Language Modeling for Spoken Dialogue System based on Sentence Transformation and Filtering using Predicate-Argument Structures

Koichiro Yoshino* and Shinsuke Mori* and Tatsuya Kawahara*
* School of Informatics, Kyoto University
Sakyo-ku, Kyoto, 606-8501, Japan

***Abstract*—We present a novel scheme of language modeling for a spoken dialogue system by effectively exploiting the back-end documents the system uses for information navigation. The proposed method first converts sentences in the document, which are written and plain style, into spoken question-style queries, which are expected in spoken dialogue. In this process, we conduct dependency analysis to extract verbs and relevant phrases to generate natural sentences by applying transformation rules. Then, we select sentences which have useful information relevant to the target domain and thus are more likely to be queried. For this purpose, we define predicate-argument (P-A) templates based on a statistical measure in the target document. An experimental evaluation shows that the proposed method outperforms the conventional method in ASR performance, and the sentence selection based on the P-A templates is effective.**

I. Introduction

The task of spoken dialogue systems has been extended from simple transactions to general information navigation based upon users' requests. It is also desired to handle not only simple keyword-based queries, which current voice search systems respond to, but also users' vague and complex requests such as tourist guide and news briefing. In these tasks, there may be no or many answers to users' questions (e.g. "what is the best spot?"), but the system should provide the most relevant information through interaction with the users. These kinds of applications can be realized through document retrieval in the corresponding domain. For example, we can turn to tourist guidebooks or relevant Wikipedia entries for tourist domain [9]. Note that an intelligent dialogue system can be realized by limiting the domain and using the knowledge on the domain [5]. We have been developing an interactive news navigator which makes dialogue based on the news article archives [13]. The system not only responds to users' queries (e.g. "Did Seattle Mariners win last night?"), but also proactively presents a piece of news which will attract users' interest.

The automatic speech recognition (ASR) module for spoken dialogue systems (SDS) needs an appropriate language model (LM) adapted to the task domain and query style. Even a very large-vocabulary ASR system cannot cover proper nouns or named entities (NEs), which are critical in information retrieval. Ideally, LM should be trained with a large-scale matched corpus, but this assumption does not hold in many realistic cases. Therefore, two approaches are commonly adopted. The first approach is mixing document texts of the target domain with a dialogue corpus of spoken-style expressions. The other is collecting relevant texts, possibly in spoken-style sentences, from Web [10], [11], [8], [1]. These approaches try to cover the target domain and spoken-style in an indirect way, but the resultant model will inevitably contain a large amount of irrelevant texts.

We investigate a direct approach that generates spoken-style sentences from the written-style document texts of the target domain. A naive method would be to transform sentences in the document to spoken question-style which are expected for spoken dialogue systems [4]. However, every sentence in the document will not correspond to queries or questions, and every phrase of the relevant sentences may not be useful. In fact, the useful information structure is dependent on the domain, and information extraction techniques have been investigated [3]. Conventionally, the templates for information extraction were hand-crafted [7], but this heuristic process is so costly that it cannot be applied to a variety of domains on the web. We have proposed a method to automatically define domain-dependent templates for information extraction, which are used for a flexible information navigation system [13].

In this paper, we extend this approach to generate an appropriate LM for a spoken dialogue system. Specifically, we propose a method of two sequential processes. First, document texts are transformed to spoken question-style sentences by using dependency and predicate-argument (P-A) structures. Second, these sentences are filtered with domain-dependent P-A templates. In this way, we can predict sentences used for information extraction and navigation which are matched to the domain and style. The proposed scheme is applied to a domain of baseball news navigation [13], and we use a newspaper article database as a back-end document set.

II. Overview of Proposed Method

The overall flow of the proposed method is depicted in **Fig. 1**. First, sentences of the newspaper articles (=documents) are parsed by JUMAN[1] and KNP[2] to generate dependency and P-A structures. Here, we focus on dependencies to verbs. A P-A structure represents a sentence with a predicate and

[1]http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman.html
[2]http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp.html

TABLE I
SENTENCE TRANSFORMATION RULES.

Predicate type	POS	Rule
Inflected word	verb	conjunctive form + "*desu*" + "*ka*"
		conjunctive form + "*mashi*" + "*taka*"
	adjective	basic form + "*desu*" + "*ka*"
		basic form + "*deshi*" + "*taka*"
	adjective verb	verb stem + "*desu*" + "*ka*"
		verb stem + "*deshi*" + "*taka*"
Event-evoking noun	general noun	original form + "*desu*" + "*ka*"
		original form + "*deshi*" + "*taka*"
	verb formed by adding "*suru*" to a noun	original form +"*shi*" + "*masu*" + "*ka*"
		original form +"*shi*" + "*mashi*" + "*taka*"

