

Improving Shared Argument Identification in Japanese Event Relation Knowledge Acquisition

Yin Jou Huang

Graduate School of Informatics
Kyoto University

huang@nlp.ist.i.kyoto-u.ac.jp

Sadao Kurohashi

Graduate School of Informatics
Kyoto University

kuro@i.kyoto-u.ac.jp

Abstract

Event relation knowledge represents the knowledge of causal and temporal relations between events. Shared arguments of event relation knowledge encode patterns of role shifting in successive events. A two-stage framework was proposed for the task of Japanese event relation knowledge acquisition, in which related event pairs are first extracted, and shared arguments are then identified to form the complete event relation knowledge. This paper focuses on the second stage of this framework, and proposes a method to improve the shared argument identification of related event pairs. We constructed a gold dataset for shared argument learning. By evaluating our system on this gold dataset, we found that our proposed model outperformed the baseline models by a large margin.

1 Introduction

Natural language understanding requires not only linguistic knowledge but also common knowledge about the real world. Event relation knowledge is a type of common knowledge of critical importance, representing the knowledge of the relation between events as well as the typical patterns of role shifting between events. Event relation knowledge is useful for natural language understanding tasks as well as natural language generation tasks which require modeling of the possible event sequences.

In this paper, we define an event to be a predicate argument structure (PAS), which consists of a

¹In this paper we adopt the Japanese case marker, *ga*, *wo*, *ni*, and *de*, which roughly corresponds to nominative, accusative, dative, and instrumental/locative cases.

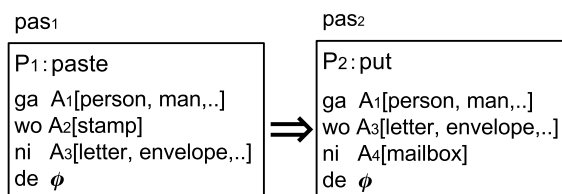


Figure 1: Event relation knowledge with shared arguments.¹

predicate and its relevant arguments. In addition, we define one unit of event relation knowledge to be a pair of successive events with one or more shared arguments. Figure 1 represents an example of event relation knowledge, which consists of two events, *pas₁* and *pas₂*.

The shared arguments correspond to the common participants of the two events, such as *A₁* and *A₃* in the above example. These shared arguments play an important role in the application of event relation knowledge since they encode the correspondence relations between case slots within a piece of event relation knowledge.

In this paper, we aim to improve the shared argument identification in Japanese event relation knowledge. Event relation knowledge acquisition in Japanese is a much more challenging task than its counterpart of English, due to several linguistic properties of Japanese. For example:

- (1) a. John attached a stamp to the letter, and he dropped it into the mailbox.
- b. John attached a stamp to the letter, and (ϕ_{he}) dropped (ϕ_{letter}) into the mailbox.

In the above example, (1-b) is the Japanese correspondence of (1-a), directly translated into English. We can observe that Japanese has an abundance of omitted arguments. In addition, Japanese lacks linguistic clues regarding the accordance in gender, number, etc., such as ‘he’ and ‘it’ in (1-a).

These linguistic properties hinder the performance of Japanese coreference resolution sys-

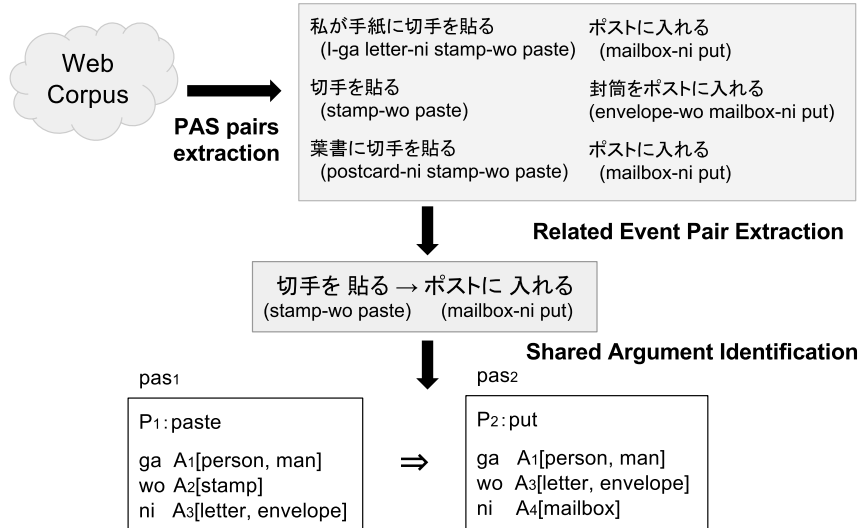


Figure 2: Two-stage approach for Japanese event relation knowledge acquisition.

