Abstract—This paper addresses issues in generating responses by extracting sentences from the Web for spoken decision-making dialogue systems. Various decision criteria are usually involved when selecting an alternative from a given set of alternatives. Such a dialogue system is required to explain the alternatives in terms of each decision criterion focusing on why the alternative is recommended. Preparation of a generation template is difficult for such explanations; thus we consider extracting and then adapting sentences retrieved from the Web. In this work, we address our preliminary experiment in selecting sentences appropriate for system responses.

I. INTRODUCTION

Over the years, a great number of spoken dialogue systems have been developed. Their typical task domains include airline information (ATIS & DARPA Communicator) [1] and railway information (MASK) [2]. Though most of them deal with database retrieval and transaction processing, the target of spoken dialogue systems has been extended to cover other tasks, such as planning [3] and tutoring [4]. We are also constructing a consulting dialogue system that assists user in making decisions [5], [6].

Spoken dialogue systems, in most cases, comprise modules of speech recognition, spoken language understanding, dialogue management, spoken language generation, and speech synthesis. In this work, we focus on issues in spoken language generation. A spoken language generation module, which translates the concept/action of the dialogue manager into natural language sentences, usually consists of a number of hand-crafted rules. These are usually costly to construct and tend to generate stereotypical sentences; thus more automated and flexible methodology is preferred.

Meanwhile, increasing number of Web text resources are being accumulated. These are garnering attention as collective intelligence, which potentially contains almost any desired kind of knowledge. We therefore consider utilizing Web text as a system response.

II. SPOKEN DECISION SUPPORT SYSTEM

We are constructing a sightseeing guidance system that assists users in making decisions [5], [7]. That is, the user selects an alternative from a given set of alternatives based on the system’s explanations.

There have been many previous studies of decision support systems in the operations research field, and the typical method that has been employed is the Analytic Hierarchy Process [8] (AHP). In the AHP, the problem is modeled as a hierarchy that consists of the decision goal, the alternatives for reaching it, and the criteria for evaluating these alternatives. In the case of our sightseeing guidance system, the goal is to decide on an optimal spot that is in agreement with the user’s preference. The alternatives are all sightseeing spots that can be proposed and explained by the system. As criteria, we adopt the “determinants” that we have defined in our tagging scheme of the Kyoto sightseeing guidance dialogue corpus [5]. The determinants include various factors that are used to plan sightseeing activities, such as “cherry blossoms”, “Japanese garden”, etc. A model hierarchy using these criteria is shown in Fig. 1.

When adopting such a hierarchy structure, the problem of deciding on the optimal alternative can be solved by estimating weights for criteria from user feedback. To realize such a decision making dialogue, the system must recommend sightseeing spots in terms of each decision criterion. Moreover, the explanation must contain the reasons for recommendation, in addition to a typical descriptive sentence.

We are thus constructing a database that comprises evaluation and explanation for the determinants. We selected several determinants that frequently appear in our dialogue corpus [5] (e.g., cherry blossoms, Japanese garden, scenery). An example from the database is shown in Table I. In the following sections, we specifically address issues in generating such a database. Handcrafting such explanations is costly, thus we consider automatic generation of sentences by extracting information from the Web.
### TABLE I

<table>
<thead>
<tr>
<th>Spot name</th>
<th>Determinant</th>
<th>Eval.</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kiyomizu temple</td>
<td>cherry blossoms</td>
<td>5</td>
<td>There are about 1,000 cherry trees on the temple grounds. Best of all, the scenery from the main temple are particularly stunning. The temple stage is built on the slope, and the views of the town from here are magnificent.</td>
</tr>
<tr>
<td>Imperial Palace</td>
<td>cherry blossom</td>
<td>5</td>
<td>A variety of cherry blossoms types are seen at location on the place grounds. In a special sighting, you can visit the cherry blossoms of “sakon”.</td>
</tr>
</tbody>
</table>

### TABLE II

<table>
<thead>
<tr>
<th>Example Human-Human Dialogue</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1: I recommend Kodaiji Temple.</td>
</tr>
<tr>
<td>U1: Kodaiji Temple.</td>
</tr>
<tr>
<td>G2: Have you ever been there?</td>
</tr>
<tr>
<td>U2: No, I haven’t.</td>
</tr>
<tr>
<td>G3: Yes, Kodaiji Temple is really beautiful. And, these are the two people that are important to explain this temple. This is Hideyoshi and his wife Nene.</td>
</tr>
<tr>
<td>U3: OK.</td>
</tr>
<tr>
<td>G4: So, if you’re not familiar with the history, let me just get you to Kodaiji Temple. It’s a trendy temple apparently.</td>
</tr>
<tr>
<td>U4: A trendy temple.</td>
</tr>
<tr>
<td>G5: A trendy temple, because all visitors come by this spot.</td>
</tr>
<tr>
<td>G6: And, the gardens are very famous here. And I wonder if I can get a better picture.</td>
</tr>
<tr>
<td>G7: Right, during the light-ups like during the cherry blossom season, and, in winter, they do lots of like light-ups and illuminations, it’s very beautiful. But regardless, the gardens are beautiful because they’re designed by a legendary tea ceremony master named Rikyu.</td>
</tr>
<tr>
<td>U6: Mmmm.</td>
</tr>
<tr>
<td>G8: ...and, he’s basically one of Japan’s best-known tea ceremony masters in Japan. So, I guess the gardens are just something you’ve got to take a look at.</td>
</tr>
</tbody>
</table>