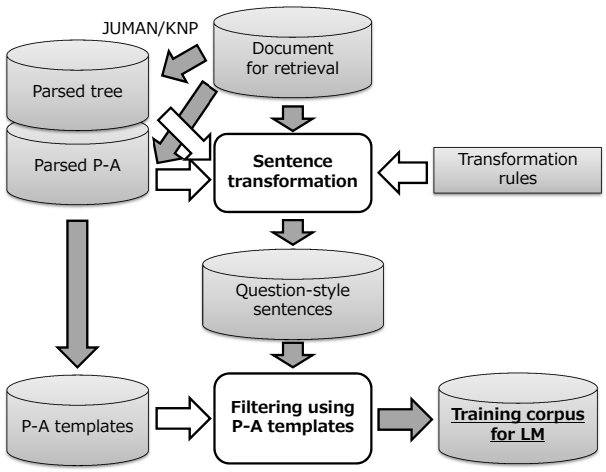

Fig. 1. Overview of proposed method.

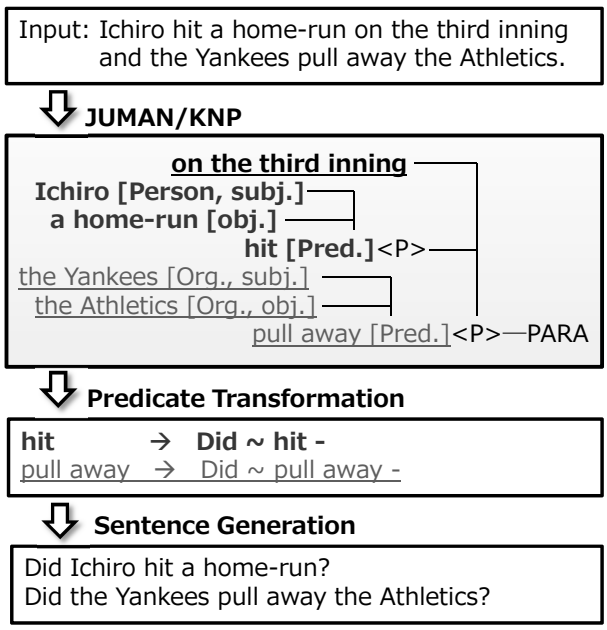

Fig. 2. Example of sentence transformation from document style to spoken question-style.

arguments with their semantic cases. The first step of the proposed method generates question-style sentences by using these structures and simple transformation rules. In the second step, the domain-dependent P-A templates are trained from the P-A structure analysis. They are used to filter the transformed question-style sentences. The resultant training corpus is expected to provide a necessary and sufficient lexicon and expressions for LM.

III. STYLE TRANSFORMATION BY USING SENTENCE STRUCTURES

We introduce sentence transformation by using dependency and P-A structures. The information navigation system poses two problems in LM for ASR. One is a gap between written and plain-style documents for retrieval and spoken question-style queries. In Japanese (and English), the main difference lies in the verbs or predicates. Thus, we identify verbs in sentences and segment the sentences into the P-A units. Simple transformation rules are applied to each unit to generate question-style sentences. The other problem is that there are many redundant phrases in documents which will not be used for queries. We focus on phrases that depends on verbs to generate natural questions.

An example of the processing is depicted in **Fig. 2**. In this example, two predicates are detected for transformation, and sub-trees of these predicates are used for question-style sentence generation. We currently use simple rules to generate yes/no questions, however, we can generate WH-type questions by using this information. For example, when the system finds the agent argument which is tagged as "Person", it can be converted to "who".

A. Predicate Transformation with Rules

First, the predicate in the sentence is identified. Predicates are classified into two types: Inflected words and event-nouns. There are three types of inflected words: verb, adjective and adjective verb. Event-nouns are similar to "Be+noun" or "Do+noun" in English [2], [6]. The transformation rules shown in **Table I** are applied according to the predicate type. When multiple rules are applicable, all of them are used to generate a variety of sentences.

B. Sentence Generation

After the transformation, sentences are generated by using sub-trees of the predicate. In the example shown in Fig. 2, the sub-trees of "hit" are "Ichiro", "a home-run" and "On the third inning". Thus, they are used to make respective sentences. The phrase "on the third inning" can be used for both questions, because it is shared with the predicates "hit" and "pull away".

IV. Sentence Selection with P-A Templates

We use P-A templates to define a useful information structure from the target document. They are used to filter the sentences generated in the previous section.

A. Extraction of Domain Dependent P-A Templates

The P-A structure automatically generated by the semantic parser provides a useful information structure. However, every P-A pair is not meaningful in information navigation; actually, only a fraction of the patterns are useful, and they are domain-dependent. For example, in the baseball domain, key patterns include "[A (agent) beat B (object)]" and "[A (agent) hit B (object)]", and in the business domain, "[A (agent) sell B (object)]" and "[A (agent) acquire B (object)]". We have proposed an automatic extraction method of P-A templates [13].