tems, and make it unsuitable to apply coreference-based methods of English event relation knowledge acquisition (Chambers and Jurafsky, 2008) directly to Japanese.

On the other hand, event relation knowledge can benefit the task of the coreference resolution. The shared arguments within an event relation knowledge provide direct clues that the case slots sharing an argument should hold co-referring arguments. These clues are particularly critical in cases in which selectional preference is not helpful, such as coreference resolution problems presented in Winograd Schema Challenge (Levesque et al., 2012; Rahman and Ng, 2012). Consider the following example:

- (2) a. グーグルが モトローラ を買収した。
彼らが破綻したからだ。
 (Google-ga acquired Motorola-wo, because they-ga went bankrupt.)
- b. A_1 -ga go bankrupt \rightarrow A_2 -ga A_1 -wo acquire

In the example of (2-a), both precedents of ‘they’, ‘Google’ and ‘Motorola’, are of the same category. While selectional preference is not helpful in this case, the event relation knowledge in (2-b) can help us resolve (2-a) correctly.

In this work, we adopted the two-stage framework for Japanese event relation knowledge acquisition (Shibata and Kurohashi, 2011). In the first stage of related event pair extraction, we adopted the method proposed by Shibata and Kurohashi (2011); and in the second stage of shared argument identification, we extended the model of Kohama et al. (2015) to incorporate all types of

shared arguments in our gold dataset. We designed a richer feature representation for shared argument learning, which considers the interaction between shared arguments and the mechanism of argument omission in depth.

In addition, we manually constructed a gold dataset for shared argument learning. With the help of linguistic experts, we established an annotation scheme for shared argument. We classified the shared arguments into three types: standard shared argument, quasi shared argument, and multiple shared argument. We evaluated our method of shared argument identification on the gold dataset. By comparing our proposed methods with several baseline models, we observed a significant improvement for shared argument identification.

2 Related Work

As a resource-rich language, coreference resolution of English has achieved a satisfying performance. Thus, several works which utilize coreference information were proposed for English event relation knowledge acquisition.

Chambers and Jurafsky (2008) introduced the concept of narrative event chains as a representation of structured event relation knowledge. Their method utilizes the coreference chains within the input text to collect events involving the same entity, which they called the *protagonist*. Among the set of events involving the same entity, event sequences that are observed a significant number of times are extracted as typical event sequences.

pas_1	pas_2	Support sentences
切手を貼る (stamp-wo paste)	ポストに入れる (mailbox-ni put)	親宛に書いた葉書きに切手を貼ってポストに入れた。 (I pasted a stamp on the postcard to my parents, and put it into the mailbox.) 手紙を書いて、封をして、切手を貼って、ポストに入れる。 (I write a letter, seal it, paste a stamp, and put it into the mailbox.)
薬を飲む (medicine-wo take)	症状が軽くなる (symptom-ga alleviate)	薬を飲み続けていると、アレルギーの症状は大分軽くなってきている。 (Taking the medicine alleviates the allergy symptom significantly.) 抗ヒスタミン系の薬を処方され、飲めば症状は軽くなります。 (I was prescribed antihistamine, the symptom alleviated after taking it.)

Table 1: Related Event Pairs.

Pichotta and Mooney (2014) used a richer representation of event than in the work of Chambers et al. and achieved an improvement in predicting performance. Instead of representing an event as a (predicate, dependency) pair, they considered an event as a structure of a predicate and arguments with subject, object, direct object relations with the predicate. With this multi-argument event representation, their model performs better in the cases of ambiguous verbs, and is more capable of capturing complex interactions between multiple entities.