### III. RESPONSE GENERATION FOR DECISION MAKING DIALOGUE SYSTEMS

A number of works have dealt with spoken language generation, and most have adopted generation templates or syntactic trees and rules [9], [10]. Such generation methods, in most cases, prepare rules for mapping from system dialogue acts (DAs) to template (natural language text with several variables). A template in bus information system, for instance, that maps the dialogue act of inform-time may be “There is a bus leaving [STOP NAME] at [DEPARTURE TIME]”. where values of variables are given by the dialogue manager. Natural language texts generated using such methodologies are, however, often formulaic, and more natural and spot-specific information is preferable.

Moreover, in our target task for decision support, the system is required to explain reasons for the recommendation, in addition to giving typical descriptive sentence. In fact, in our dialogue corpus [5] between a professional tour guide and a tourist, many of the guide’s utterances contain reasons (defending expressions) why the spot is recommended. An example dialogue between the guide and the user in our corpus is shown in Table II.

In this example, the guide gives reasons to recommend Kodaiji temple in terms of several perspectives (in this episode, “history” and “Japanese garden”), exemplifying named entities (NEs). It is difficult to generate such sentences using methods using generation templates, so we consider use of Web text for response generation.

In the sense that the system extracts sentences containing reasons from a large document set, our motivation is similar to that of why question-answering in order to answer a question “Why the spot is famous for [determinant]”. Most of conventional why question-answering systems use hand-crafted rules [11]. The goal of our research, however, is collecting evidence that supports the recommendation, thus the target sentences do not necessarily need to contain such expressions.

### IV. APPROACH

In this work, we address a “spoken language generation” method to construct a database for sightseeing decision support systems by extracting and adapting Web-based information. In our framework, the system:

1. Retrieves Web pages about the target sightseeing spot
2. Extracts candidate sentences from the Web pages that contain determinant-specific cue words
3. Checks if the Web texts are suitable as responses
4. Transforms the sentence (written in written style) to spoken style[12]

We assume here that existing web pages described specific sightseeing spots that are readily available. The main issue addressed in this paper is the above step 3.

### V. CLASSIFIER BASED SENTENCE SELECTION

#### A. Data collection and annotation

We selected 3 determinants of “Japanese garden,” “buildings,” and “cherry blossoms” and annotated the candidate sentences if they were appropriate for explaining recommendation of the spot in terms of the determinants. Annotators² were requested to select sentences naturally followed by

²Two annotators annotated the sentences, but we did not cross-labeled these sentences.
- Sentences annotated as appropriate reasons for recommendation
  - The Hojo garden, constructed by Motonobu Kano, is a famous dry garden style that adopted painterly techniques. (Japanese garden)
  - The current building is said to have been built in the Edo era, and it has been designated as a national important cultural property. (building)
  - The Fudan in the west garden is a rare cherry tree that continues to bloom from autumn to spring. (cherry blossoms)

- Sentences annotated as NOT appropriate as reasons for recommendation
  - There are two gardens at this spot, one is called the Kairyushiki garden and is located between the two main buildings. (Japanese garden)
  - There is a famous lantern, which is said to have been presented by Lord Mitsukuni, in the temple garden. (Japanese garden)
  - Today’s building was constructed. (building)

the sentences “I recommend that you visit the spot to see determinant because” with a little editing. They were requested to judge based on objective factors in the sentence rather than subjective expressions. Thus, sentences such as “I strongly recommend that you to visit Nijo-castle to see its beautiful garden” and “cherry” for the determinant of sentences that contain cue phrases. In Japanese, case markers indicate a focus (topic) of the candidate sentences.

TABLE III

<table>
<thead>
<tr>
<th>determinant</th>
<th># of sentence</th>
<th># of appropriate sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese garden</td>
<td>435</td>
<td>174</td>
</tr>
<tr>
<td>building</td>
<td>611</td>
<td>165</td>
</tr>
<tr>
<td>cherry blossom</td>
<td>206</td>
<td>112</td>
</tr>
<tr>
<td>Total</td>
<td>1,252</td>
<td>451</td>
</tr>
</tbody>
</table>

From several web sites, we collected 2,092 web pages describing sightseeing spots (e.g., web site of our Kyoto Keitai Project3). From the collected web pages, we extracted sentences that contain cue phrases that are manually given per determinant (e.g., “garden” for the determinant of Japanese garden and “cherry” for the determinant of cherry blossom). We resultantly obtained a test set consisting of 1,252 candidate sentences. The specification of the test set is listed in Table III. All candidate sentences are manually annotated on whether or not they are appropriate as reasons for recommendation.

