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
Otaru University of Commerce
Midori 3-5-21,Otaru
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 greenundant 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悉皆調査を実施し，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 retrievial 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 $\mathbf{C}$, and we judge whether each clause C_i is a demand or not. Figure 1 illustrates a clausal structure of Example(1).

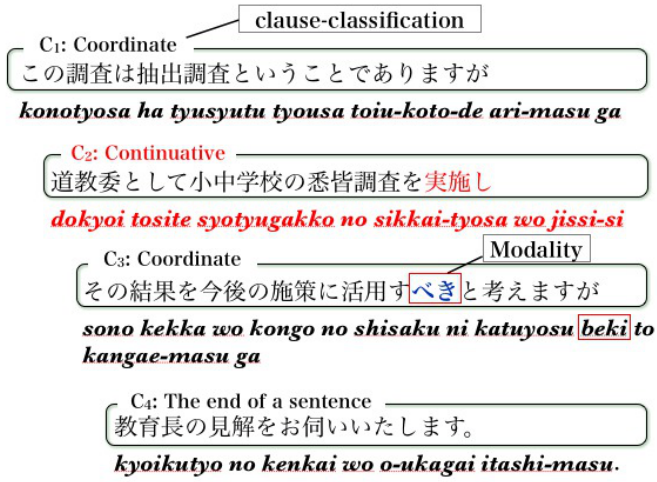


Fig. 1. **Demand utterances in local assembly minutes**

In order to judge demand, modality words such as modal verbs and evaluative expressions are important. The clause including modality in Figure 1 is C_3 which contains the word “べき (*beki*): should”. C_3 is judged as a demand because C_3 consists of an assemblyman's demand for effective utilization of investigation results.

However, in clause-level demand estimation, judging using only modality words is difficult. For instance, although there is no modality word in C_2, we can consider it a demand because it consists of an assemblyman's demand for a complete survey of elementary schools. To deal with such clauses, information on syntactic patterns such as the dependency relation of C_2 and C_3 is necessary.

To obtain such syntactic information, we propose using clause-classification for clause-level demand retrieval. For example, the clause-classification of C_2 is Continuative, which shows a dependency relation of C_2 and C_3. We assume that a higher score can be obtained by applying clause-classification and modality words to the features of supervised learning than by only using the BagOfWords feature, which is generally applied.

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 gathegreen from Hokkaido and Otaru City in Japan based on this definition.

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 considegreen 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 to a mayor or government official for an explanation. It does need to be analyzed because judging this kind of demand with supervised machine learning is already highly accurate. Therefore, in the following corpus analysis, we exclude this category.

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].

[1]http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html

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

INPUT clause-classification set **cc**, modal verb set $\mathbf{m_{mv}}$, positive word set $\mathbf{m_{pw}}$ 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 $cc \leftarrow$ ""
    for each clause C_j after C_i
    (j = i + 1 to $|\mathbf{C}|$ and $(j-i) \leq 3$) **do**
      $cc = cc +$ " → " $+cc_{C_{j-1}}$
      {Store transition of clause classification}
      if $\exists \mathbf{m_{mv}} \in Cj$ **then**
        number of transition $t \leftarrow j - i$
        INCREMENT $number_{mv,t,cc}$
        {Count the frequency of co-occurences of modal verbs with cc, t}
        break
      else if $\exists \mathbf{m_{pw}} \in Cj$ **then**
        INCREMENT $number_{pw,t,cc}$
        {Count the frequency of co-occurences of positive words with cc, t}
        break
      end if
    end for
  end for
end for

TABLE I. TARGET MODALITIES AND CLAUSE-CLASSIFICATIONS

$\mathbf{m_{mv}}$(modal verbs)	"べき" (*beki*) : should do... "てほしい" (*tehoshi*) : want you to do ... "なければ" (*nakereba*) : unless you do... "ください" (*kudasai*) : would you please do... "いただき" (*itadaki*) : I would like you to do...
$\mathbf{m_{pw}}$(positive words)	"重要" (*juuyo*) : be important to do... "必要" (*hitsuyo*) : be necessary to do... "普通" (*futsuu*) : be normal to do ... "急務" (*kyuumu*) : be urgently necessary... "不可欠" (*fukaketsu*) : be indispensable to do... "責務" (*sekimu*): have a duty to do... "責任" (*sekinin*) : have a responsibility to do... "意見" (*iken*) : have the opinion that you do... "希望" (*kibo*): have a hope that you do... "当然" (*tozen*) : take it granted for you to do... "求め" (*motome*) : need you to do... "こそ" (*koso*) : be exactly...
cc(Clause-classification)	"連用" (*renyou*) : continuative "並列" (*heiretsu*): coordinate "理由" (*riyu*) : reason "条件" (*joken*): conditional "時間" (*jikan*): time "譲歩" (*joho*) : concessive "補足" (*hosoku*) : noun (complement) "間接疑問" (*kansetsugimon*) indirect interrogative: "連体" (*rentai*) : attributive "文末" (*bunmatsu*): the end of a sentence