There are several works proposed for Japanese event relation knowledge acquisition utilizing the co-occurrences of events. Abe et al. (2008) proposed a pattern-based method which utilized a pre-defined set of lexico-syntactic co-occurrence patterns to perform bootstrapping for event relation learning. Their work focused on the acquisition of related event pairs, but not the relations between the arguments of the related events.

Shibata and Kurohashi (2011) proposed a two-stage approach for Japanese event relation knowledge acquisition (Figure 2). In the first stage, related event pairs are extracted from large-scale corpora by association rule mining. In the second stage, shared arguments of the event pairs are identified heuristically based on case slot similarity scores.

Kohama et al. (2015) improved the work of Shibata and Kurohashi (2011) by utilizing crowd-sourced data for shared argument learning. They proposed a joint model that simultaneously predicts the shared argument configuration and disambiguates the meaning of the predicates. However, their work failed to identify the shared arguments accurately for two reasons. First, the crowd-sourced data they used is very noisy and lacks a well-defined standard of labeling. Second, the features used in their model are not sufficient for capturing the characteristics of shared arguments.

3 Shared Argument Identification

In this section, we introduce our method of shared argument identification. In Section 3.1, we first introduce the acquisition of related event pairs, which are the inputs to our shared argument identification model. We introduce the gold dataset used for model learning in Section 3.2. In Section 3.3, we describe the selection of case frames. These case frames will be used to model different meanings of predicates in our model. The remaining of the section will be dedicated to the description of our proposed methods of shared argument identification.

3.1 Related Event Pairs

Our work is based on the two-stage framework of event relation knowledge proposed by Shibata and Kurohashi (2011). We adopt the first stage of related event pair extraction proposed in their work to obtain the related event pairs, which will be the input to our shared argument identification model.

Here, we briefly describe the first stage of related event pair extraction. Starting from the web corpus, we first extract the PAS pairs with syntactic dependency, and use the Apriori algorithm to pick out the related event pairs efficiently (Figure 2). In order to improve the quality of the extracted event pairs, we apply an additional filtering step based on the clause relations between event pairs as suggested in Kohama et al. (2015).

Table 1 shows several examples of related event pairs extracted in this process. Each event pair R consists of two PASs, pas_1 and pas_2 , and the sentences containing both pas_1 and pas_2 are regarded as the support sentences of R . These support sentences contain many valuable clues for the task of shared argument identification. Thus, the event pair R along with its support sentences will serve as the input to our shared argument identification model.

Type	Event Pair	Shared Argument
Standard	切手を手紙に貼る 手紙をポストに入れる (stamp-wo letter-ni paste) (letter-wo mailbox-ni put)	n-w
Quasi	牛を飼う 牛乳でチーズを作る (cow-wo raise) (milk-de cheese-wo make)	w-d'
Multiple	観光客が町を/に訪れる 町が賑わう (tourist-ga town-wo/ni visit) (town-ga be crowded)	w/n-g

Table 2: Types of Shared Arguments.

Type	Shared Argument	Standard Shared Argument Set
Standard	$n-w$	$\{n-w\}$
Quasi	$w-d'$	$\{w-d, \phi\}$
Multiple	$n-n/w$	$\{n-n, n-w\}$

Table 3: Transforming different types of shared arguments to their standard shared argument sets.

3.2 Gold Dataset

We manually constructed a gold dataset for learning shared argument identification model. In this work, we train and evaluate our proposed model on this gold dataset.

This dataset contains 809 related event pairs, with each of the event pair annotated with its shared argument configuration. Three annotators with linguistic background participated in the construction of this dataset.

Type of Shared Arguments

The gold dataset contains the following types of shared arguments (Table 2):

1. Standard Shared Argument:

The arguments shared between one case slot of the first event and another case slot of the second event. This type of shared argument represents the fact that arguments of the two cases should correspond to an identical real world entity.

In this work, we only consider the four main cases of ga (が), wo (を), ni (に), and de (で). From now on, we use the shorthand notation of g , w , n , and d to represent these four main cases. The first example in Table 2 has a standard shared argument between the first ni-case and the second wo-case, which both correspond to the entity ‘letter’. we use the notation $n-w$ to represent it.

