Speech Act Identification using an Ontology-Based Partial Pattern Tree

Chung-Hsien Wu, Jui-Feng Yeh, and Ming-Jun Chen

Department of Computer Science and Information Engineering
National Cheng Kung University, Tainan, Taiwan, R.O.C.
{chwu, jfyeh, mjchen}@csie.ncku.edu.tw

Abstract
This paper presents an ontology-based partial pattern tree to identify the speech act in a spoken dialogue system. This study first extracts the key words/concepts in an application domain using latent semantic analysis (LSA). A partial pattern tree is used to deal with the ill-formed sentence problem in a spoken dialogue system. Concept expansion based on domain ontology is adopted to improve system performance. For performance evaluation, a medical dialogue system with multiple services, including registration information, clinic information and FAQ information, is implemented. Four performance measures were separately used for evaluation. The speech act identification rate achieves 86.2%. A Task Success Rate of 77% is obtained. The contextual appropriateness of the system response is 78.5%. Finally, the correct rate for FAQ retrieval is 82% with an improvement of 15% in comparison with the keyword-based vector space model. The results show the proposed ontology-based partial pattern tree is effective for dialogue management.

1. Introduction
Conversation with the machine using spoken language is a very tremendous vision of the computer technology. Dialogue management is arguably the core functionality of the spoken dialogue systems because the dialogue manager tends to have the controlling function of the system as a whole and maintains a model of the evolving dialogue context [1]. In these decades, several practicable dialogue systems, such as air travel information service system, weather forecast system, automatic banking system, automatic train timetable information system, and the Circuit-Fix-it shop system, have been developed. The dialogue control in these systems can be categorized into three approaches: finite state-based, frame-based and agent-based [2]. Among these approaches, there are two major issues about the dialogue research: the first is the robustness for speech recognition and the other is speech act identification. Since the speech act plays an important role in the development of dialogue management for dealing with complex applications, speech act identification [4] will be the most important topic with respect to the methods used to control the dialogue with the users.

Although the understanding of spontaneous spoken language is an open issue in natural language processing and artificial intelligence, in practice, the concept of understanding is situation-dependent. Dialogue systems usually collect the information from the users by filling the semantic frames or slots. There are many strategies to obtain the content desired in the semantic frames/slots recently. The plan-based approaches adapted to user’s interest gather the information items. The object-oriented approaches inherit the attribute from the ancestor classes. However, neither the plan-based nor the object-oriented approaches are both capable to work effectively. This is because that these approaches cannot identify precise speech act from the user’s utterances using only semantic frames or slots. In addition, more accurate identification is needed for multiple services in order to distinguish different speech acts of utterances with different word orders. This study proposes an approach to identify the speech act in a complex dialogue situation like multiple services in the same domain. In this study, first, the latent semantic analysis is used to obtain the key word set from the domain-dependent training corpus. Because of the variety of the spontaneous characteristics and the sparseness of the training data, the partial pattern tree (PPT) [3] is employed to model the ill-formed sentences from the training corpus. Based on the PPT, an utterance scoring algorithm is used to estimate the similarity between input utterance and the partial pattern sequence in the PPT. As a partial pattern sequence in the PPT with the highest score has been matched, the speech act of the user’s utterance can be determined.

On the other hand, the discourse information is generally stored in the semantic frame or slots for speech act identification. However, dialogue systems cannot obtain the exact semantic meaning due to the lack of the knowledge or conceptual representation in the input utterances. It is manifest that words are the semantic building block in the linguistic research, so the conceptual representation of a word can provide auxiliary aid for the comprehension of the input utterances. This study introduces the ontology which provides the conceptual space for domain application. Therefore, key concept expansion covering synonyms and relations defined in the ontology is used for semantic inference [5].

2. Ontology-based partial pattern tree
In speech act identification, word matching is not suitable for spontaneous speech due to the difference between the word orders in two spoken sentences. Consequently, in this study, a partial pattern tree (PPT) is used to partially match the words between the ill-formed input utterance and the speech act patterns in the PPT for robust speech act identification.

2.1. Key word set selection using latent semantic analysis
Latent semantic analysis is a novel approach for automated document indexing which is based on a latent class model for factor analysis of count data. There are two reasons for our approach using the latent semantic analysis. First, key words are the important information for speech act representation. Second, latent semantic analysis that projects the higher dimensions into the latent semantic space from the original space provides a method for dimensionality reduction. This study constructs a SpeechAct-by-Word matrix first. The
mapping is performed by decomposing the SpeechAct-by-Word matrix $A$ into the product of three matrices, $W$, $S$, and $SA$ using singular value decomposition.

$$A_{td} = W_{td} S_{nx} (SA_{dx})^\top \approx W_{td} S_{nx} (SA_{dx})^\top = \tilde{A}_{td}$$

where $n = \min (t, d)$. The matrices, $W$ and $SA$, have orthonormal columns. This means that column vectors have unit length and all vectors are orthogonal to each other. In fact, these vectors are the eigenvectors. The diagonal matrix $S$ contains the singular values of $A$ in a descending order. The $i$-th singular value indicates the amount of variation along the $i$-th axis. The LSA approximation of $A$ is computed by thresholding all but the largest $r$ singular value in $S$. Besides dimensionality reduction, the LSA keeps the key words while filters out the irrelevant words.

