

Discriminative Approach to Predicate-Argument Structure Analysis with Zero-Anaphora Resolution

Kenji Imamura, Kuniko Saito, and Tomoko Izumi

NTT Cyber Space Laboratories, NTT Corporation

1-1 Hikarinooka, Yokosuka, Kanagawa, 239-0847, Japan

{imamura.kenji,saito.kuniko,izumi.tomoko}@lab.ntt.co.jp

Abstract

This paper presents a predicate-argument structure analysis that simultaneously conducts zero-anaphora resolution. By adding noun phrases as candidate arguments that are not only in the sentence of the target predicate but also outside of the sentence, our analyzer identifies arguments regardless of whether they appear in the sentence or not. Because we adopt discriminative models based on maximum entropy for argument identification, we can easily add new features. We add language model scores as well as contextual features. We also use contextual information to restrict candidate arguments.

1 Introduction

Predicate-argument structure analysis is a type of semantic role labeling, which is an important module to extract event information such as “*who did what to whom*” from a sentence. There are many arguments called *zero pronouns* that do not appear in the surface of a sentence in Japanese. In this case, predicate-argument structures cannot be constructed if we only rely on the syntactic information of a single sentence. Similar phenomena also happen in English noun predicates, in which arguments of noun predicates sometimes do not exist in the sentence due to things such as ellipses (Jiang and Ng, 2006). To correctly extract the structures from such sentences, it is necessary to resolve what zero pronouns refer to by using other information such as context.

Although predicate-argument structure analysis and zero-anaphora resolution are closely related, it was not until recently that these two tasks were lumped together. Due to the developments of large annotated corpora with predicate-argument and coreference relations (e.g.,(Iida et al., 2007))

and with case frames, several works using statistical models have been proposed to solve these two tasks simultaneously (Sasano et al., 2008; Taira et al., 2008).

In this paper, we present a predicate-argument structure analysis that simultaneously resolves the anaphora of zero pronouns in Japanese, based on supervised learning. The analyzer obtains candidate arguments not only from the sentence of the target predicate but also from the previous sentences. It then identifies the most likely arguments based on discriminative models. To identify arguments that appear in the sentence and are represented by zero pronouns without distinction, the analyzer introduces the following features and techniques: the language model features of noun phrases, contextual features, and restrictions of candidate arguments.

2 Predicate-Argument Structure Analyzer

2.1 Procedure and Models

The procedure of our predicate-argument structure analyzer is as follows. The input to the analyzer is an article (multiple sentences) because our target is to identify arguments spread across sentences.

1. First, each sentence is individually analyzed and segmented into base phrases by a morphological analyzer and a base phrase chunker. In Japanese, a base phrase is usually constructed by one or more content words (such as base noun phrases) and function words (such as case particles). In addition, dependency relations among base phrases are parsed by a dependency parser. In this paper, base phrases and dependency relations are acquired from an annotated corpus (i.e., correct parses).
2. Next, predicates are extracted from the base phrases. In general, a predicate is determined

	Name	Note
Baseline Features	Predicate	Form and POS of the predicate
	Noun	Form and POS of the head-word of the candidate phrase
	Particle	Form and POS of the particle of the candidate phrase
	Path	Dependency relation between the predicate and the candidate phrase
	Passive	Passive auxiliary verbs that the predicate contains
	PhPosit	Relative phrase position between the predicate and the candidate phrase
	SentPosit	Relative sentence position between the predicate and the candidate phrase
Additional Features (c.f., Sec. 2.2 and 2.3)	LangModel	Language model scores
	Used	Flag whether the candidate phrase was used as arguments of previous predicates
	SRLOrder	Order in Salient Referent List

Table 1: Features Used in this Paper

based on parts of speech such as verbs and adjectives. In this paper, the predicates are also provided from an annotated corpus.

3. Concurrently, noun phrases and their headwords are extracted as candidate arguments from base phrases. If an argument of a predicate is a zero pronoun, it is likely that the argument itself has appeared in previous sentences. Therefore, the analyzer collects not only all phrases in the sentence but also some phrases in the previous sentences. We also add the special noun phrase NULL, which denotes that the argument of the predicate is not required or did not appear in the article (i.e., exophoric).
4. Next, features needed for an argument identifier are extracted from each pair of a predicate and a candidate argument. Features used in this paper are shown in Table 1. Baseline features are roughly those of the predicate, the noun phrase, and their relations (on the phrasal/sentential sequence and the dependency tree). For binary features, we use all combinations of these features listed above.
5. Finally, the argument identifier selects the best phrases for nominative, accusative, and dative cases from the candidate arguments (Figure 1).