B. Features used for classification

We then train a classifier that checks if the candidate sentences are appropriate as system responses. We use the following word and syntactic features obtained from the sentences.

1) Word feature: We use the following three word features.
   1) Surface: Nouns appear in the candidate sentence. We selected 3,000 words4 that frequently appear in the collected web pages. The element of the vector that corresponds to a specific noun is set to the number of times the noun appears.
   2) WC: The number of times word classes (WC) appear in the candidate sentence. We used Bunruigoihyo5 Japanese thesaurus to define word class, and selected 815 word classes that frequently appear in our web pages. The element of the vector that corresponds to a specific word class is set to the number of times the word class appears.
   3) NE: The number of times the named entity (NE) appears in the candidate sentence. The element of the vector that corresponds to a specific NE is set to the number of times the NE appears. We used nine entity classes defined in CaboCha Japanese dependency structure analyzer6. (e.g., person’s name, location, date)

The surface feature would be the most distinguishable feature when a number of training data is applicable. However, Our training data is small compared to other text classification tasks, thus we consider backing off to more extensive features of word class and named entity.

2) Syntactic feature: In general, when explaining the reasons for something, the focus or the subject of the utterance must concern the topic. Sentences whose focus is not the determinant, for instance Teiennno tonarini tatsu hondoha sekaisannni shiteisareteiru (“The main building next to the Japanese garden is designated a World Heritage Site.”), are not appropriate as reasons. A subject of a sentence is not clear (it is not always at the top of the sentence) in Japanese; thus we consider using the following syntactic features to check the focus (topic) of the candidate sentences.

4) CM: Case markers (CMs) of the bunsetsu7 containing cue phrases. In Japanese, case markers indicate a semantic role of a noun or pronoun contained in a bunsetsu. The semantic parser KNP is used to segment the sentence into phrase units called bunsetsu and to annotate the case markers of the bunsetsu. The element of the vector is set to “1,” if the case marker corresponds to the specific marker types. We used 26 case markers, which are defined by KNP. (e.g., ga [nominative], wo [accusative], ni [dative], to [with, that]. )
5) NE-DEP: Named entities contained in the bunsetsu depend on/depended by the bunsetsu containing the cue phrases. It is used to check if the NEs have strong relation with the topic of the sentence.

Using the above features, an SVM classifier was trained by TinySVM8.

C. Evaluation

1) Evaluation of classification using training data matched to determinant: First, we conducted an evaluation assuming several training data matched to the determinant are available.

3http://www2.nict.go.jp/pub/whatsnew/press/h21/090908/090908.html
4Threshold was experimentally set to 10.
5http://www.kokken.go.jp/kanko/goihyo/
6http://chasen.org/~taku/software/cabochoa/
7A bunsetsu is defined as a basic unit of Japanese grammar and it consists of a content word (or a sequence of nouns) followed by a case marker.
8http://chasen.org/~taku/software/TinySVM/
This assumes a situation with several available database entries (sightseeing spots in our case) that have explanatory sentences for the determinants, and then adding entries.

We evaluated the performance of classification by using a 10-fold cross validation by splitting the annotated data into 10 (set-1, ···, set-10), that is, one set was used as a test set to evaluate accuracy, and the nine others were used as training data. Classification results in F-measure defined by

$$F = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

are shown in Table IV. We show the performance of the case in which only a surface feature is used, one in which word features are used, one which syntactic features are used, and one in which feature selection is conducted from those in Subsection V-B.

Even when only the simple surface feature was used, we achieved a reasonable performance, but we achieved a better performance by combining syntactic features (cf. result by using feature selection and selected features). The results of precision and recall by the best feature were Japanese garden: (66.9%, 56.6%), building: (65.2%, 52.1%), cherry blossoms: (70.0%, 75.0%) respectively.

2) Evaluation using training data of different determinants: We then evaluated the classification performance for when no determinant specific training data is available. This assumes a situation where the system developer is adding determinants ("history," "scenery," etc.) that the system has not handled to the database.

We trained the classifier without using determinant specific training data. That is, one determinant dataset (e.g., Japanese garden) was used as test data and the others (e.g., building and cherry blossom) were used as training data. Classification results in F-measure are shown in Table V.

Classification performance using only the surface feature was significantly degraded. This result indicates that the word sets often used to provide explanation vary with determinants. On the other hand, by concatenating word class and NE features, we obtained higher performance. Using syntactic features in combination with word features, also enabled us to achieve much higher performance. These results suggest that the syntactic features are relatively general in extracting sentences that describe reasons for recommending in terms of determinants.

### VI. Conclusion

This paper addressed a framework for generating natural language response by extracting Web text for spoken decision support dialogue systems. We also conducted a preliminary experiment to check if the extracted sentence is appropriate as a response using SVM classifiers. We conducted evaluations assuming two situations of constructing a database for the system, and confirmed that we can obtain a reasonable performance in classification. Though we evaluated a simple binary classifier, The framework is wished to be extended to select appropriate sentences matched to dialogue contexts [13], [6].

### References


