pages 1–18.
- Yuta Hayashibe, Mamoru Komachi, and Yuji Matsumoto. 2011. Japanese Predicate Argument Structure Analysis Exploiting Argument Position and Type. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 201–209.
- Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. Deep Semantic Role Labeling: What Works and What’s Next. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 473–483.
- Ryu Iida, Mamoru Komachi, Kentaro Inui, and Yuji Matsumoto. 2007. Annotating a Japanese Text Corpus with a Predicate-Argument and Coreference Relations. In *Proceedings of the 1st Linguistic Annotation Workshop*, pages 132–139.
- Kenji Imamura, Kuniko Saito, and Tomoko Izumi. 2009. Discriminative Approach to Predicate-Argument Structure Analysis with Zero-Anaphora Resolution. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 85–88.
- Yasuhiro Ishihara and Koichi Takeuchi. 2016. Construction of Japanese Semantic Role Labeling System Using Hierarchical Tag Context Tree Extracted from Tail Expressions of Dependency Elements. *Journal of Information Processing*, 57:1611–1626. (in Japanese).
- Daisuke Kawahara, Sadao Kurohashi, and Kōiti Hasida. 2002. Construction of a Japanese Relevance-tagged Corpus. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*, pages 2008–2013.
- Paul Kingsbury, Martha Palmer, and Mitch Marcus. 2002. Adding Semantic Annotation to the Penn TreeBank. In *Proceedings of the Human Language Technology Conference*.
- Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. 2014. Balanced Corpus of Contemporary Written Japanese. *Language Resources and Evaluation*, 48:345–371.
- Yuichiro Matsubayashi and Kentaro Inui. 2018. End-to-End Japanese Predicate-Argument Analysis Taking into Account Relations Among Multiple Predicates. In *Proceedings of the 24th Annual Meeting of the Association for Natural Language Processing*, pages 101–104. (in Japanese).
- Suguru Matsuyoshi, Satoshi Satoh, and Takehito Utsuro. 2007. A Dictionary of Japanese Functional Expressions with Hierarchical Organization. *Journal of Natural Language Processing*, 14(5):123–146.
- Vinod Nair and Geoffrey E. Hinton. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on Machine Learning, ICML’10*, pages 807–814, USA. Omnipress.
- Kyoko Ohara, Seiko Fujii, Hiroki Saito, Shun Ishizaki, Toshio Ohori, and Ryoko Suzuki. 2003. The Japanese FrameNet Project: A Preliminary Report. In *Proceedings of Pacific Association for Computational Linguistics 2003*, pages 249–254.
- Kyoko Ohara, Junya Kato, and Hirokazu Saito. 2011. Annotation of Japanese FrameNet to BCCWJ. In *Proceedings of the Workshop of Japanese Corpus in Grant-in-Aid for Scientific Research on Priority Areas*, pages 513–518. (in Japanese).
- Hiroki Ouchi, Hiroyuki Shindo, Kevin Duh, and Yuji Matsumoto. 2015. Joint Case Argument Identification for Japanese Predicate Argument Structure Analysis. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 961–970.
- Hiroki Ouchi, Hiroyuki Shindo, and Yuji Matsumoto. 2017. Neural Modeling of Multi-Predicate Interactions for Japanese Predicate Argument Structure Analysis. In *Proceedings of the 55th annual meeting of the Association for Computational Linguistics*, pages 1591–1600.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–105.
- Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes. In *Proceedings of the Joint Conference on EMNLP and CoNLL: Shared Task*, pages 1–40.
- Michael Roth and Mirella Lapata. 2016. Neural Semantic Role Labeling with Dependency Path Embeddings. In *Proceedings of 54th Annual Meeting of the Association for Computational Linguistics*, pages 1192–1202.
- Mamoru Saito. 1989. Scrambling as Semantically Vacuous A’-movement. In Mark R. Baltin and Anthony S. Kroch, editors, *Alternative Conceptions of Phrase Structure*, pages 182–200. University of Chicago Press.
- Ryohei Sasano and Sadao Kurohashi. 2011. A Discriminative Approach to Japanese zero Anaphora Res-

- olution with Large-Scale Lexicalized Case Frames. In *In Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pages 758–766.
- Tomohide Shibata, Daisuke Kawahara, and Sadao Kurohashi. 2016. Neural Network-Based Model for Japanese Predicate Argument Structure Analysis. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 1235–1244.
- Mihai Surdeanu, Sanda Harabagiu, John Williams, and Paul Aarseth. 2003. Using Predicate-Argument Structures for Information Extraction. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 8–15.
- Hirotoishi Taira, Sanae Fujita, and Masaaki Nagata. 2008. A japanese predicate argument structure analysis using decision lists. In *Proceedings of EMNLP*, pages 523–532.
- Koichi Takeuchi, Masayuki Ueno, and Nao Takeuchi. 2015. Annotating Semantic Role Information to Japanese Balanced Corpus. In *Proceedings of MAPLEX 2015*.
- Kristina Toutanova, Aria Haghighi, and Christopher D Manning. 2008. A Global Joint Model for Semantic Role Labeling. *Computational Linguistics*, 34(2):161–191.
- Nianwen Xue and Martha Palmer. 2004. Calibrating Features for Semantic Role Labeling. In *Proceedings of Empirical Methods in Natural Language*, pages 88–94.
- Jie Zhou and Wei Xu. 2015. End-to-end Learning of Semantic Role Labeling Using Recurrent Neural Networks. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*
Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pages 1127–1137.