### 2.2. Speech act identification by ontology-based partial pattern tree

In general, each utterance can be represented as a sequence of functional phrases and a semantic word from the word key set as $S_s = \{FP_r, FP_r', \cdots, FP_r, SW, FP_{SW}, \cdots, FP_{SW, SW_r}\}$, where $SW_r$ denotes the semantic word and $FP_r$ denotes the $j$-th function word in the utterance $S_s$. $NB$ and $NA$ represent the number of the function words before and after the semantic word, respectively. Therefore, $2^{NB} \times NA$ partial pattern sequences containing the semantic word, $SW_r$, will be generated according to the definition of PPT. For example, if the training utterance is “ABC” with the semantic word “B,” four partial pattern sequences will be generated as: “ABC,” “AB,” “BC” and “B.” Fig. 1 shows the example for the partial pattern tree.

![Partial Pattern Tree](image)

Figure 1: Example for the partial pattern tree.

The PPT is basically an integrated tree structure of the partial pattern sequences generated from the training sentences. Each partial pattern sequence is tagged with one speech act. Each internal node representing a word in the partial pattern tree is denoted as $IN_i = \{PH_i, FR_i, N_i, SON_i\}$, where $PH_i$ is the word in this internal node. $FR_i$ is the frequency of the internal node. $N_i$ is the number of the descending internal nodes. $SON_i$ is the pointer linked to its son. Each external node represents one partial pattern sequence which corresponds to one speech act in the PPT. The data structure of the external node is defined as $EN_i = \{PP_r, Pr_r, P(SA^k)\}$, where $PP_r$ is the reference partial pattern sequence, $Pr_r$ is the pattern pointer set, $P(SA^k)$ is the probability of the $k$-th speech act with respect to the $i$-th partial pattern sequence. The algorithm for constructing the partial pattern tree is described as follows:

#### PPT Construction Algorithm

**Step 1:** Initialization: Create a root node, $R$

**Step 2:** Recursion: For each training partial pattern sequence $PP_i = \{Ph_1, Ph_1', \cdots, Ph_n\}$, where $N_i$ is the number of words in $PP_i$, compute Step 2.1 to Step 2.5

**Step 2.1:** According to the sequence of words in the partial pattern sequence, traverse the PPT and stop at $IN_i$, such that the path from the root node to node $IN_i$ matches the prefix of $PP_i$.

**Step 2.2:** If all the words in $PP_i$ have been used to traverse the PPT, then go to Step 2.5

**Step 2.3:** According to the suffix phrase $Ph_i$ of $PP_i$, a subtree, which links $Ph_i$, is created and the pointer of this subtree is added to the internal node $IN_i$.

**Step 2.3.1:** Create a new internal node, $IN_i$.

**Step 2.3.2:** $Ph_i$ is the word sequence of $IN_i$.

**Step 2.3.3:** Add $IN_i$ into the son’s linked list of $IN_i$.

**Step 2.4:** If the external node does not exist in the partial pattern tree, then create a new external node $NE_i$ and set the word sequence $NE_i$ to the partial pattern $PP_i$.

**Step 2.5:** Increase the frequency $FR_i$ of the internal node

Using the constructed PPT, speech act identification can be performed by matching the input utterance and the partial pattern sequence in the PPT. Dynamic programming is applied to deal with the problem of word order which is important for the user’s intension. If the input utterance is $P_i = \{a_1, a_2, \ldots, a_t\}$ and $PP_j = \{b_1, b_2, \ldots, b_j\}$ is the external node representing the $j$-th partial pattern sequence. Therefore, the similarity of these two sequences can be obtained using the Needleman-Wunsch algorithm [6] shown in the following steps:

**Initialization:** Create a matrix with $I + 1$ columns and $J + 1$ rows. The first row and first column of the matrix can be initially filled with 0. That is,

$$Sim(i, 0) = 0, \text{ if } i = 0 \text{ or } j = 0$$

**Score pathways through array:** Assign the values to the remnant elements in the matrix as the following:

$$Sim(i, j) = \max \left\{ \begin{array}{l} Sim(i - 1, j - 1) + Sim_{text}(a_{i-1}, b_{j-1}), \\ Sim(i - 1, j) + Sim_{text}(a_{i-1}, b_j), \\ Sim(i, j - 1) + Sim_{text}(a_i, b_{j-1}) \end{array} \right.$$ (3)