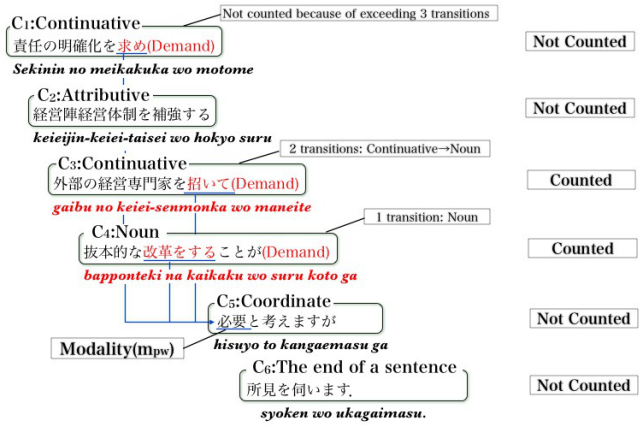
In order to ascertain the relationship between clause-classification and modality type, the target modality words are classified into modal verbs like "べき (*beki*): should" and words which represent the positiveness of an action like "必要 (*hitsuyo*)", which means "necessary". Such positive words are 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.

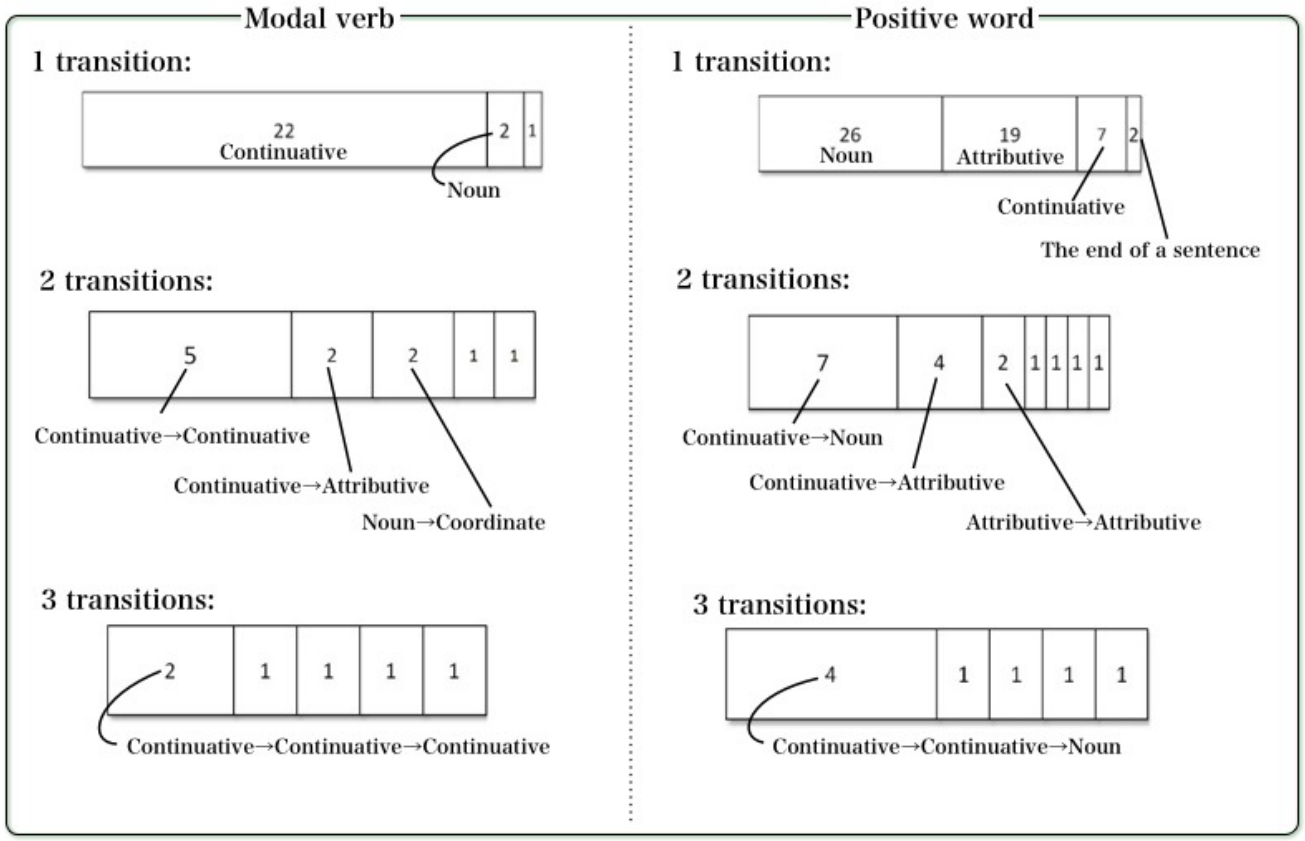


Fig. 3. **The result of Corpus Analysis**

- 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, [2] and the SVM kernel is a 2-degree polynomial kernel[3].

[2]http://www.csie.ntu.edu.tw/cjlin/libsvm

[3]The C(cost) parameter is 0.01.

The estimation label consists of three categories: "No Demand" (negative example), "Demand", and "Demand an Explanation"[4]. 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 $\mathbf{x}$. 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[5]

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].

[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]"ください (*kudasai*)" and "いただき (*itadaki*)" were excluded for improved recognition.

TABLE II. FEATURE LIST

(a)	$\mathbf{x}_{\mathrm{BOW}}$: BagOfWords of C_t. x_{POS} : Part-of-speech trigram of the last word of C_t and two foremost words of C_{t+1}.
