Abstract: This paper proposes a new method for analyzing conversation model recorded during the PBL instruction. The theory to analyze the conversation log is based on Speech Act Theory. According to this theory, each sentence could be categories into different intentional behavior. Using the categories sentences, this research cut the entire log into segments and determines which type the segments belongs to. A conversation log in a group discussion is analyzed by the algorithm designed in this paper and is transformed into Chart for teacher to understand the characteristics of this group in the end of this research.

1. Introduction

Problem Based Learning (PBL) (Barrows, 1996) is a teaching strategy for making students learning by practicing real problem. One of the characteristics in PBL is that teacher usually grouping students into several small teams. Each team should discuss with each other to find a workable solution and solve their problem by themselves. In this situation, teacher should observe the discussion progress in each team and give them advices when the team has difficulty to solve the problem or discuss the issue in the wrong direction.

Recently, some researchers tried to put the PBL instruction into web and use various Internet applications to support some learning activities of PBL. Chat-room is a type of synchronous communication tool for supporting on-line discussion during the instruction of PBL. The advantages of using chat-room for replacing face-to-face discussion is that every word the team discussed could be easily recorded in computers. Teachers could read those records and analyze the characteristics of students after the discussion.

This paper tries to propose a computational method to analyze the discussion record in chat-room and finds out some quantitative data for teachers. Section 2 introduces some web-based PBL instruction system. Speech Act Theory also mentioned in section 2. They way to use speech act theory for modeling conversation is analyzed in section 3. Section 4 shows discussion-analysis algorithm based on the analysis in section 3. An example would be demonstrated in section 5 and section 6 would give a brief conclusion.

2. Web-Based PBL Instruction and Speech Act Theory

PBL uses real life problems to make learners know what they learned and how to use. More and more reseachers construct suitable web environment for supporting PBL instruction. Problem Based Internet Assisted Learning System (PBIALS) is one example of web-based PBL learning system. (Chen, Kuo, Chang and Yang, 2003) Students could use the system to read teaching materials, discuss problems, construct brain-storm-maps, upload homework and produce their whiteboard.
Another web-based learning software tool, CoMMIT, is used by Glenn for realizing web-based PBL instruction. (Lautenbacher, Campbell, Sorious and Mahling, 1997) There are three major modules in CoMMIT. Students could use the Authoring Module to represent the relationships between educational materials by graphical-drawing tool. They could also use Student Module for sharing course contents by delivery engine. The teacher could use the Instructor Support Module to monitor students’ behavior and evaluate each team according to the recorded data of students.

Most of these web-based PBL instruction system have synchronous (like chat-room) or asynchronous (like discussion board) for students to discuss how they could solve the problem. All these discussions can be considered as conversations among students. To analyze conversation, most researches mentioned about Speech Act Theory. (Austin, 1962; Searle, 1969) The Speech Act Theory separates a conversation into two levels, locutionary level and illocutionary level. Locutionary level only means the meaning utterance itself, but illocutionary level focuses on the conventional force in the conversation.

Searle classified those illocutionary behaviors into five types. Some other researchers also categorized illocutionary behaviors in different ways. Leech offered four classes of illocutionary functions with social goal. (Leech, 1983) Another researcher, Bales, defined four categories of interaction among groups, including Question, Information, Positive, and Negative. (Hare, Borgatta and Bales, 1965) In Conversational Role Model (COR Model), the researchers also defined several illocutionary interaction behaviors focused on how to present the modules in final state machine. (Hoppe, King, Schiele and Tiilen, 1989) This paper would use these researches to analyze the conversation module in PBL instruction.