In this paper, we use maximum entropy models normalized for each predicate to each case. That is, the identifier directly selects the best phrase that

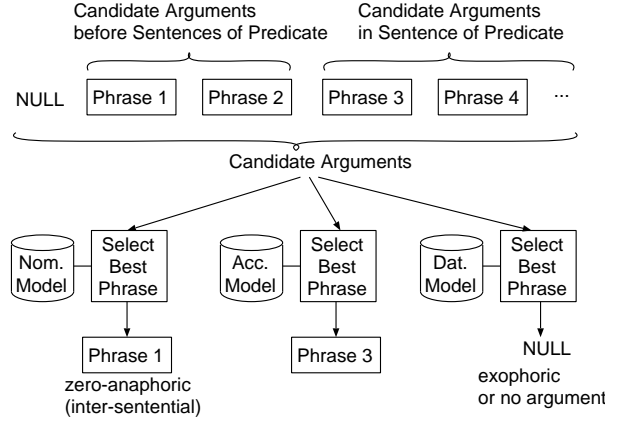


Figure 1: Summary of Argument Identification

satisfies the following equations from the candidate arguments:

$$\hat{n} = \operatorname{argmax}_{n_j \in N} P(d(n_j) = 1 | X_j; M_c) \quad (1)$$

$$P(d(n_j) = 1 | X_j; M_c) = \frac{1}{Z_c(X)} \exp \sum_k \{\lambda_{c_k} f_k(d(n_j) = 1, X_j)\} \quad (2)$$

$$Z_c(X) = \sum_{n_j \in N} \exp \sum_k \{\lambda_{c_k} f_k(d(n_j) = 1, X_j)\} \quad (3)$$

$$X_j = \langle n_j, v, A \rangle \quad (4)$$

where n , c , and v denote a noun phrase of an argument, the case, and the target predicate, respectively, N denotes a set of candidate arguments, $d(n)$ is a function that returns 1 iff the phrase n becomes the argument, and M_c denotes the model of the case c . In addition, $f_k(d(n_j) = 1, X_j)$ is a feature function, λ_{c_k} denotes a weight parameter of the feature function, and A denotes an article in which all sentences are parsed.

As shown, our analyzer can assign the best noun phrases to arguments regardless of whether they appear in the sentence or not by collecting candidates spread across multiple sentences. Furthermore, because the identifier is regarded as a selector based on the discriminative models, our analyzer has two properties: 1) New features can be easily added. 2) The precision can be improved by restricting the candidate arguments appropriately.

When we analyze predicate-argument structures and zero-anaphora resolution, syntactic information sometimes does not help because referents of zero pronouns do not appear in the sentence of the predicate. To overcome this problem,

we introduce additional information, i.e., language model scores and contextual information.

2.2 Language Models

Even if syntactic information does not help to identify arguments, we can expect that a certain noun phrase might be the correct argument of the predicate when we put it in place of the zero pronoun and the sentence becomes meaningful. Therefore, we add language model scores as features of the identifier. Because the appearance order of argument phrases is not strongly constricted in Japanese, we construct generation models that reflect dependency relations among a predicate, its case and a noun phrase. That is, we regard generation probabilities $P(n|c, v)$ acquired from the dependency tree as the scores of language models.

The language models are built from large plain texts by using a dependency parser. First, predicates and the base phrases that directly depend on the predicates are acquired from parsed sentences. Next, case particles and headwords are extracted from the base phrases. Finally, generation probabilities are computed using maximum likelihood estimation. Good-Turing discounting and backoff smoothing are also applied. Here, it is necessary to assign generation probabilities to NULLs. Regarding the training corpus that will be described in Section 3, the NULL rates of the nominative, accusative, and dative cases were 16.7%, 59.9%, and 81.6%, respectively. We assign these rates to the backoff term $P(\text{NULL}|c)$.

Using the language models, generation probabilities of the noun phrases are computed for every case of the predicate, and features that maintain the logarithms of language model scores are added ('LangModel' features in Table 1). Thus, the values of these feature functions are real.

2.3 Usage of Context

Centering theory claims that noun phrases that have been used once tend to be used again within the same context. We adopt this claim and add two different kinds of features. One is the feature that indicates whether a candidate has been used as an argument of predicates in the preceding sentences ('Used' features). However, the Used features are affected by the accuracy of the previous analyses. Thus, we also adopt the Saliency Reference List (Nariyama, 2002), which only uses explicit surface case markers or a topic marker, and added

	Training	Development	Test
# of Articles	1,751	480	695
# of Sentences	24,225	4,833	9,272
# of Predicates	67,145	13,594	25,500
# of Arguments			
Nom.	56,132	11,969	21,931
Acc.	26,899	5,566	10,329
Dat.	12,332	3,147	5,944

Table 2: Corpus Statistics

their priority order to the List as another feature ('SRLOrder' feature).

Another way to adopt contextual information is to restrict the candidate arguments. When we analyzed the training corpus from the viewpoint of zero pronouns, it was found that 102.2 noun phrases on average were required as candidate arguments if we did not stipulate any restrictions. When the candidate arguments we had restricted to those that had been used as arguments of the predicate appeared in a previous *one* sentence (namely, noun phrases appeared in more than one sentence before have a chance to remain), then the number of candidate arguments significantly decreased to an average of 3.2 but they covered the 62.5% of the referents of zero pronouns.

