Text Modeling in Adaptive Educational Chat Room

Muna Hanini	Nidal A.M Jabari	Radwan Tahboub
Palestine Polytechnic University, Informatics Palestine	Palestine Technical Colleges-Arroub, Computer Department Palestine	Palestine Polytechnic University, CE/CS Department Palestine

ABSTRACT

This paper discussed a model for monitoring the text and expressions used by the student during the educational chat session with his colleagues about a particular topic assigned to be discussed during a particular session. Monitoring is done through text modeling process by evaluating the expressions used by the student within the session time. This method is used to improve the general idea of the educational Chat among students and make it formalized which will lead to make the students behave more seriously during the session chat. Experiments performed as well as previous studies have produced mathematical equations based on the parameters extracted by analyzing and mining the messages submitted by the peers. Text model will be added to the time and peer model to construct the final student model in the educational chat room.

Keywords

 Model, Chat, AVCM, Student, Time, Text.

1. INTRODUCTION

Student modeling is a methodology to extract the student characteristics and evaluation process within the virtual classroom [1]. In [2] the virtual classroom model based on student modeling and course sequencing was discussed. This model offers online presentation, online testing and online peer to peer chat .In [3] also, it was discussed the student modeling in adaptive educational chat room by itself and in [4] it was discussed the time modeling in adaptive chat room.

This paper will discuss the text modeling in the educational adaptive chat room as one component of the student modeling in educational chat room which consists of text modeling, time modeling and peer evaluation modeling as described in [3]. Text model means extracting the evaluation criteria which we may use to evaluate the text and expressions that the student used while chatting with his peer in the virtual classroom.

When we succeed to formulate a mathematical equation for student model in adaptive chat room, we will add a new component to the student modeling within the chat room besides the last model which was related to the time investment during the session by the student [4].All these models will be integrated with the AVCM student modeling components [2].

2. BACKGROUND AND THEORITICAL FRAMEWORK

2.1 AVCM

 AVCM [2] is a virtual classroom model which applies adaptive e-learning and consists of three main facilities:

1. Adaptive presentation

2. Adaptive testing

3. Adaptive chat.

Several models underlined these models to perform their functions required by AVCM as described in [2]:

1. Domain Model (DM).

2. Student Model (SM).

3. Activity Model (AM).

4. Resource Model (RM).

5. Nodes Selector.

6. Concept Score Evaluator.

7. Cognitive Style Evaluator.

8. Chat Room Interface Adapter.

9. Peer Evaluation.

AVCM student modeling process relies mainly on two methods:

- Static Student Modeling (Stereotype).
- Dynamic modeling (Overlay Model).

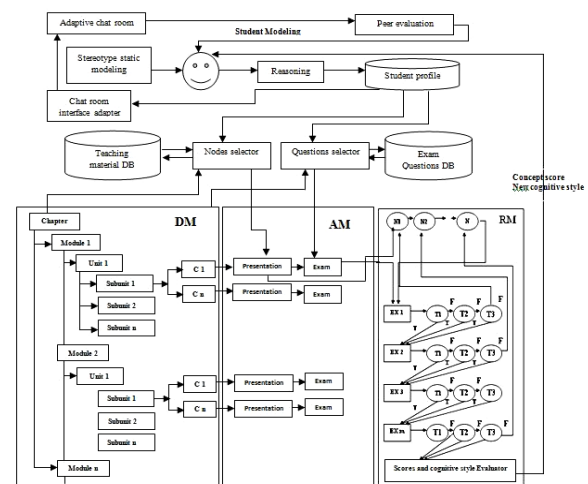AVCM overall architecture is shown in figure 1.

Fig 1: AVCM Overall Architecture

2.2 AVCM Chat Model

Chat is a peer-to-peer service for communication with others [5]. Chat room in AVCM is an additional separated system in AVCM [6] which can get advantage from Student modeling in AVCM. Adaptive chat room is a collaborative tool for methodologies can be used to model the students chat when they want to discuss among each others. These tools can give evidence about the student knowledge level for each student participating in chat room relying on student concept evaluation [2].The most suitable methodology in our case

Chat takes an input from both student and domain model, and adapts the chat room according to the input score. The student who has less knowledge than the current student will appear in red color. The student who has equal knowledge to the current student knowledge will be given a yellow color. Finally, the student who gets higher knowledge will have green color. This way makes it easier for the current student to determine which student he should discuss with the current concept [2].

