Clause Classification-based Retrieval of Demand from Local Assemblies Minutes

Fumitoshi Ashihara  
Graduate School of  
Information Science and Technology  
Hokkaido University  
Kita 14, Nishi 9, Sapporo  
Hokkaido, 060-0814, Japan  
Email: ashihara@media.eng.hokudai.ac.jp

Yasutomo Kimura  
Department of  
Information and Management Science  
Otani University of Commerce  
Midori 3-5-21, Otani  
Hokkaido, 047-8501, Japan  
Email: kimura@res.otaru-uc.ac.jp

Kenji Araki  
Graduate School of  
Information Science and Technology  
Hokkaido University  
Kita 14, Nishi 9, Sapporo  
Hokkaido, 060-0814, Japan  
Email: araki@media.eng.hokudai.ac.jp

Abstract—In this paper we present a method for estimating clauses that express demands within the questions of city assemblymen, in order to achieve effective use of local assembly minutes which are available to the public via the Web. One of the problems of our task is that modality is not connected to a clause. Therefore, we apply clause-classification to solve this problem. Experimental results show the effectiveness of our approach in comparison with the baseline. We obtained an F-measure increase of 4.6 points with only modality and an F-measure increase of 9.3 points with both modality and clause-classification.

Keywords—demand expression, modality, clause-classification

I. BACKGROUND

In recent years, local governments use websites to disclose administrative information to the public, and information about local councils is also available on the Web. Local councils have the functions of determining the basic principles of the local government and monitoring and evaluating the government. In the local assembly minutes, discussions on topics such as political issues and budget decisions are recorded during all meetings and sessions. However, despite the fact that local assembly minutes are openly available on the Web, this record is rarely read, because of technical contents and spoken language description which is not elaborated on. Therefore, effective use of local assembly minutes is needed for people concerned with local governments.

Previously, there have been some studies in Japan which used minutes as data. Yamamoto studied paraphrasing for gree-nundant expressions[1] and Kimura attempted to extract word phrases relating to political issues[2]. However, most studies have not focused on the opinions of city assemblymen, which is the main theme of the meetings. Recently, research has been conducted on opinion mining that focuses on understanding the user’s opinion from Web data[3]. In political text, there has been research by Diermeier that attempts to classify US senators’ ideological positions in order to provide information to help voters make more informed decisions[4]. In local assembly minutes, also, understanding city assemblymen’s opinions is an important factor for grasping the theme of the meeting.

In this study, we focus on city assemblymen’s utterances that include demands, which have been studied as “demand expressions” in CGM(Customer Generated Media) data and as “request intentions” in the Japanese Public Comment system[5][6]. Example (1) below shows an example of a city assemblyman’s demand utterance.

(1) This research is based on random-sampling, but you should conduct a complete survey of elementary schools, and utilize the results in the next policy. I would like to ask the superintendent of education’s opinion. (In original Japanese: この調査は抽出調査ということでありますが、道教委として小中学校の被調査者を実施し、その結果を今後の調査に活かすべきと考えますが、教育長の見解をお伺いいたします。)

In local assembly minutes, this kind of utterance expresses the arguments made by city assemblymen. Therefore, mining demand utterances is the key to understanding political issues in local governments.

A. Clause-Level Approach

In order to achieve retrieval of assemblymen’s demands, it is necessary to judge whether the target sentence includes a demand expression or not. Ohtsuka et al. applied the supervised learning utilizing BagOfWords to this problem[6]. On the other hand, Kanayama and Nasukawa applied automatic syntactic pattern induction in addition to recognizing the targets of the demands[5]. In contrast to these studies, our local assembly minutes text data tends to be long and complex because it consists of spoken language. Accordingly, even if we judge whether a sentence includes a demand or not, the extracted sentence would be not easy to understand.

To deal with this problem, we attempt to judge whether clauses, rather than sentences, include demands or not. Since clauses are more fine-grained units than sentences, and are meaningful to an extent, judging demand per clause enables us to extract segments which are easy to understand from local assembly minutes.

We call this task “clause-level demand retrieval”. In this task, a sentence is divided into clauses C, and we judge whether each clause C_i is a demand or not. Figure 1 illustrates a clausal structure of Example(1).
A. Definition of Demand

In this subsection, we defined the annotation criteria of demand as follows referencing the researches of Ohtsuka which describes the definition of demand in Japanese[6].

- Contents of the clause are hoped for by the speaker.
- The agent of the verb is the person or organization solving the issue.