In the previous study, the extraction method based on Naive Bayes classifier was shown to be effective. In this method, probability of domain t (e.g. baseball) given word w_i is defined as,

$$P(t|w_i) \quad = \quad \frac{C(w_i,t) + D_t\gamma}{C(w_i) + \gamma}. \tag{1}$$

Here, $C(w_i)$ is a count of word w_i, and $C(w_i,t)$ is count of word w_i in domain t. γ is a smoothing factor which is estimated with a dirichlet prior [12] and D_t is a normalization coefficient of the corpus size of the domain t.

$$D_t = \frac{\sum_j C(w_j,t)}{\sum_k C(w_k)}. \tag{2}$$

The evaluation score of a P-A template is calculated as a mean of the scores of its components: predicate (p), argument (a) and its semantic case (s). We define two ways of this calculation. One uses a pair of a predicate and a semantic case as one word, the other uses an argument and a semantic case as one word.

$$\begin{cases} NB_{ps\_a}(t|P\text{-}A_i) = \sqrt{P(t|w_{ps}) \times P(t|w_a)} \\ NB_{p\_sa}(t|P\text{-}A_i) = \sqrt{P(t|w_p) \times P(t|w_{sa})} \end{cases} \tag{3}$$

The statistical method often encounters the problem of data sparseness due to mismatch between the training set and the test set especially in the named entities (NEs). To solve this problem, we conduct clustering to NEs which appear in the training set.

P-A structure

s = "**Ichiro hit a home-run on the third inning and the Yankees pull away the Athletics.**"
P-A = ["[Person]/subject/hit",
"a home-run/object/hit",
"[Organization]/subject/pull away",
"[Organization]/object/pull away"]

P-A templates

Score	Argument	case	Predicate
0.99599	middle relievers	subject	lose one's stuff
0.99519	relief pitcher	subject	lose one's stuff
0.98115	home-run	object	hit
0.78062	[Person]	subject	hit
0.70589	[Organization]	subject	pull away
0.70007	[Organization]	object	pull away
0.09994	share price	subject	slide
0.09994	charge	subject	increase
	...		

Scoring with P-A templates

NB_s = (0.98115+0.78062+0.70589+0.70007) / 4
= 0.7919325

Fig. 3. Example of Scoring.

B. Filtering with P-A templates

For each sentence generated by the method described in Section III, we calculate an evaluation score (NB_s) by taking an average for constitute P-A pairs.

$$NB_s = \frac{\sum_{i=1}^{n} NB(t|P\text{-}A_i)}{n}. \tag{4}$$

An example of the scoring is shown in **Fig. 3**. The input sentence s has four P-As. We calculate an average of their scores which are given by the corresponding P-A templates.

Then, the sentences are sorted with NB_s, and we make selection of the sentences of high scores for training of LM. With this method, we can select sentences which are more relevant to the target domain and more likely to be asked by users.

V. Experimental Evaluation

We evaluate the LM constructed by the proposed method. We prepared 201 questions. The target document set is a collection of the Mainichi Newspaper articles of ten years (2000-2009). There are 176,852 sentences which are tagged with the Japanese professional baseball domain. As a result of parsing, 500,523 predicates are extracted from these sentences (342,322 inflected words and 158,201 event-noun). Test-set word perplexity and word error rate (WER) are used as evaluation measures. We use adjusted perplexity to fairly compare LMs which have different sizes of vocabulary. To define the adjusted perplexity, we fixed the vocabulary size to 16,239. This is derived from the entire training corpus with cut-off 5.

For reference, we construct an LM with the conventional mixing method. We use the above newspaper corpus and 481,243 question-style sentences categorized as baseball in Yahoo! QA corpus [3], which is a set of queries collected through a web site.

[3]This corpus is provided by Yahoo!JAPAN and National Institute of Informatics, Japan.