2. Quasi Shared Argument:

Quasi shared argument is a pair of arguments which are closely related to each other in the context of the given event relation knowledge. As can be seen from the example in Table 2, the arguments of the first wo-case

and the second de-case are ‘cow’ and ‘milk’, respectively. These two arguments are considered to be closely related since the milk in the context corresponds to the specific milk which is produced by the cow in the same context.

We attached an apostrophe (') to denote a quasi shared argument.

3. Multiple Shared Argument:

Multiple shared argument occurs when more than two case slots share the same argument. As can be seen from the example in Table 2, the argument ‘town’ is shared between three cases: wo-case or ni-case of the first event, and the ga-case of the second event.

We use the symbol ‘/’ to separate different case slots of the same predicate which share arguments.

Preprocessing of Gold dataset

In this work, we only focus on the identification of standard shared arguments. For utilizing the gold dataset with other shared argument types, we perform a pre-processing to the gold annotation before model training. We transform each shared argument configuration into its corresponding standard configuration set.

First, we define the corresponding standard shared argument set for each shared argument in the following manner (Table 3):

1. For each standard shared argument, we transform it into the standard shared argument set containing only itself.
2. For each quasi shared argument, we transform it into the standard shared argument set containing a null shared argument (ϕ) and its

Shared Argument Configuration	Standard Configuration Set
$[g-g]$	$\{[g-g]\}$
$[g-g\ w-d']$	$\{[g-g, w-d], [g-g]\}$
$[g-g\ n-n/w]$	$\{[g-g, n-n], [g-g, n-w]\}$
$[g-g\ w-d'\ n-n/w]$	$\{[g-g, n-n], [g-g, n-w], [g-g, n-n, w-d], [g-g, n-w, w-d]\}$

Table 4: Transforming shared argument configuration to corresponding standard configuration set.

standard counterpart in which all the apostrophe (') mark is removed. See the second example in Table 3.

- For each multiple shared argument, we transform it into the standard shared argument set containing all the shared arguments that could be entailed from it. See the third example in Table 3.

For a given shared argument configuration, we first transform each of its containing shared argument into its corresponding standard shared argument set in the above manner. By taking the product of these standard shared argument sets, we obtain the corresponding standard configuration set of the shared argument configuration. See Table 4 for examples.

3.3 Case Frame Selection

Selectional preferences provide important clues for the task of share argument identification. Case frames are good sources of selectional preference information, and it handles the issue of predicate ambiguity by clustering the usage of each predicate by their meanings. In turn, the meaning of a case frame is represented by the argument distribution in each case slot of its corresponding case frame.

In this work, we consider wide-coverage case frames constructed automatically from a huge raw corpus as the source of selectional preference information (Hayashibe et al., 2015). For each event pair $R(pas_1\ pas_2)$, we select 10 relevant case frames for both pas_1 and pas_2 by utilizing the supporting sentences S of R . Here, we describe the method for selecting relevant case frames for each event pair, which are used in our proposed models.

Given a case frame cf , we denote the bag-of-words (BoW) representation of arguments within each case slot of cf as follows:

$$V^g, V^w, V^n, V^d$$

We denote the BoW representation of arguments appearing in the corresponding case slots of the

support sentences S as follows:

$$U^g, U^w, U^n, U^d$$

We define the relevance score of cf with respect to R as follows:

$$rel(cf, R) = \sum_{x=\{g,w,n,d\}} \cos(U^x, V^x) \quad (1)$$

which is the sum of cosine similarity scores between the BoW representation of case slots in the four main cases.

Finally, we rank all the case frames in descending order with respect to relevance score and take the top 10 of them as relevant case frames. Table 5 represents the first five relevant case frames of the predicate 訪れる (visit) of the following event pair:

観光客が訪れる 賑わう
(tourist-ga visit be crowded)

3.4 Joint Prediction of Shared Argument and Case Frame

As mentioned in Section 3.3, case frames provide important information of selectional preferences. However, the gold data does not provide the appropriate case frame of each predicate. To tackle this problem, we propose a model of shared argument identification that simultaneously predicts the appropriate case frame for each predicate.