3. Conversation Segments

Based on the classical Discourse Theory, a conversation could be separated into three phases, which are opening, body, and closing. (Piwek, P. and K. van Deemter, 2003) A conversation may be opened with greeting or pleasantry and closed by saying goodbye. The subject of this conversation would usually be presented in the body phrase. In this research, the definition of opening is the beginning of discussion. In general, the beginning of conversation usually exchange greet with small talk, such as everyone says hello to other participants... etc. The meaning of body in this research represents the process of discussion. The participants will discuss some problems and maybe generate conclusions that are all participants identified of some topic. For example, participants exchange views about topic ... etc. Closing in this paper represents the end of discussion. For example, students say goodbye to other participants... etc

For analyzing how tutors and learners interact in the conversation with PBILS, this research tries to classify types of the conversation segments. Before clustering those sentences into different segment types, what dimension of segment types exists in the discussion of PBL lesson is the first topic this paper focuses about. In Schegloff’s research, two people’s conversation sequence can be divided into several adjacency pairs. (Schegloff, 1968) This research tries to expand the architecture of adjacency pairs for more than two speakers and focus on the conversation in PBL instruction.

To cut the entire conversation into smaller slices, the discussion sentences can be divided into segments, which is similar to adjacency pairs. Each segment has three dimensions, including:

1. **Participant Role**: In a conversation, the participants may have difference roles, like teacher and student.
2. **Participant Number**: Schegloff’s research only focused on two speakers chat. But, most discussions in the classroom have more than two speakers. The number of speaker, participant number, becomes another dimension for determining types of segment. In this research, the participant number dimension can be separated into three types, including “one participant”, “two participants”, and “more than two participants”.
3. **Information Seeker**: During the discussion of PBL instruction, students may ask teacher questions which they don’t understand; teachers may also propose some questions for guiding students the correct way of solving problems. For this condition, each segment of conversation may have different information seeker. Because the participant roles of this research only have teacher and student, the information seeker would also have these two types.

According to these three features of conversation segments, six segment types could be classified during the discussion in Table 1. The first column is the notation of each type related to the second column, Segment Type. The last three columns are the three features of conversation segment, including number of participants, positions of participants, and information seeker.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Segment Type</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Numbers of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>participants</td>
</tr>
<tr>
<td>S1</td>
<td>Student-Soliloquy</td>
<td>1</td>
</tr>
<tr>
<td>S2</td>
<td>Teacher-Soliloquy</td>
<td>1</td>
</tr>
<tr>
<td>S3</td>
<td>Discussion</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>More than 2</td>
</tr>
<tr>
<td>S4</td>
<td>Guidance to single student</td>
<td>2</td>
</tr>
<tr>
<td>S5</td>
<td>Guidance to multiple students</td>
<td>More than 2</td>
</tr>
</tbody>
</table>

Table 1. Segment Types and its Features
The first two segment types, *Student-Soliloquy* and *Teacher-Soliloquy*, have only one participant. The difference between these two types is the position of the participant which is also the information seeker in the discussion segment. Other types have more than one participant. The *Discussion* segment type means that the participants are all students in the discussion segment, including the information seeker. If the teacher is also in the discussion segment, he/she could ask some question to guide students to the correct to solve the problem. The types $S_1$ and $S_2$ are discussion segments which teacher uses question in the beginning to lead student to consider the correct direction to solve the problem. If students are confused for some concepts, they could become the information seeker to ask teacher for help, which is the *Ask-for-help* segment type in Table 1.

The most difficult part of determining dimension value mentioned above is to identify who is the *Information Seeker* in this segment. The way to solve this problem is to identify each sentence of the conversation with speaker’s behavior. In the previous session, some researchers analyzed different interactions in the conversation. This paper defines four types of *Participants’ Behavior* based on Balesian Categories and COR Model which is listed in Table 2.

**Table 2.** The corresponding of *Participants’ Behavior* to Balesian Categories and COR Model

<table>
<thead>
<tr>
<th>Participants’ Behavior</th>
<th>Balesian Categories</th>
<th>COR Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>Question</td>
<td>Request</td>
</tr>
<tr>
<td>Information</td>
<td>Information</td>
<td>Promise, Accept, Offer, Assert, Reject, Withdraw, Inform</td>
</tr>
<tr>
<td>Evaluation</td>
<td>Positive</td>
<td>Evaluate</td>
</tr>
<tr>
<td>Chat</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

The first type of *Participants’ Behavior* is *Question*. This behavior comes from the Question Category of Bales Interaction and also be similar to the action of Request in COR Model. The main idea of *Question* type is to bring up a problem in the discussion. The second type of *Participants’ Behavior*, *Information*, also comes from the Information Category of Bales Interaction. In this behavior, the speaker would reply some opinions to deal the problem in the discussion. This behavior could also response to most of actions in the COR Model.