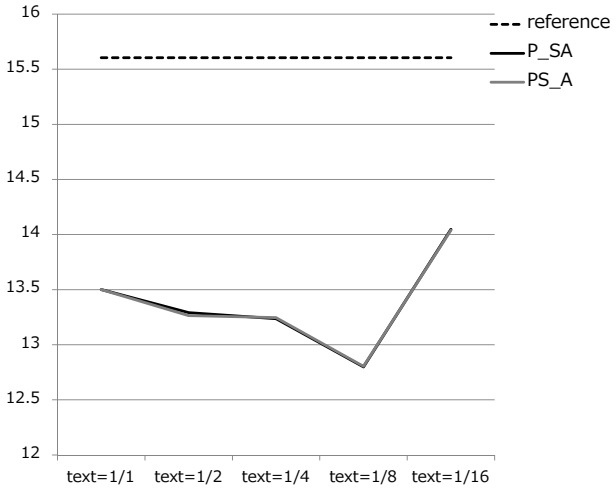

Fig. 4. Test-set word perplexity

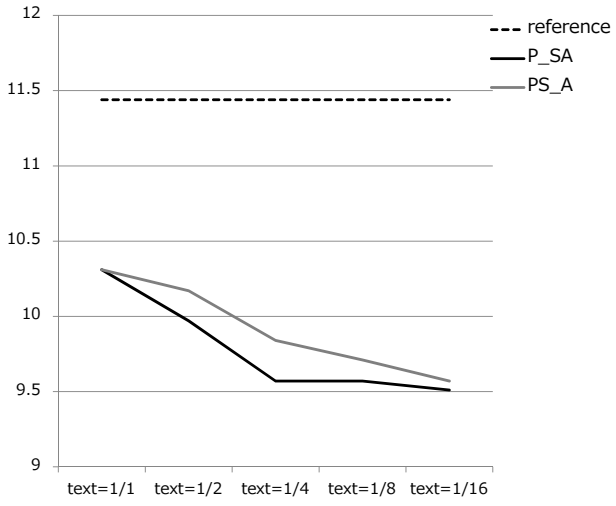

Fig. 5. Word error rate (WER) of ASR

To evaluate the effectiveness of filtering, we experiment by using sentences that rank in 50%, 25%, 12.5% and 6.25% of all. We show the graph of the perplexity in **Fig. 4** and WER in **Fig. 5**. In these graphs, PS_A and P_SA correspond to the methods for calculating the evaluation score NB_s which are defined in formula (3). The horizontal axis shows the quantity of the training data. The left-most case (text=1/1) applies the transformation only. The figures show that the proposed method outperforms the reference method and reduces the WER by 16.9% and perplexity by 18.0% at an optimal point.

Compared with the no-filtering case (text=1/1), the proposed sentence selection method reduces perplexity by 5.2% and WER by 7.8%. Thus, the proposed method is shown to be effective for LM construction for the domain-specific information navigation system.

VI. Conclusion

We have proposed a novel scheme of language modeling for the information navigation system. It consists of two processes: sentence transformation based on dependency analysis and simple rules, and sentence filtering based on P-A templates. This proposed method performs better than the conventional method without using a spoken question-style corpus. It is also shown that the sentence selection based on the P-A templates is effective. We plan to apply the method to a variety of domains in a large scale.

References

[1] Ivan Bulyko, Mari Ostendorf, Manhung Siu, Tim Ng, Andreas Stolcke, and Özgür Çetin. Web resources for language modeling in conversational speech recognition. *ACM Trans. Speech Lang. Process.*, 5(1):1:1–1:25, 2007.

[2] Jane Grimshaw. Argument structure. *MIT Press*, 1990.

[3] Ralph Grishman. Discovery methods for information extraction. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pages 243–247, 2003.

[4] Varga Istvan, Kiyonori Otake, Kentaro Torisawa, Stijn De Saeger, Teruhisa Misu, Shigeki Matsuda, and Jun'ichi Kazama. Similarity based language model construction for voice activated open-domain question answering. In *Proc. IJCNLP2011*, 2011.

[5] Tatsuya Kawahara. New perspectives on spoken language understanding: Does machine need to fully understand speech? In *Proc. IEEE-ASRU*, pages 46–50, 2009.

[6] Mamoru Komachi, Ryu Iida, Kentaro Inui, and Yuji Matsumoto. Learning based argument structure analysis of event-nouns in japanese. In *Proc. of the PACLING*, pages 120–128, 2007.

[7] L.Ramshaw and R.M.Weischedel. Information extraction. In *IEEE-ICASSP*, volume 5, pages 969–972, 2005.

[8] Teruhisa Misu and Tatsuya Kawahara. A bootstrapping approach for developing language model of new spoken dialogue system by selecting web texts. In *INTERSPEECH*, pages 9–13, 2006.

[9] Teruhisa Misu and Tatsuya Kawahara. Bayes risk-based dialogue management for document retrieval system with speech interface. *Speech Communication*, 52(1):61–71, 2010.

[10] Ruhi Sarikaya, Agustin Gravano, and Yuqing Gao. Rapid language model development using external resources for new spoken dialog domains. In *Proc. ICASSP*, volume 1, pages 573–576, 2005.

[11] Abhinav Sethy, Panayiotis G. Georgiou, and Shrikanth Narayanan. Building Topic Specific Language Models from Webdata Using Competitive Models. In *Proc. Interspeech*, pages 1293–1296, 2005.

[12] Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581, 2006.

[13] Koichiro Yoshino, Shinsuke Mori, and Tatsuya Kawahara. Spoken dialogue system based on information extraction using similarity of predicate argument structures. In *Proc. of SIGDIAL*, pages 59–66, 2011.