In [3, 4] the student modeling within this adaptive educational chat room is discussed .The process relies on four models:

1. Time model
2. Text model
3. Peer model
4. Interface model.

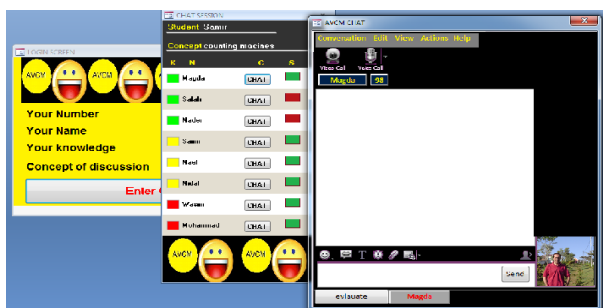AVCM chat interface is shown in figure 2.

Fig 2: AVCM Chat Interface

These models are responsible to extract the parameters which are necessary to model and evaluate the student during the chat session. Mathematical equation was extracted for time model in [4] as the following:

$$\text{Time model} = 46.813 + 0.574 * (\text{posts count} / \text{Complete session time} + \text{TIDL}) \ldots\ldots (2.2.1)$$

Where "posts count" means the total count of messages the student submitted during the chat session," complete session time" is predefined by the supervisor while TIDL the total IDL time during the session.

2.3 Text Mining and Analysis

Data mining, in general, is the process to extract patterns from a large amount of data. The process relies on AI methods to extract the information and change it to manageable and useful to use. It uses database management systems, inference, visualization and other computer science concepts to perform its tasks [7]. In peer to peer chat process, both parties write number of messages. These messages are usually related to one or more concept. In this case, the research is looking for methodology to analyze the text submitted in the chat room to evaluate its relation with the current concept of discussion. Chat posts is usually classified as being a part of conversation, or free topic chat being about a particular topic [8], collaborative-task oriented, academic seminar or presentation chat, practice chat and evaluation chat. All these are collaborative task-oriented chat and evaluation chat. We are interested in this type because we are looking for estimating the degree of the relation of the chat posts, and the current educational concept. In chat posts, we face many chat characteristics which should be taken in consideration during this study [9]:

1. Chat text and its component.
2. Chat initializes, which are abbreviations and acronyms used in chat, are used to express some words instead of spoken language.
3. Emoticon usage: Some symbols used to express emotions.
4. Abbreviated speech: Misused words are more common in chat room.
5. Mentions: Sounds like buzz and poke.
6. using of different languages.

Currently, a number of monitoring systems are available, mainly divided into two categories network based and client based.

Hui, He and Dong in [10], proposed chat monitoring system called IM analysis. It uses text mining techniques for chat text analysis. Procedures introduced on chat are for retrieval of general browsing. It discovers the social interactions of IM users with their contacts, in addition to topic analysis for the current chat.

Figure 3 shows the general methodology used for data mining and analysis of text mining solutions.
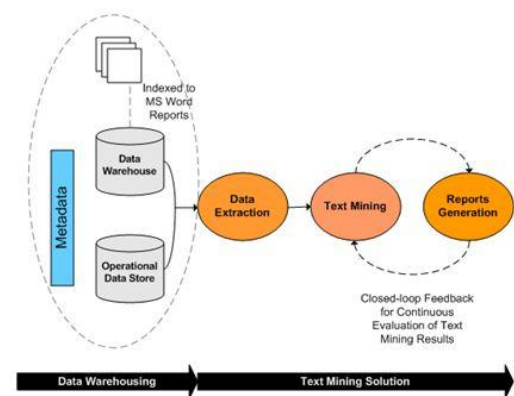

Fig 3: Text Mining Methodology

Typical text mining as tasks describe in [10] includes:

1. Text categorization.
2. Text clustering.
3. Concept entity extraction.
4. Production of granular taxonomies,.
5. Sentiment analysis.
6. Document summarization.
7. Entity relation modeling (i.e., learning relations between named entities).

2.4 Regression Analysis and Statistics

This section describes the usage techniques and benefits of using regression analysis in statistical procedures. In this research regression analysis is used widely to extract the necessary parameters.