Model

We adopt a maximum entropy (MaxEnt) classifier model.

Given a related event pair $R(pas_1 \rightarrow pas_2)$ and its supporting sentences S , the conditional probability of a shared argument configuration \mathbb{A} and case frame pair cf_1, cf_2 is modeled as:

$$P(\mathbb{A}, cf_1, cf_2 | R, S; \mathbf{w}) = \frac{\exp\{\mathbf{w} \cdot \phi(\mathbb{A}, cf_1, cf_2, R, S)\}}{Z} \quad (2)$$

In the above equation, $\phi(\mathbb{A}, cf_1, cf_2, R, S)$ is the feature representation of the shared argument configuration, \mathbf{w} is the model parameter, and Z is the normalization constant. In Table 6 we summarized the features used, under the example of shared argument $n-w$.

Rank	Case Frame	Relevance Score
1	[観光客, 人] が [地, 日本] を [実際] に 訪れる [tourist, person]-ga [place, Japan]-wo [practically]-ni visit	0.966
2	[数人, 人] が [事務所, 京都] を [激励, 視察] に 訪れる [people, person]-ga [office, Kyoto]-wo [encourage, inspection]-ni visit	0.807
3	[観光客, 大統領] が [中国, 台湾] を [視察, 見学] に 訪れる [tourist, president]-ga [China, Taiwan]-wo [inspection, field trip]-ni visit	0.760
4	[客, 観光客] が [店, ショップ] を [目当, 実際] に 訪れる [guest, tourist]-ga [store, shop]-wo [goal, practically]-ni visit	0.748
5	[人, 観光客] が [博物館, 美術館] を [見学] に 訪れる [person, tourist]-ga [museum, art museum]-wo [field trip]-ni visit	0.742

Table 5: Relevant case frames of 訪れる (visit).

Feature	Description
Configuration	Binary feature indicating the existence of the shared argument $n-w$.
Post-predicate	Binary feature indicating the existence of argument in w case of pas_2 .
Core	Binary features indicating if n case of cf_1 and w case of cf_2 are core cases. If a case slot takes argument in more than 10% of the time in the selected case frame, we define it as a core case.
Case slot similarity	The cosine similarity between the vocabulary distribution of n case of cf_1 and w case of cf_2 .
Normalized case slot similarity	Case slot similarity of $n-w$ normalized over the similarities of all case slots of cf_1 . Same for cf_2 .
Conflict	The ratio of support sentences in S that holds different arguments in the first n case and the second w case.
Context	We collect words that appear in S but not within the event pair as context words. We calculate the relative probability of each context word to appear in the first n case compared to other main cases, and similar for the second w case. A tf-idf weighted sum of this probability is added as feature.

Table 6: Features for shared argument $n-w$.

Prediction

During the prediction phase, the shared argument configuration $\hat{\mathbb{A}}$ and case frame pair \hat{cf}_1, \hat{cf}_2 that gives the highest probability is chosen:

$$(\hat{\mathbb{A}}, \hat{cf}_1, \hat{cf}_2) = \operatorname{argmax}_{\mathbb{A}, cf_1, cf_2} P(\mathbb{A}, cf_1, cf_2 | R, S; \mathbf{w}) \quad (3)$$

For each related event pair R , we choose 10 relevant case frames for each predicate of concern as candidate of cf_1 and cf_2 , as described in Section 3.3.

Model Training

In the training phase, the most probable case frame pair (\hat{cf}_1, \hat{cf}_2) and the model parameter \mathbf{w} are updated alternatively. Also, the most probable gold configuration \hat{g} among the standard configuration set is also updated along with the case frame pair.