The next type of *Participants’ Behavior*, *Evaluation*, is to judge the information mentioned in the discussion and can be easily correspond to the Evaluate action of COR Model. This evaluation can be identified into two types, which are *Positive* and *Negative* categories of Balesian Interaction. The last type is *Chat*. This type of behavior does not belong to any categories of Balesian Interaction or COR Model but usually exists in the conversation or discussion, such as some greeting, gossip or chat which are not related to the problem proposed in the discussion. For this *Participants’ Behavior*, this research also defines a new segment type, *Chat* (and denoted as $S_0$), in Table 2.

During the discussion, question and answer always exists in the conversation. To simplify the analysis, each segment has only one information seeker. For this reason, a segment begins with a question, or the *Question* behavior mentioned in Table 2. The following conversation classified into *Information* and *Evaluation* behavior would be the same segment of the problem. If the conversation is cut in with a *Chat* or *Question* behavior, a new segment would be created.

To identify what participants’ behavior the sentence is, some keywords should be defined first. A *Question* behavior should have an interrogative sentence which consists with interrogative words. Another behavior, *Evaluation*, has two conditions, presenting positive and negative opinions. Based on this definition, a sentence which belongs to *Evaluation behavior* should have affirmative or negative words. The way to identify *Information* and *Chat* behavior is if there’s any related keyword exists in the sentence. If a sentence consists with keywords related with the discussion problem, the behavior of this sentence is *Information*. Other conditions would be considered as *Chat* behavior.

4. **Algorithm**

Based on the previous analysis, Figure 1 shows an algorithm of discussion analysis. To understand the entire algorithm, Table 3 lists all the definitions of needed variables. *SemanticsWords*, *PragmanticsWords_Question*, and *PragmanticsWords_Evaluation* are three predefined variable. Each of them are used for identifying *information*, *question*, and *evaluation* behaviors in Table 1. The whole algorithm begins with variable *DiscussionLog*, which records the conversation and speakers.

**Table 3.** Lists of variables definition used in discussion-analysis algorithm in Figure 1.
There are two major procedures in the algorithm. The first procedure is to identify each sentence into its Participants’ Behavior. This procedure only focuses on the conversation part in the DiscussionLog. For this reason, in the beginning of the algorithm separates the DiscussionLog into two part, ConversationVector and UserVector. After that, the algorithm analyze each sentence in the ConversationVector to determine which Participants’ Behavior it is. The rule to categories these sentences based on the three predefined variable mentioned in the beginning of this section. The analyzed data would be stored in the variable ConversationTypeVector.

The other procedure in the algorithm is to separate the entire discussion log into segments and identify which type the segment belongs to. In this procedure, Question and Chat behavior would be determined as the beginning of a segment. When a new segment begins, the procedure would also record the speaker of the beginning sentence as the information seeker in this segment and record this information into SegmentInformationSeeker variable. Each sentence belongs to the segment would be recorded in to DiscussionSegment variable and the speakers involve in this segment of conversation would be also stored in the variable SegmentParticipantList. The final step of this procedure is to determine the type of the segments. The rules in this step based on the analysis in Table 1, which is related to the variable SegmentInformationSeeker and the number of elements in SegmentParticipantList.