The role of using statistics especially in science and engineering is an important factor in finding results [11]. [11] Described the role of statistics in experiments as described in figure 4.
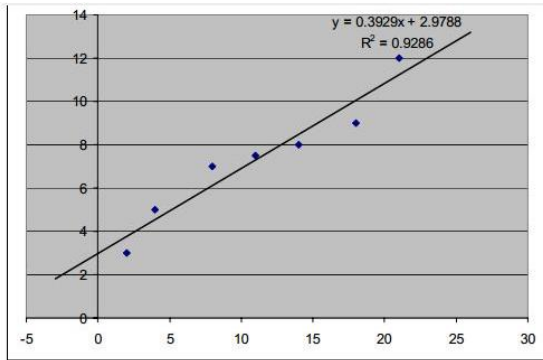
Fig 4: Regression Analysis Concept

3. TEXT MODEL IN ADAPTIVE EDUCATIONAL CHAT ROOM

It is needed to find solution for the text evaluation of the chat session conversation between two students. Consequently, it is need to design an algorithm to solve this problem. Chat conversation is unstructured text. It is very difficult to analyze if it does not converted to structured text. Structured information is easier to search, manage, organize, share and create reports for computers as well as people. This allows computers and people to better manage the information, and allows known techniques and methods to be used. Text mining analysis of the chat text gives minor number of expressions. These expressions are compared to the keywords which the field expert gave and calculate the degree of matching between them.

The text in the chat conversation usually has many expressions; some of them can be used for the text modeling, while the others are not used at all. Even the used expressions have different distance from the concept of the discussion, which is the core of the research modeling process. These expressions are extracted from the experiment that conducted in PTCA on 18 students twice using two different concepts. The results are extracted using regression analysis discussed before and described in the following paragraph.

First experiment held in PTCA on 18 students, 4 out of them failed to apply the experiment efficiently, while the others applied it successfully discussing "relations and keys" in relational database system. From the experiment, we found that the expressions used in the chat conversation are either used or unused. We divided the used expressions into eight clusters as seen in table 1, while table 2 shows the levels description.

Table 1. Text Clustering in Chat.

Category	Level	Target tag
Main concept	Level 1	Self
Related directly to the main concept	Level 2	Self
Useful words	Level 3	Self
Positive expressions	Level 4	Both
Agreement expressions	Level 5	Both
Enquiry expressions	Level 6	Both
Respect expressions	Level 7	Self
Negative expressions	Level 8	Peer
Unused expressions	Level 9	Both

Table 2. Levels Description

Cluster	Content (database)
Level 1	Relation , keys , degree
Level 2	Table, one, many, primary, foreign, composite, database. information ,field , record , operation , process , insert , query , redundancy , merge , map , related , attribute , unique , redundancy , number , text , element , system , binary , empty , Null , triple , circular , access , SQL , compound . (and all other derivatives).
Level 3	Example, instance, suppose, declaration, describe, allowed, specified, See, Look, Refer, Moreover, However, Anyhow, Despite, In spite, yet, classify. (and all other derivatives).
Level 4	Clear, Ok, good, excellent, understand, Thanks, yes, amnesty, correct, true, of course.
Level 5	Not clear, Bad, I don't know, wrong.
Level 6	That is right, You are right, I think so too, agree with you, Exactly, Right, well, absolutely true.
Level 7	Can , How , what , when , where , who , which , Ask you , Give me , Describe me ,
Level 8	Hello, high, how are you, good morning, good day, good afternoon, good evening, have a nice time, gentle.
Level 9	All other words.

As shown in table 2, the text analysis in chat depends on two main sections:

- Main concept and related to the concept.
- Other expressions.

The main and related section is filled by the expert while the other expressions are fixed and can be expanded once it is needed to add new expressions.

Level 1 to level 8 in addition to the total words used in the chat session are extracted using database system through matching keywords in the database with the chat expressions and count the number of matching expressions for each level. The system should be database oriented and component based system. Figure 5 describes the main component of the model.
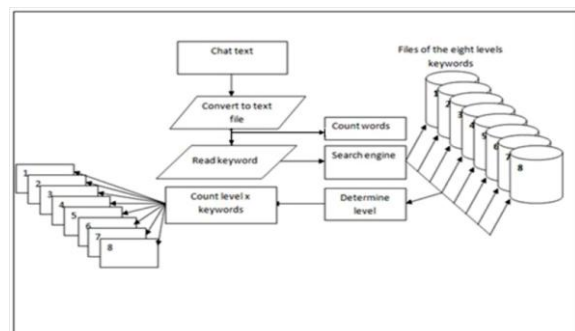

Figure 5: Proposed Architecture of the System.