The training algorithm is summarized below:

1. Initialize model parameter \mathbf{w} randomly.
2. Use the current parameter \mathbf{w} to update the most probable gold configuration and the most probable case frame pair $(\hat{g}, \hat{cf}_1, \hat{cf}_2)$:

$$\hat{g}, \hat{cf}_1, \hat{cf}_2 = \operatorname{argmax}_{g, cf_1, cf_2} P(g, cf_1, cf_2 | R, S; \mathbf{w}) \quad (4)$$

3. Use $(\hat{g}, \hat{cf}_1, \hat{cf}_2)$ to update model parameter \mathbf{w} . The following is the objective function, in which the superscripts of g , cf_1 , and cf_2 denote the id of the event pairs, and N is the total number of training objects:

$$L = \sum_{n=1}^N \log P(g^{(n)}, cf_1^{(n)}, cf_2^{(n)} | R, S; \mathbf{w}) - \alpha \|\mathbf{w}\|^2 \quad (5)$$

$$\hat{\mathbf{w}} = \operatorname{argmax}_{\mathbf{w}} L \quad (6)$$

(Hyper-parameter α is set to 1.0.)

4. Back to 2 until convergence. The convergence condition is that the most probable $(\hat{g}, \hat{cf}_1, \hat{cf}_2)$ for all event pairs are the same as the previous iteration. If the convergence condition is not satisfied after 15 iterations, we terminate the training process.

3.5 Shared Argument Learning with Combined Case Frame

Here, we introduce another model for learning shared arguments which uses the combined case frames.

The joint reference model (Section 3.4) picks exactly one case frame for each predicate. On the other hand, the combined case frame model combines the relevant case frames by taking the weighed sum of them by the relevance scores with respect to the event pair. This method does not decide the most appropriate case frame of each predicate. Instead, all of the relevant case frames are considered, and case frames with higher relevance scores have larger influence on the feature representation.

Combined Case Frame

A combined case frame is obtained by combining the relevant case frames according to their relevance scores. The calculation of the relevance scores of each case frame is described in Section 3.3.

Given a set of relevant case frames CF , we defined the combined case frame \widetilde{cf} as follows:

$$\widetilde{cf} : \widetilde{V}^g, \widetilde{V}^w, \widetilde{V}^n, \widetilde{V}^d \quad (7)$$

$$\widetilde{V}^x = \sum_{cf \in CF} rel(cf, R) \times V_{cf}^x, \forall x \in \{g, w, n, d\} \quad (8)$$

in which V_{cf}^x is the vocabulary distribution vector of cf .

Model

Similar to the joint prediction model presented in section 3.4, we adopt a MaxEnt classifier model. Given an event pair $R(pas_1 \quad pas_2)$ and its supporting sentences S , we model the conditional probability of shared argument configuration \mathbb{A} as:

$$P(\mathbb{A}|R, S; \mathbf{w}) = \frac{\exp\{\mathbf{w} \cdot \phi(\mathbb{A}, \widetilde{cf}_1, \widetilde{cf}_2, R, S)\}}{Z} \quad (9)$$

In the above equation, ϕ is the feature representation as summarized in Table 6, \mathbf{w} is the model parameter, and Z is the normalization constant.

The training algorithm is similar to the one described in Section 3.4. In the training phase, the most probable gold configuration \hat{g} and the model parameter \mathbf{w} are updated alternatively until convergence.

4 Experiments

4.1 Settings

The case frames used in the experiments are built from a web corpus of four billion sentences, with the method proposed by Hayashibe et al. (2015).

We use Classias (Okazaki, 2009) as the implementation of maximum entropy classifier and L-BFGS (Nocedal, 1980) as the optimization algorithm for learning. We train and evaluate our proposed models by a 5-fold cross-validation test on the gold shared argument dataset.

4.2 Evaluation and Result

We apply three evaluation metrics: precision, recall, and F-score (F_1) for the evaluation of our shared argument identification models.

Model	Precision	Recall	F_1
Baseline[g-g]	0.731	0.717	0.724
Baseline[Kohama+15]	0.729	0.733	0.731
Joint	0.747	0.786	0.766
Combined	0.753	0.748	0.750

Table 7: Evaluation result.

We compared our proposed models with two baseline models. The first baseline model, denoted as Baseline[g-g] in Table 7, is the majority classifier which gives the output of g-g regardless of the event pair given. The second baseline model, denoted as Baseline[Kohama+15], is the model proposed by Kohama et al. (2015).