Input: DiscussionLog, SemanticsWords, PragmaticsWords_Question, PragmaticsWords_Evaluation
Output: DiscussionSegments, SegmentTypes

1. Build ConversationVector
2. Build UserVector
3. For each sentence i in ConversationVector
   4. If the sentence i has any words in PragmaticsWords_Question
      ConversationTypeVector[i] = “Question”
   5. Else if the sentence i has any words in PragmaticsWords_Evaluation
      ConversationTypeVector[i] = “Evaluation”
   6. Else if the sentence i has any words in SemanticsWords
      ConversationTypeVector[i] = “Information”
   7. Else
      ConversationTypeVector[i] = “Chat”
4. For each sentence i in ConversationTypeVector
   5. If ConversationTypeVector[i] is “Chat”
      Create a new DiscussionSegments[j]
   6. SegmentType[j] = “Chat” // S0 in Table 1
16. Else if ConversationTypeVector[i] is “Question”
17. Create a new DiscussionSegments[j]
19. Create a new SegmentSpeakersList[j]
20. SegmentSpeakersList[j] = UserVector[i]
21. Else
22. If UserVector[i] is not exist in SegmentSpeakersList[j]
23. Insert a new element in SegmentSpeakersList[j]
24. For each segment j in DiscussionSegment
25. If SegmentInformationSeeker[j] is “Teacher”
26. If the number of elements in SegmentSpeakerList[j] is 1
27. SegmentType[j] = “Teacher-Soliloquy” // S2 in Table 1
28. Else if the number of elements in SegmentSpeakerList[j] is 2
29. SegmentType[j] = “Guidance to single student” // S4 in Table 1
30. Else
31. SegmentType[j] = “Guidance to multiple students” // S5 in Table 1
32. Else if SegmentInformationSeeker[j] is “Student”
33. If “Teacher” exists in SegmentSpeakerList[j]
34. SegmentType[j] = “Ask-for-help” // S6 in Table 1
35. Else
36. If the number of elements in SegmentSpeakerList[j] is 1
37. SegmentType[j] = “Student-Soliloquy” // S1 in Table 1
38. Else
39. SegmentType[j] = “Discussion” // S3 in Table 1

Figure 1. The algorithm of discussion analysis

Figure 2 shows a small example of the discussion-analysis algorithm. There are four input variables, including DiscussionLog, SemanticWords, PragmaticsWords_Question, and Pragmaticswords_Evaluation. In step 1 and 2, ConversationVector and UserVector are created in the beginning. After that, step 3~11 classified each sentences into its Participants’ Behavior. When the ConversationTypeVector is build, step 12~23 would determine how to divide the entire conversation into segments and store the result in DiscussionSegment. These steps also record the information of information seeker and speakers in each segment. The last few steps, step 24~39, would determine which type the segment is based on the information of SegmentInformationSeeker and SegmentSpeakerList.

Input: DiscussionLog = “Jerry:  What is the instruction SELECT?
Jason:  Maybe the instruction be used to select data from database.
Teacher:  hmmm...right!
Jerry:  What is the instruction CREATE?”
SemanticWords = “Select, Database, Create”
PragmaticsWords_Question = “What, ?”
PragmaticsWords_Evaluation = “hmmm, Maybe”
1 ConversationVector = “  What is the instruction SELECT?
Maybe the instruction be used to select data from database.
hmmm...right!
What is the instruction CREATE?”
2 UserVector = “Jerry, Jason, Teacher, Jerry”
3~11 ConversationTypeVector = “Question, Evaluation, Evaluation, Question”
12~23 DiscussionSegment[1] = “  What is the instruction SELECT?
Maybe the instruction be used to select data from database.
hmmm...right!”
SegmentSpeakersList[1] = “Jerry, Jason, Teacher”
The result in Figure 2 is

\[ \text{DiscussionSegment}[1] = \text{“What is the instruction SELECT?”} \]

Maybe the instruction be used to select data from database.

\[ \text{hmmm...right!”} \]

\[ \text{SegmentType}[1] = \text{“Ask-for-Help”} \]

\[ \text{DiscussionSegment}[2] = \text{“What is the instruction CREATE?”} \]

\[ \text{SegmentType}[2] = \text{“Student-Soliloquy”} \]