(b)	Relative position of C_t, $\mathbf{x}_{\mathrm{d_i}}(i = 1, 2, ..., 10)$ $\begin{cases} 1 & \mathrm{ceil}(\frac{t}{\lvert C\rvert} \times 10.0) = i. \\ 0 & (otherwise) \end{cases}$
(c)	$\mathbf{x}_{\mathrm{fe}}$ $\begin{cases} 1 & \mathbf{t} \in (C_t \cup C_{t+1}). \\ 0 & (otherwise) \end{cases}$ where $\mathbf{t}$ is a functional expression of the modal verb in “Tsutsuji”[6]. This is represented as a semantic code such as D11, which contains words meaning “should”.
(d)	x_{irg} $\begin{cases} 1 & \textbf{interrogatives} \text{ such as なに” (}\textit{nani}\text{): “what”, or “どうやって (}\textit{do-yatte}\text{) : “how”} \notin (C_{t-1} \cup C_t \cup C_{t+1}). \\ 0 & (otherwise) \end{cases}$
(e)	$\mathbf{x}_{\mathrm{adverbial}}$ $\begin{cases} 1 & \text{“ため” (}\textit{tame}\text{): “in order to” or “向け” (}\textit{muke}\text{): “towards”} \notin C_t. \\ 0 & (otherwise) \end{cases}$
(f)	$\mathbf{x}_{\mathrm{dp}}$ $\begin{cases} 1 & (\mathbf{m}_{\mathrm{mv}} \cup \mathbf{m}_{\mathrm{pw}}) \in (DP_1 \cup DP_2 \cup DP_3). \\ 0 & (otherwise) \end{cases}$ where DP_0 = a verb of C_t, DP_i = dependent word[7] from DP_{i-1}.
(g)	$\mathbf{x}_{\mathrm{m}}$ $\begin{cases} 1 & (\mathbf{m}_{\mathrm{mv}} \cup \mathbf{m}_{\mathrm{pw}}) \in (C_{t+1} \cup C_{t+2} \cup C_{t+3}). \\ 0 & (otherwise) \end{cases}$
(h)	$\mathbf{x}_{\mathrm{pt}}$ $\begin{cases} 1 & \text{The patterns in Table III match.} \\ 0 & (otherwise) \end{cases}$