For example, $C_3$ in Figure 1 is judged as a demand because the agent of the verb of $C_3$ is the local government, and the contents are hoped for as the clause contains the modality word “べき (beki): should”. Moreover, $C_2$ is judged as a demand because the agent of $C_2$ is also the local government, and the contents are hoped for. Using this definition, indirect demands such as “This room is hot”, which sometimes means “Open the window”, are excluded, because this sentence does not meet the second criterion. The reason for the necessity of this exclusion is that these indirect demands may cause deviations in human judgment, and it is necessary to prevent targeting utterances which the speaker (assemblyman) did not intend as a demand.

Based on this definition of demand, we annotated a corpus of Japanese local assembly minutes. The average agreement value of three persons was $\kappa = 0.45$, which is lower than in the studies of Ohtsuka et al. and Kanayama and Nasukawa, because this annotation is not a sentence-level agreement but a clause-level agreement[5][6]. However, the 0.45 kappa value is generally considered as medium agreement; thus, we concluded that this annotation is feasible.

In addition, it should be noted that in the annotated demands, there is a category of “Demand an Explanation”. This contains no political issues, but is an assemblyman’s demand for effective utilization of investigation results.

B. Structure of this Paper

In Section II, we count co-occurrences of modality and clause-classification to investigate what kind of clause-classification should be taken into account for supervised learning. Next, in Section III, we describe an experiment to investigate whether the modality and clause-classification pattern is effective for demand retrieval or not, based on the results of Section II. Finally, in Section IV, we discuss and conclude the results of our research.

II. DEMAND IN LOCAL ASSEMBLY MINUTES

In this section, we will discuss demand utterances in local assembly minutes. First, we will define “demand” in order to determine annotation criteria and calculate agreement value. Secondly, we will analyze a corpus of local assembly minutes gathered from Hokkaido and Otaru City in Japan based on this definition.

B. Corpus Analysis

In this section, we investigate co-occurrences of clause-classification and modality, which should be the features of clause-level demand retrieval. This analysis is conducted from the following viewpoints.

- What kind of clause-classification has co-occurrence with modality word?
- What kind of clause-classification transitions from modality words within demand clauses?

The first viewpoint is necessary for narrowing down the number of clause-classifications; the second is important for confirming the transition of clause-classification when the target clause does not contain a modality word. To clarify these viewpoints, we conducted a statistical investigation shown in Procedure 1. In addition, each $C_i$ in Procedure 1 consists of words split using MeCab.1

Procedure 1
Count the frequency of co-occurrences of modality and clause-classification in each transition number.

INPUT clause-classification set cc, modal verb set m_{mv}, positive word set m_{pw} from TABLE I.
OUTPUT the number of each kind of modality, transition number, clause-classification number_{x,y,z}

for each sentence s in the dataset do
    for each clause C_i of clauses C in s do
        clause-classification cc ← “”
        for each clause C_j after C_i (j = i + 1 to |C| and (j−i) ≤ 3) do
            cc = cc+ “ ” + cc_{C_{j−1}}
        end for
        {Store transition of clause classification}
        if ∃m_{mv} ∈ C_j then
            number of transition t ← j − i
            INCREMENT number_{mv,t,cc}
            {Count the frequency of co-occurrences of modal verbs with cc, t}
            break
        else if ∃m_{pw} ∈ C_j then
            INCREMENT number_{pw,t,cc}
            {Count the frequency of co-occurrences of positive words with cc, t}
            break
        end if
    end for
end for

TABLE 1. TARGET MODALITIES AND CLAUSE-CLASSIFICATIONS

<table>
<thead>
<tr>
<th>m_{mv} (modal verbs)</th>
<th>cc</th>
<th>m_{pw} (positive words)</th>
<th>cc (Clause-classification)</th>
</tr>
</thead>
</table>
| "べき" (beki): should do
   "でほしい" (dehoosi): want you to do...
   "なければならない" (nakereba): unless you do...
   "くださき" (kudasai): would you please do...
   "いったい" (itaideki): I would like you to do...
| "重要" (jyuuou): be important to do...
   "必要" (hituyou): be necessary to do...
   "普通" (jitsuu): be normal to do...
   "不可欠" (fukaketsu): be indispensable to do...
   "責任" (sekinin): have a duty to do...
   "責任ある" (senshin): have a responsibility to do...
   "意見" (iken): have the opinion that you do...
   "希望" (kibou): have a hope that you do...
   "当然" (touou): take it granted for you to do...
   "求め" (motome): need you to do...
   "こそ" (kosou): be exactly...
| "適用" (renyou): continuous
   "並用" (heiretusu): coordinate
   "理由" (ryou): reason
   "条件" (joken): conditional
   "時間" (jikan): time
   "課歩" (kose): concessive
   "補足" (hosou): noun (complement)
   "間接質問" (kansetsusoumon): indirect interrogative:
   "連体" (renni): attributive
   "文末" (bunmatsu): the end of a sentence

In order to ascertain the relationship between clause-classification and modality type, the target modality words are classified into modal verbs like “べき (beki): should” and words which represent the positiveness of an action like “必要 (hituyou)”, which means “necessary”. Such positive words are used as in the sentence below.