5. Demonstration

After constructing the discussion-analysis algorithm, this paper collects the chat-room conversation records from Chen’s research. (Chen 2004) The students were separates into several groups and discussed how to fabricate an ideal computer. This section would analyze one group’s conversation as an example of the previous discussion. Figure 3 is a snapshot of the eighth group’s discussion log. Column A shows the record index of the discussion log for confirming the correct sequence of the conversation. The second column records what group this conversation belongs to. After getting the eighth group’s discussion log, the DiscussionLog in algorithm comes from Column C and Column D, which are also UserVector and ConversationVector in the algorithm.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.220</td>
<td>第八组</td>
<td>何家亮</td>
<td>都可以</td>
</tr>
<tr>
<td>10.220</td>
<td>第八组</td>
<td>何家亮</td>
<td>不好顯示卡啊</td>
</tr>
<tr>
<td>10.222</td>
<td>第八组</td>
<td>杜政業</td>
<td>我分tCPU</td>
</tr>
<tr>
<td>10.228</td>
<td>第八组</td>
<td>杜政業</td>
<td>你們大家意見如何</td>
</tr>
<tr>
<td>10.229</td>
<td>第八组</td>
<td>陳冠良</td>
<td>用什麼相機再拍</td>
</tr>
<tr>
<td>10.231</td>
<td>第八组</td>
<td>杜政業</td>
<td>呵呵</td>
</tr>
<tr>
<td>10.233</td>
<td>第八组</td>
<td>杜政業</td>
<td>你會修卡副嗎</td>
</tr>
<tr>
<td>10.240</td>
<td>第八组</td>
<td>陳冠良</td>
<td>數理數理,ok</td>
</tr>
<tr>
<td>10.240</td>
<td>第八组</td>
<td>陳冠良</td>
<td>哈哈</td>
</tr>
<tr>
<td>10.244</td>
<td>第八组</td>
<td>呂夢琪</td>
<td>我記憶體機90</td>
</tr>
<tr>
<td>10.248</td>
<td>第八组</td>
<td>杜政業</td>
<td>老師說CPU理</td>
</tr>
<tr>
<td>10.250</td>
<td>第八组</td>
<td>teacher</td>
<td>數理數理</td>
</tr>
<tr>
<td>10.255</td>
<td>第八组</td>
<td>teacher</td>
<td>讓到討論區發表嗎</td>
</tr>
<tr>
<td>10.258</td>
<td>第八组</td>
<td>何家亮</td>
<td>哈哈</td>
</tr>
<tr>
<td>10.266</td>
<td>第八组</td>
<td>何家亮</td>
<td>哈哈</td>
</tr>
</tbody>
</table>

**Figure 3. Discussion Log**

The next step after getting the DiscussionLog from the original data, the discussion-analysis algorithm could operates the DiscussionLog and divide the log file into segments as shown in Figure 4. The sentences in fourth line and the last line in Figure 4 also show that the algorithm figures out which type the segment belongs to.
This conversation analysis report could be rearranged into graph or chart for teachers and make teachers to observe the characteristics of each group. Figure 5 calculates the amount of each segment type in the eighth group. This chart shows that the character of teacher is to guide students for solve the problem in the correct way. Sometimes, teacher should answer the question if students get confused. There is only one time that the teacher speaks by himself. The times of student discussion is equal to the times that teacher guide single and multiple student. It shows that this group could solve the problem by discussion but still need guidance from teacher.

![Figure 5. Simple statistics of segment types from the eighth group](image)

Figure 6 is another example in the third group. Most of the segments type is “guidance to single student” and no “Discussion” segment type exists. Teachers should encourage students finding solution by them self and discussing with teammates more often.

![Figure 6. Simple statistics of segment types from the third group](image)

6. Conclusion

This paper analyzed conversation log based on the Speech Act Theory for on-line PBL instruction. The first procedure for analyzing conversation log is to identify the Participants’ Behavior of each sentence. After knowing the behaviors hidden in the sentences, the entire log file would be divided into several segments. The types of segments can be categories according to the number of participants, positions of participants, and the information seeker in this segment. This research also proposed a basic algorithm for analyzing discussion log based on the previous analysis.

In the end of paper, the algorithm processed a conversation log in one group and showed a simple statistics for the segment
types. To discuss the transition among each type of segments could be another interesting issue for the future work. Observing when the teacher should interfere with students’ discussion could be another further research in this paper.

Reference