Applying the experiment on the model described in figure 5 means that all students' chat text should be converted from the chat window into text file for easier manipulation. The words in the text file should initially be counted and stored in variables, and they should be read sequentially, one word by one word. These words are considered as keywords. The

search engine will look for the keyword and its derivatives in the database files designed for this propose. If found it means it is a used keyword, otherwise it will be considered as not used keyword. Finally the system will have a count for each parameter or level which can be used for text modeling in the chat. Table 3 shows a sample of text analysis from the first experiment.

Table 3. Sample from the Text Analysis (Experiment 1)

Number	Total Words	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Used words	Unused words
Adaptive 1	191	12	36	13	8	1	0	8	1	79	112
Adaptive 11	193	17	48	6	7	1	4	5	1	89	104

From the last table we can understand that the average of used words is 40% from the total words, while the unused words are 60% for all students.

No significant relation between number of words and the other levels. It means that the student may talks a lot but without significant expressions.

It is reasonable that level 1 and level 2 are joined together because we found uniform relation between their count and usage. At the same time we need to find the rate of each level occurrence from the total used words during the session.

In table 4, the rate of levels can be seen after mixing the first two levels into one level and divide into the number of words in chat session for each student. Table 4 is a sample.

Table 4: Sample of Expressions Rate in Chat

Number	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Adaptive 1	25.1%	6.8%	4.2%	0.5%	0.0%	4.2%	0.5%
Adaptive 11	33.7%	3.1%	3.6%	0.5%	2.1%	2.6%	0.5%

To understand the relation between each level and the total words used in chat we got the percentage rate for each level in reference to the total number of words, and at the same time, we calculated the rate for the same levels to used expressions only. Logically, it found that using the used expressions as the reference base is better than the total words of the chat seen in table 4 since it is looking for the expressions used by the student and useful not the unused expressions. In order to use regression analysis for extracting the parameters used in text model, we asked three experts in database systems to read the chat text for all students and give us their own score .We used the average score from theses scores.

Table 5 shows a sample for the expert score for each level which we need to extract the text model equation using regression analysis.

Table 5. Sample of Expert Score

Number	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Adaptive 1	26.8	9.0	4.6	0.6	0.0	4.6	0.3
Adaptive 11	42.8	1.3	3.3	0.5	1.0	1.2	0.3

Now we have two tables have contained the experiment result and the expert's evaluation. In order to get the proper value for each parameter, the regression analysis is needed to be used which has the following form:

$$Y = b + aX \qquad (3.1)$$

The following table shows the equations for each level using the last predefined procedure.

Table 6: Regression Analysis Result

Level 1 equation	(-11.489+1.6x)
Level 2 equation	(-1.127+1.413x)
Level 3 equation	(-0.8+1.256)
Level 4 equation	(0.074+1.416x)
Level 5 equation	(-0.491+1.173x)
Level 6 equation	(-0.927+1.44x)
Level 7 equation	(-0.447+1.59x)

From the experiment we extracted the occurrence rate for each level to the total occurrences for all levels .The result is shown in table 7.

Table 7: Levels occurrences

Level	Rate
Level 1	0.601
Level 2	0.132
Level 3	0.065
Level 4	0.028
Level 5	0.051
Level 6	0.096
Level 7	0.027

From the last two tables we can write down the equation for text modeling in adaptive chat room.

Text Model = 0.601*level 1 equation
+ 0.132 * level 2 equation
+ 0.065 * level 3 equation
+ 0.028 * level 4 equation
+ 0.051 * Level 5 equation
+ 0.096 * level 6 equation
+0.027 * level 7 equation …………… (3.2)

Text Model = 0.601*(-11.489 + 1.61*level 1)
+ 0.132*(-1.127 + 1.413 * level 2)
+ 0.065*(-0.8 + 1.256 * level 3)
+ 0.028*(0.074 + 1.416 * level 4)
+ 0.051*(-0.491 + 1.173 * level 5)
+ 0.096*(-0.927 + 1.44 * level 6)
+ 0.027*(-0.447 + 1.59 * level 7) ……. (3.3)

Simplifying this equation, it gets the final Text model equation as the following equation 4.10:

Text Model (exp1) = -7.71 + (0.97 * Rlevel1)

+ (0.19 * Rlevel2)

+ (0.08 * Rlevel3)

+ (0.04 * Rlevel4)

+ (0.06 * Rlevel5)

+ (0.14 * Rlevel6)

+ (0.04 * Rlevel7) ……………… (3.4)

Referring to the second experiment we performed on the same students but on different concept. The new concept is the mathematical operation on binary numbers.