TABLE III. CLAUSE-CLASSIFICATION AND MODALITY COMBINATION PATTERNS

	Conditions for pattern formation
A: Continuative Pattern	$x_{\mathrm{c_1}}$ The clause-classification cc of C_t is “continuative” and $\exists \mathbf{m}_{\mathrm{mv}} \in C_{t+1}$.
B: Attributive Pattern	$x_{\mathrm{a_1}}$ The clause-classification cc of C_t is “attributive” and $\exists \mathbf{m}_{\mathrm{pw}}$ at the head of C_{t+1}.
C: Noun Pattern	$x_{\mathrm{n_1}}$ The clause-classification cc of C_t is “noun” and $\exists \mathbf{m}_{\mathrm{pw}} \in C_{t+1}$.
D: 2, 3 Transitions Pattern	$x_{\mathrm{c_2}}$: The clause-classification cc of C_t is “continuative” and cc of C_{t+1} is “continuative” and $\exists \mathbf{m}_{\mathrm{mv}} \in C_{t+2}$. $x_{\mathrm{c_3}}$: The clause-classification cc of C_t is “continuative” and cc of C_{t+1} is “continuative” and cc of C_{t+2} is “continuative” and $\exists \mathbf{m}_{\mathrm{mv}} \in C_{t+3}$. $x_{\mathrm{a_2}}$: The clause-classification cc of C_t is “continuative” and cc of C_{t+1} is “attributive” and $\exists \mathbf{m}_{\mathrm{pw}}$ at the head of C_{t+2}. $x_{\mathrm{a_3}}$: The clause-classification cc of C_t is “continuative” and cc of C_{t+1} is “attributive” and cc of C_{t+2} is “attributive” and $\exists \mathbf{m}_{\mathrm{pw}}$ at the head of C_{t+3}. $x_{\mathrm{n_2}}$: The clause-classification cc of C_t is “continuative” and cc of C_{t+1} is “noun” and $\exists \mathbf{m}_{\mathrm{pw}} \in C_{t+2}$. $x_{\mathrm{n_3}}$: The clause-classification cc of C_t is “continuative” and cc of C_{t+1} is “continuative” and cc of C_{t+2} is “noun” and $\exists \mathbf{m}_{\mathrm{pw}} \in C_{t+3}$.

- (d): Exclude indirect interrogative clause.
- (e): Exclude meaningless adverbial clauses which include “向け” (*muke*): “towards”, and clauses which are not judged as demands including “ため” (*tame*): “in order to”.

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]

.

[6]http://kotoba.nuee.nagoya-u.ac.jp/tsutsuji/

[7]We 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 V. EXPERIMENTAL RESULTS

Not demand	Precision	Recall	F-measure
i (Baseline)	96.4	99.1	97.7
ii	96.8	99.0	97.8
iii	97.0	99.0	98.0
Demand An Explanation	**Precision**	**Recall**	**F-measure**
i (Baseline)	93.0	88.6	90.6
ii	92.4	88.9	90.5
iii	92.4	88.6	90.4
Demand	**Precision**	**Recall**	**F-measure**
i (Baseline)	83.8	46.9	60.0
ii	81.9	53.7	**64.6**
iii	85.1	58.9	**69.3**

The result for "Demand" shows improvement with feature set iii. Compagreen 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

[1] Kazuhide Yamamoto and Yasuhiro Adachi. Summarizing spoken language used in the minutes of the national diet. *Journal of Natural Language Processing*, Vol. 12, No. 1, pp. 51–78, 2005. (in Japanese).

[2] Yasutomo Kimura and Satoshi Sekine. Annotation and definition of political issues focusing on headword. *Proceedings of The Eighteenth Annual Meeting of The Association for Natural Language Processing (NLP-2012) F1-6*, 2012. (in Japanese).

[3] Bakhtawar Seerat and Farouque Azam. Opinion mining: Issues and challenges (a survey). *International Journal of Computer Applications*, Vol. 49, No. 9, pp. 42–51, 2012.

[4] Daniel Diermeier, Jean-Francois Godbout, Bei Yu, and Stefan Kaufmann. Language and ideology in congress. *Annual Meeting of the Midwest Political Science Association*, 2007.

[5] Hiroshi Kanayama and Tetsuya Nasukawa. Textual demand analysis: Detection of users' wants and needs from opinions. *COLING '08 Proceedings of the 22nd International Conference on Computational Linguistics*, Vol. 1, pp. 409–416, 2008.

[6] Hiroko Otsuka, Masao Uchiyama, and Hitoshi Isahara. Request-intention judging criterion for open-ended questionnaires. *Journal of Natural Language Processing*, Vol. 11, No. 2, pp. 21–66, 2004.

[7] Takehiko Maruyama, Hideki Kashioka, Tadashi Kumano, and Hideki Tanaka. Rules for automatic clause boundary detection and their evaluation. *In Proceedings of the Nineth Annual Meeting of the Association for Natural Language proceeding*, pp. 517–520, 2003. (in Japanese).

[8] Moe Eguchi, Chitose Sao, Kentaro Inui, and Yuji Matsumoto. Extended modality analysis that integrates modality, actuality and valuation. *Proceedings of The Sixteenth Annual Meeting of The Association for Natural Language Processing (NLP-2010)*, Vol. 37, No. 2, pp. 852–855, 2010. (in Japanese).