**Construct alignment:** Determine the actual alignment with the maximum score $Sim(P_i, PP_j)$.
It is difficult to obtain the exact matching between the input utterance and the partial pattern sequence due to the versatility of the spoken language, especially in the word sense. To solve this problem, ontology is employed for similarity measure. Two basic relations: hypernym and synonym are introduced. The similarity is defined as follows:

\[
\text{Sim}_{\text{onto}}(a_i, b_j) = \begin{cases} 
1 & \text{if } a_i = b_j \\
\frac{1}{2} & \text{if } a_i \text{ and } b_j \text{ are hypernyms} \\
1 - \left(\frac{1}{2}\right)^n & \text{if } a_i \text{ and } b_j \text{ are synonyms} \\
0 & \text{otherwise}
\end{cases}
\]

where \( l \) is the number of levels between \( a_i \) and \( b_j \). The variable \( n \) is the number of their common synonyms in the synonym set. Finally, the speech act for the input utterance \( P_i \) is determined according to the following equation.

\[
S(A) = \arg \max_{k,j} \left( P(SA^k) \times \text{Sim}(P_i, PP_j) \right)
\]

where \( P(SA^k) \) is the probability of the \( k \)-th speech act with respect to the \( j \)-th partial pattern sequence estimated in the construction of PPT.

3. Experiments

To evaluate the proposed ontology-based PPT, a dialogue system for medical application was developed as shown in Figure 2. There are three services provided in the medical domain. The registration information service helps patients register and consult the doctor’s information. The clinic information service is a decision support oriented application system. It provides the suggestion about the department of clinic care according to the input utterance. For this service we built a medical knowledge base with 1222 axioms. The content of the axioms is formatted as shown in Figure 3. The FAQ information service is a health care education system. It provides suitable medical documents inquired by the users. The speech acts in these three services are very diverse and manually classified into 12 speech acts as shown in Figure 4.

3.1. Development of spontaneous corpus

There are 3 corpora collected for experiments: telephone corpus, wizard of Oz (WOZ) corpus and testing corpus. In total, 4098 turns in 364 dialogues were recorded via the telephone line from National Cheng-Kung University Hospital, Taiwan. There are 11.2 turns in each dialogue on average. 234 turns in 34 dialogues were collected via WOZ and 502 turns in 56 dialogues were collected as the testing corpus.
distributions of the number of turns in a dialogue are shown in Figure 5.

![Graph showing distributions of number of turns in a dialogue](image)

Figure 5: Distributions of the number of turns in a dialogue

### 3.2. Dialogue control evaluation

To evaluate the dialogue flow control using the proposed method, fifty users who did not participate in this research project were asked to test the system. There are 4 measures used for evaluation [7][8]: Task Success Rate, Average Number of Turns, Contextual Appropriateness, and Intention detection Rate. The approach using semantic frames were implemented for comparison and the results are listed in Table 1. According to the results, we can see the variability in the average number of turns. The number of turns in the FAQ information service is shorter than the multiple information services. The FAQ service achieves the best score in the speech act identification rate due to the clear behavior compared with others services. However, we can find there are some confusion between the registration information service and the clinic information service in speech act identification. The task success rate of FAQ service is higher than that of other services. However, dramatic decrease happens in the integrated system due to the poor performance of speech act identification. Speech act identification is strongly related to contextual appropriate response. In addition, Table 2 shows that the ontology-based partial pattern tree can effectively improve the speech act identification rate and task success rate especially in multiple services in a dialogue system.

<table>
<thead>
<tr>
<th>Evaluation Measures</th>
<th>Speech act Identification Rate (%)</th>
<th>Task Success Rate (%)</th>
<th>Average Number of Turns</th>
<th>Contextual Appropriateness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration Module</td>
<td>87.3%</td>
<td>92%</td>
<td>8.40</td>
<td>82.0%</td>
</tr>
<tr>
<td>Clinic Query Module</td>
<td>84.6%</td>
<td>80%</td>
<td>9.27</td>
<td>75.1%</td>
</tr>
<tr>
<td>FAQ Module</td>
<td>92.4%</td>
<td>88%</td>
<td>4.80</td>
<td>85.2%</td>
</tr>
<tr>
<td>Integrated System</td>
<td>80.6%</td>
<td>68%</td>
<td>9.80</td>
<td>74.3%</td>
</tr>
</tbody>
</table>

### 4. Conclusions

This paper presents an approach for speech act identification using latent semantic analysis and ontology-based partial pattern tree. This approach was applied to deal with the complex dialogue application: the system with multiple services in the medical domain. According to the experiments, we can find the proposed approach outperforms the traditional approach using the semantic frames/slots especially in speech act identification rate and task success rate. The results show the ontology-based partial pattern tree is superior to the traditional dialogue management.

### References