Chat Text Analysis for Experiment 2 will have the same procedure as experiment 1 .The following equation is the text model extracted from experiment 2.

Text Model (exp2) = 0.601*(-12.595+1.395*level 1)

+ 0.046*(0.431+1.648* level 2)

+ 0.098*(-0.487+1.506* level 3)

+ 0.061*(-1.045+1.469* level 4)

+ 0.019*(0.581+2.167* level 5)

+ 0.156*(7.408+1.415* level 6)

+ 0.019*(2.311+0.39* level 7) ….(3.5)

Now we need to calculate the final form of Text Model .By taking the average of the two experiments the final form of the text model got as the following:

Text Model Final Form = -6.91 + (0.9 * Rlevel 1) **+ (0.14 * Rlevel 2)** **+ (0.11 * Rlevel 3)** **+ (0.06 * Rlevel 4)** **+ (0.06* Rlevel 5)** **+ (0.18 * Rlevel 6)** **+ (0.02 * Rlevel 7)** **……………… (3.6)**

While Rlevel i= leveli / used words

Level i=count of level words occurence during the chat session

4. CONCLUSION AND DISCUSSION

The text model described and ensures that the student used his chat expressions efficiently in the chat session. Chat expressions are from several levels and each level should have rate of occurrences. Two of these levels are related directly to the concept of chat while the others are general words affect the chat session quality .Using these levels and applying two experiments we got some parameters.

Text Model Average = -6.91 + (0.9 * Rlevel 1)

+ (0.14 * Rlevel 2)

+ (0.11 * Rlevel 3)

+ (0.06 * Rlevel 4)

+ (0.06* Rlevel 5)

+ (0.18 * Rlevel 6)

+ (0.02 * Rlevel 7) …………… (3.7)

This equation can be used in any chat session to model and evaluate the student through the expressions he used during the session.

In the future work, we are going to find a methodology for peer evaluation model in order to complete the student model in AVCM chat tool.

5. REFERENCES

[1] Brusilovsky, P. "Adaptive hypermedia for education and training," Adaptive Technologies for Training and Education, Cambridge University Press, Cambridge, UK. 2012. pp. 46-68.

[2] Jabari, N.A.M., Hariadi, M. & Purnomo, M. H. "Intelligent Adaptive Presentation and E-testing System based on User Modeling and Course sequencing in Virtual Classroom". IJCA 50 (9): 0975-8887. 2012.

[3] Hanini, M. Tahboub, R. and Jabari, N. A. M. " Student Modeling In Adaptive Educational Chat Room". JATIT (Journal of Theoretical and Applied Information Technology). 31st December 2013. Vol. 58 No.3

[4] Tahboub, R. and Hanini, M. "Time Modeling in Educational Chat Room". (European Journal of Scientific Research). April, 2014. Volume 122. No1, pages 27-35.

[5] T. Davey, A. Envall, M. Gernerd, T. Mahomes, M. Monroe, J. Nowak, M. Patricoski, and J. Weiler, Instant messaging: Functions of a new communicative tool," Anthropology 427 Doing Things With Words, Master thesis, 2004.

[6] Jabari and Haridi" Adaptive Web-Based Virtual Classroom based on Student Modeling". ITS. 2010.

[7] J. Han, M. Kamber, and J. Pei, 2012. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers is an imprint of Elsevier. 225 Wyman Street, Waltham, MA 02451, USA, c 2012 by Elsevier Inc. All rights reserved, third edition, isbn 978-0-12-381479-1 ed.

[8] K. C. Chen, "Text mining ecomplaints data from eauction store with implications for internet marketing research," Journal of Business & Economics Research, USA, 2009. vol. 7, no. 5, May.

[9] H. Dong, S. C. Hui, and Y. He, "Structural analysis of chat messages for topic detection," Online Information Review, 2 May 2006. vol. 30, no. 5, p. 496516.

[10] S. C. Hui, Y. He, and H. Dong, \Text mining for chat message analysis," Cybernetics and Intelligent Systems, 2008 IEEE Conference, pp. 411-416, 2008.

[11] R. L. Mason and R. F. Gunst, Statistical Design and Analysis of Experiments With Applications to engineering and Science.2003. A JOHN WILEY & SONS PUBLICATION, Southern Methodist University, second edition.