- “法人2税をふやすには、新設を主なターゲットとしていく必要があると考えます。”
  (In English: “In order to increase these two kinds of corporation tax, I think it is necessary to proceed by targeting mainly the establishment of new corporations.”)

Although this kind of expression is not often used as demand on the Web or in a normal corpus, it is frequently used as demand expression in local assembly minutes.

The clause-classifications cc shown in Table I were analyzed with CBAP, which analyzes Japanese clause-classifications and clause boundaries of a sentence using surface information[7].

Figure 2 shows an example of this statistic. In this case, “2 transitions: Continuative→Noun” and “1 transition: Noun” are counted. The conditions of analysis are shown below.

In English:
"It is necessary for you to demand (Hokkaido Air System) to clarify responsibilities, engage an external management consultant to enhance the management system and reform drastically. I would like to ask your opinions."

Fig. 2. Method of taking statistics of the transition of clause-classification to modality

- The number of data is 5,670 annotated clauses.
- The data are taken from the regular meetings of Hokkaido and Sapporo local governments in 2011.
- The number of transitions is limited to three, because the number of data with more than three transitions was two or lower.

C. Discussion

Figure 3 shows the results of the corpus analysis.

From these results, we can state the existence of the following tendencies.
Fig. 3. The result of Corpus Analysis

- Modality words (modal verbs) appears in the clause that follows the continuative clause.
- Modality words (positive words) appears in the clause that follows the attributive and noun clauses.
- If modality appears in the clause following several clauses, the clause-classification transitions with the continuative clause.

From these tendencies, we confirmed that it is possible to narrow the clause-classification candidates down to combinations of modality and clause-classification, which are “modal verb and continuative” and “positive word and (attributive or noun)”, and the transition of demand clause to modality through the continuative.

III. EVALUATING CLAUSE-LEVEL DEMAND RETRIEVAL

In this section, we evaluate clause-level demand retrieval with an SVM classifier as supervised learning. The purpose of this experiment is to determine whether the combination of modality and clause-classification is effective or not.

A. Experiment Settings

We use LibSVM as the machine learning library, and the SVM kernel is a 2-degree polynomial kernel.

The estimation label consists of three categories: “No Demand” (negative example), “Demand”, and “Demand an Explanation”. The dataset is 8,640 clauses, which consist of 5,670 data used in the statistical investigation and 2,880 additional data.

Table II shows the features. These features are binary features whereby if the condition is true, the value of the feature is 1, and if false, the value is 0.

We adopted (a) to (f) as baseline features, which contain BagOfWords and other syntactic patterns. Our proposed modality words and clause-classification patterns (h) are shown in Table III. For the modality pattern features, we used modality words in Table I.

The reasons for applying baseline features (a) to (f) are as follows.

- (a) $x_{POS}$: Adopt syntactic connection pattern between $C_t$ and $C_{t+1}$.
- (b): Incorporate the tendency that demand clauses are positioned around the last part of the sentence.
- (c) and (f): Compare to other syntactic feature with reference to Eguchi et al.’s research [8].

---

2 http://www.csie.ntu.edu.tw/cjlin/libsvm
3 The C(cost) parameter is 0.01.
4 Although “Demand an Explanation” is not directly related to the current task, we analyzed it for topic change detection, which is a future task.
5 "くだささい (kudasai)" and "いただきます (itadaki)" were excluded for improved recognition.
(a) $x_{\text{BOW}}$: BagOfWords of $C_1$.
$x_{\text{POS}}$: Part-of-speech trigram of the last word of $C_i$ and two foremost words of $C_{i+1}$.

(b) Relative position of $C_i$, $x_d(i = 1, 2, ..., 10) \begin{cases} 1 & \text{ceil} \left( \frac{i}{10} \times 10.0 \right) = i. \\ 0 & \text{(otherwise)} \end{cases}$

(c) $x_{\text{fe}} \begin{cases} 1 & t \in (C_i \cup C_{i+1}). \\ 0 & \text{(otherwise)} \end{cases}$
where $t$ is a functional expression of the modal verb in "Tsutsuji" 6.
This is represented as a semantic code such as D1, which contains words meaning "should". 