By using these characteristics, our analyzer restricts the candidate arguments to those that are of the same sentence, and those that were used as the arguments of another predicate in a previous sentence.

3 Experiments

3.1 Experimental Settings

Corpora: We used the NAIST Text Corpus version 1.4b (Iida et al., 2007) and the Kyoto Text Corpus 4.0 as the annotated corpora. We could obtain dependency and predicate-argument structures because these corpora were annotated to almost the same newspaper articles. We divided them into training, development, and test sets as shown in Table 2.

Argument Identification Models: Maximum entropy models were trained using the training set. In these experiments, we used the Gaussian prior, and the variance was tuned using the development set. Candidate argument restrictions were applied during both training and decoding.

Language Models: Language models were trained from twelve years of newspaper articles (Mainichi Shinbun newspaper 1991-2002, about

Case	Type	# of Args.	Prec.	Rec.	F
Nom.	Dep.	14,287	85.2%	88.8%	87.0%
	Zero-Intra	4,581	58.8%	43.4%	50.0%
	Zero-Inter	3,063	47.5%	7.6%	13.1%
	Total	21,931	79.4%	68.0%	73.2%
Acc.	Dep.	9,316	95.6%	92.2%	93.9%
	Zero-Intra	742	53.7%	21.6%	30.8%
	Zero-Inter	271	25.0%	0.4%	0.7%
	Total	10,329	94.3%	84.7%	89.2%
Dat.	Dep.	5,409	91.1%	72.6%	80.8%
	Zero-Intra	396	0.0%	0.0%	0.0%
	Zero-Inter	139	0.0%	0.0%	0.0%
	Total	5,944	91.1%	66.1%	76.6%

Table 3: Results on the Test Set

5.5M sentences) using the method described in Section 2.2. However, we eliminated articles that overlap the NAIST Corpus.

Evaluation: We evaluated the precision and recall rates, and F scores, all of which were computed by comparing system output and the correct answer of each argument. We also evaluated the rate at which all arguments of a predicate were completely identified as predicate-argument accuracy.

3.2 Results

The results are shown in Table 3. This table shows accuracies of the argument identification according to each case and each dependency relation between predicates and arguments. The predicate-argument accuracy on the test set was 59.4% (15,140/25,500).

First, focusing on the F scores of the Dep. relations, which denote a predicate and an argument in the same sentence and directly depend upon each other, scores of over 80% were obtained for all cases. Compared with Taira et al. (2008), they were higher in the nominative and accusative cases but were lower in the dative case. Overall, we obtained F scores between 73.2% and 89.2%.

Next, focusing on the intra-sentential (Zero-Intra) and inter-sentential (Zero-Intra) zero-anaphora, the analyzer identified arguments at some level from the viewpoint of precision. However, the recall rates and F scores were very low. The Zero-Inter recall rate for the nominative case, in which zero pronouns are centered, was only 7.6%. This is because our method preferred NULL phrases over unreliable phrases appearing before the predicate sentence. In fact, the analyzer output only 488 arguments, although the answer

was 3,063. To control the NULL preference is a future work for our analyzer.

4 Discussions and Conclusions

We proposed a predicate-argument structure analysis that simultaneously conducts zero-anaphora resolution. By adding noun phrases as candidate arguments that are not only in the sentence of the target predicate but also outside of the sentence, our analyzer identified arguments regardless of whether they appear in the sentence or not. Because we adopted discriminative models for argument identification, we can easily add new features. By using this property, we added language model scores as well as contextual features. We also used contextual information to restrict candidate arguments. As a result, we achieved predicate-argument accuracy of 59.4%, and accuracies of argument identification were F-scores of 73.2%–89.2%.

Verifying argument structures by language models evokes selectional preference of case frames. Sasano et al. (2008) has proposed statistical models using case frames built from 1.6 B sentences. Because the amount of the resources used in our study is quite different, we cannot directly compare the methods and results. However, because our analyzer has scalability that can freely add new features, for our future work, we hope to adopt the case frames as new features and compare their effect.

References

- Ryu Iida, Mamoru Komachi, Kentaro Inui, and Yuji Matsumoto. 2007. Annotating a Japanese text corpus with predicate-argument and coreference relations. In *Proceedings of the Linguistic Annotation Workshop in ACL-2007*, pages 132–139.
- Zheng Ping Jiang and Hwee Tou Ng. 2006. Semantic role labeling of nombank: A maximum entropy approach. In *Proceedings of EMNLP-2006*, pages 138–145.
- Shigeo Nariyama. 2002. Grammar for ellipsis resolution in Japanese. In *Proceedings of TMI-2002*, pages 135–145.
- Ryohei Sasano, Daisuke Kawahara, and Sadao Kurohashi. 2008. A fully-lexicalized probabilistic model for Japanese zero anaphora resolution. In *Proceedings of COLING-2008*, pages 769–776.
- Hirotohi Taira, Sanae Fujita, and Masaaki Nagata. 2008. A Japanese predicate argument structure analysis using decision lists. In *Proceedings of EMNLP-2008*, pages 523–532.