The experiment results are summarized in Table 7. In addition, several event relation knowledge acquired are shown in Table 8.

4.3 Discussion

Comparison with Baseline Models

As can be observed from Table 7, both of our proposed models outperformed the baseline models by a large margin.

Compared to the model proposed by Kohama et al. (2015), we use a richer feature representation for shared argument configuration. In their work, a shared argument is represented by the vocabulary distribution similarity between two case slots, such as the similarity between case frames, or the similarity between arguments in the supporting sentences. However, by considering only the distributional similarities between two case slots, their method overlooked two important intrinsic properties of the shared argument identification task:

Event Pair	Gold Annotation	System Output	Error Type
熟成させる 出荷される (ripen) (ship)	w-g	w-g	-
ジュースが安くなる 買う (juice-ga become cheaper) (buy)	g-w	g-w	-
肌に与える 若返らせる (skin-ni give) (rejuvenate)	w/g-g n-w	g-g n-w	-
切手を貼る ポストに入れる (stamp-wo paste) (mailbox-ni put)	g-g n-w	g-g	1
迫害される 殺される (suffer persecution) (be killed)	g-g n-n	n-g	2
明るくなる 太陽が顔を出す (become brighter) (sun-ga face-wo appear)	ϕ	g-w	3

Table 8: Evaluation results of the proposed and baseline models for shared argument identification.

Case	Arguments
ga	私, 誰, 人, ママ, 夫, 自分, 母, .. (I, who, people, mom, self, mother, ..)
wo	茶, 私, 子供, 花, 模様, ..., 手紙, ..., 封筒, .. (tea, I, child, flower, pattern, ..., letter, ..., envelope, ..)
ni	中, 風呂, 部屋, 手, 家, ポスト, .. (interior, bathroom, room, hand, house, mailbox)
de	<数量>+人, <時間>, 急須, 白, 湯, 鉛筆, .. (<number>+people, <time>, teapot, white, hot water, pencil)

Table 9: Example of bad case frame

1. Interaction of shared arguments:

Different pieces of shared arguments are not independent, and shared arguments that share a case slot have repulsive effects on each other. For example, if a shared argument configuration already includes $g-g$, then it would be unlikely that $g-w$ also exists in the same configuration. We add the normalized case slot similarity feature which considers not only the case slot similarity of a pair of case slots, but also the relative similarity of them, to account for this property.

2. The mechanism of argument omission in related event pairs:

High vocabulary distribution similarity indicates the existence of shared arguments, but not vice versa. Consider the following example:

ジュースが安くなる ジュースを 買う
(juice-ga become cheaper juice-wo buy)

Although there exists a shared argument of $g-w$, the vocabulary distributions of the two corresponding case slots are quite different. To address this property, we add the context feature which considers each context word and the relative probability of them to appear in each of the main case slots.

Comparison Between Proposed Models

The major difference between the two proposed models lies in how case frames for feature construction are decided.

As can be observed from Table 7, the joint prediction model achieved a better F-score than the combined case frame model. We conclude that deciding one best case frame is a better way for modeling the selectional preference of a predicate, compared to combining case frames with respect to the relevance scores. The result also verified the effectiveness of the joint model of case frames and shared arguments.

Error Analysis

In the following are several patterns of error observed in the system output. Examples of each error type are presented in Table 8.

1. **Error due to case frame granularity (Error Type 1):**

Our proposed model jointly predicts the most appropriate case frame along with the shared argument configuration. By selecting a single case frame for each predicate, we are able to model the selectional preference of the predicates accurately. However, the automatically constructed case frames do not always provide the granularity suitable for our task. If a coarse-grained case frame is selected dur-

ing prediction phase, the prediction of shared argument will also be affected.

For the example shown in Table 8, an appropriate case frame of the second predicate ‘put’ should contain words that supports $n-w$ shared argument in the wo-case. Table 9 represents the most appropriate case frame of the predicate ‘put’ among all the case frames of this predicate. It can be observed that although the wo-case contains words relevant to the $n-w$ shared argument, such as ‘letter’ and ‘envelope’, there are other irrelevant words dominating this case. These kind of broad, somewhat noisy case frames hinder the performance of our shared argument identification model.