(d) $x_{\text{interrogatives}}$ such as "what" (nami), "how" (do-yatte), and "in order to" (muke): "in order to" $\notin (C_{i-1} \cup C_i \cup C_{i+1})$.

(e) $x_{\text{adverbial}} \begin{cases} 1 & \text{ "in order to" or "towards" } \notin C_i. \\ 0 & \text{(otherwise)} \end{cases}$

(f) $x_{\text{dp}} \begin{cases} 1 & (m_{\text{mov}} \cup m_{\text{pw}}) \in (DP_1 \cup DP_2 \cup DP_3), \text{ where } DP_0 = \text{ a verb of } C_{i-1}. \\ 0 & \text{(otherwise)} \end{cases}$

(g) $x_{\text{m}} \begin{cases} 1 & (m_{\text{mov}} \cup m_{\text{pw}}) \in (C_{i+1} \cup C_{i+2} \cup C_{i+3}). \\ 0 & \text{(otherwise)} \end{cases}$

(h) $x_{\text{pt}} \begin{cases} 1 & \text{The patterns in Table III match.} \\ 0 & \text{(otherwise)} \end{cases}$

### TABLE III. Clause-Classification and Modality Combination Patterns

<table>
<thead>
<tr>
<th>Conditions for pattern formation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Continuative Pattern</td>
</tr>
<tr>
<td>B: Attributive Pattern</td>
</tr>
<tr>
<td>C: Noun Pattern</td>
</tr>
<tr>
<td>D: 2, 3 Transitions Pattern</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

- (d): Exclude indirect interrogative clause.
- (e): Exclude meaningless adverbial clauses which include "muke": "towards", and clauses which are not judged as demands including "towards".

The feature sets for comparison with the baseline are ii (consider only modality) and iii (consider modality and clause-classification), as shown in Table IV.

### Table IV. Feature Set For Experiment

| i (Baseline: No modality and clause-classification) | [a]~[f] |
| ii (Only modality) | i and [g] |
| iii(Combination patterns of modality and clause-classification) | ii and [h] |

6http://kotoba.nuee.nagoya-u.ac.jp/tsutsuji/  
7We obtained Japanese dependency structure by using analyzer Cabocha.  
http://code.google.com/p/cabocha/
For these three feature sets, we evaluate precision, recall, and F-measure by 10-fold cross-validation.

B. Experiment Results

Table V shows the results of the experiment.

<table>
<thead>
<tr>
<th>TABLE V. EXPERIMENTAL RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not demand</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>i (Baseline)</td>
</tr>
<tr>
<td>ii</td>
</tr>
<tr>
<td>iii</td>
</tr>
<tr>
<td>Demand An Explanation</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>i (Baseline)</td>
</tr>
<tr>
<td>ii</td>
</tr>
<tr>
<td>iii</td>
</tr>
<tr>
<td>Demand</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>i (Baseline)</td>
</tr>
<tr>
<td>ii</td>
</tr>
<tr>
<td>iii</td>
</tr>
</tbody>
</table>

The result for “Demand” shows improvement with feature set iii. Compared to the baseline, it was observed that an improvement of 4.6 points is shown in feature set ii, and an improvement of 9.3 points is shown in feature set iii. In particular, an improvement of 12.0 points in recall was observed in feature set iii. From these results, we can say that the number of recognized positive data is increased by the clause-classification and modality pattern.

IV. Conclusion

To solve this problem, we designed “clause-level demand retrieval”, which is suitable for spoken language, and conducted judgment of clauses containing demands using supervised learning. Within this approach, we applied features that consider the combination of modality and clause-classification for improving accuracy.

The first part of this paper described a corpus analysis relating to demand expressions in local assembly minutes. Thereby, the tendencies of clause-classification and modality patterns were made clear. With this analysis, we suggested using the patterns of modality and clause-classification in an SVM classifier. The last part of this paper described an experiment that confirmed the effectiveness of the combined patterns of clause-classification and modality. The experimental results showed an improvement of estimation with our method.

On the other hand, it still remains difficult to estimate clauses that have a large number of transitions, and other positive instances which we cannot tackle with our modality lexicon. Our perspective is that the positive instances which we cannot estimate often emerge in conditional clauses and with complex modality patterns. In future work, we will increase the amount of training data and the modality lexicon so as to tackle complex patterns.

ACKNOWLEDGMENT

This work was supported by KAKENHI Grant Number 22300086.

REFERENCES