2. Error due to event participants with similar characteristics (Error Type 2):

Our method relies largely on selectional preference information for identifying shared arguments. Thus, the prediction performance of our system is not very good for event pairs containing multiple participants with similar characteristics.

For the example shown in Table 8, our model wrongly identified the shared argument $n-g$. Although both cases are expected to hold human participants, the entity in the first ni-case should correspond to the victim of both actions ‘persecute’ and ‘kill’, while the second ga-case should hold the entity of the perpetrator of the two actions. In the scenario of the above event pair, there are two participants of similar characteristics, which are both expected to be human. Since selectional preference cannot effectively distinguish between these similar participants, our model often has difficulty dealing with event pairs with multiple similar participants.

3. Error due to fixed expression (Error Type 3)

In a fixed expression, an argument often takes on a different meaning than it usually does. Fixed expressions within events sometimes cause problems in shared argument identification. For the example shown in Table 8, the system output is as follows:

顔が明るくなる 太陽が顔を出す

(face-ga become brighter sun-ga face-wo appear)

Independently, both PASs shown above are plausible. However, the first PAS, ‘face-ga

become brighter’, means showing a cheerful look; while the second PAS, ‘sun-ga face-wo appear’, means sun rising. Although both expression contains the argument ‘face’, the shared argument of $g-w$ does not exist.

5 Conclusion

This paper proposed a method for shared argument identification in event relation knowledge acquisition. By addressing several problems of the previous works, we improved the shared argument identification model significantly. We proposed a richer feature representation of shared argument configuration which is more suitable for model learning. In order to incorporate different types of shared argument in the gold dataset, we update the most appropriate gold configuration along with case frames during the training process. We evaluated our model on a manually annotated gold dataset, and our model outperformed the baseline models by a large margin.

Our proposed model jointly predicts the shared argument configuration and the appropriate case frames. By comparing the result of our proposed model with the combined case frame model, we verified the effectiveness of this joint model to predict the appropriate case frames.

References

- Shuya Abe, Kentaro Inui, and Yuji Matsumoto. 2008. Acquiring event relation knowledge by learning cooccurrence patterns and fertilizing cooccurrence samples. *Proceedings of the 3rd International Joint Conference on Natural Language Processing* pages 479–504.
- Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised learning of narrative event chains. In *Proceedings of the 46th Annual Meeting of the Association of Computational Linguistics*. Association for Computational Linguistics, Columbus, Ohio, pages 789–797.
- Yuta Hayashibe, Daisuke Kawahara, and Sadao Kurohashi. 2015. Japanese case frame construction considering varieties of case patterns. Technical report.
- Shotaro Kohama, Tomohide Shibata, and Sadao Kurohashi. 2015. Argument alignment learning in event knowledge. In *Proceedings of the 21th Annual Meeting of the Association for Natural Language Processing (In Japanese)*. Association for Natural Language Processing, Kyoto, Japan, pages 1065–1067.

- Hector Levesque, Ernest Davis, and Leora Morgenstern. 2012. The winograd schema challenge. In *13th International Conference on the Principles of Knowledge Representation and Reasoning*. Association for the Advancement of Artificial Intelligence, pages 552–561.
- Jorge Nocedal. 1980. Updating quasi-newton matrices with limited storage. *Mathematics of computation* 35(151):773–782.
- Naoaki Okazaki. 2009. Classias: a collection of machine-learning algorithms for classification.
- Karl Pichotta and Raymond Mooney. 2014. Statistical script learning with multi-argument events. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, Gothenburg, Sweden, pages 220–229.
- Ataf Rahman and Vincent Ng. 2012. Resolving complex cases of definite pronouns: The winograd schema challenge. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, Jeju Island, Korea, pages 777–789.
- Tomohide Shibata and Sadao Kurohashi. 2011. Acquiring strongly-related events using predicate-argument co-occurring statistics and case frames. In *Proceedings of 5th International Joint Conference on Natural Language Processing*. Asian Federation of Natural Language Processing, Chiang Mai, Thailand, pages 1028–1036.