Identification of Agreement and Disagreement in Chat Conversations

Rasid Cetin

Politehnica University of Bucharest
313 Splaiul Independentei,
Bucharest, Romania
cetinrasid3@gmail.com

Ştefan Trăuşan-Matu

University Politehnica of Bucharest
313 Splaiul Independentei,
Bucharest, Romania
and
Research Institute for Artificial Intelligence
and
Academy of Romanian Scientists
stefan.trausan@cs.pub.ro

ABSTRACT

A recent pedagogical approach is computer-supported collaborative learning (CSCL) where by means of collective interaction through collaborative interfaces on internet, learning is accomplished by joint knowledge building. One of the most relevant tools that a CSCL uses is chat, which is an effective way to transfer information and collaborate. The purpose of this paper is to offer a technique of analyzing the composition of chats with a design to find the points of convergence and divergence between participants. The points of convergence that are detected in a chat discussion could emphasize a moment where a fact is validated, or a better solution for a problem is discovered or a new concept was assimilated by someone. Going in the same direction, the points of divergence also provide valuable information, because they capture moments where a fact is invalidated or a new idea about an existing topic is introduced. The starting point of this method has its origins in Bakhtin's dialogism and Trausan-Matu's polyphonic model.

KEYWORDS: CSCL; chat; convergence; divergence; polyphony; dialogism; pattern

ACM Classification Keywords

I.2.7 Natural Language Processing: Text analysis.

INTRODUCTION

Dialogue is an exchange of ideas and opinions on a particular issue, and its structure resembles that of a game, with turns and utterances (the unit of analysis), where every participant waits for his turn to say something, or to play [1]. An utterance is the act of saying. For example, if we have a sentence "Today it is not tomorrow", and someone read it with loud voice two times, he emits two utterances of the same sentence. An utterance is an emitted "piece of language", which has a moment of silence from the utterer before and after it. The essence of an utterance is given by the fact that it has a time, a place, a speaker, a language in which the message transmitted is encoded, but has no special form or content.

Computer-supported collaborative learning (CSCL) is a recent pedagogical approach on which, by means of collective interaction through collaborative interfaces on internet, for example chat (instant messaging), learning is accomplished by joint knowledge building [2]. One advantage of instant messaging is that the participants may introduce utterances at very close moments, even concomitantly, thus a higher degree of interaction is achieved than in face-to-face conversation [3].

System validation is done by using a corpus of chat conversations performed by students. The corpus consists in chat conversations between more than three students. They had to find advantages and disadvantages for how chats, blogs, forums, wiki, and google wave help in collaborative learning. The conversation had to be divided in two parts. The first part consisted in a debate, where to every student is assigned one technology and they had to persuade the others that their technology is the best choice and in the same time to emphasize the disadvantages of using the other technologies. In the second part, the students should reach to a common point [4]. Because of this format of the discussion, various points of convergence and divergence should arise between participants.

The idea of collaborative learning is that a group of people discuss for trying to jointly understand new topics or to find meanings to some concepts or to find solutions to some problems. One advantage is the appearance of controversies, which marks the fact that some participants have conflicting ideas with others. These moments are beneficial because some flaws in the way of thinking are detected and they may be corrected. Another advantage is that because of the high degree of interaction, new ideas are generated and shared between the participants, thus the construction of knowledge is enabled.

There are several applications of the approach presented in this paper. First, it may enhance the collaboration in the classroom by helping the teacher to manage the students. By using this program, the students can have their individualized learning plan that is created by the teacher after keeping track

of their achievements, of how far they are from accomplishing the targets that were set, taking account of their involvement inferred from their agreements and disagreements. Also, collaboration between students can be improved through this technology, because the teacher can group the students using the application, avoiding, for example, to put too many students in the same group if they have a profile with frequent disagreements. An idea that should be investigated in the future is that groups should have students with balanced agreement/disagreement profiles.

In the next section is presented the theoretical background which consists in Trausan-Matu's polyphonic model [3, 8, 9], based on Bakhtin's ideas about dialogism and polyphony [7]. It is also presented the concept of points of divergence and convergence. In the third section are presented some implementation details of the system. In the fourth section are shown some results of the actual system and in the last section are drawn some conclusions.

STATE OF THE ART

The Polyphonic Model

Dialogue is a result of a mental and verbal effort which appears when a group of people tries to endorse their way of thinking by providing personal assertion offered as evidence. In order to give voice to their ideas, each participant attempts to choose the right words. Commonly, the dialogue resulted from an educational interaction is constrained by certain restrictions that guides the discussion and not let it to divert from the main topics.

In Bakhtin's theory, the dialogue (as any language communication, including texts) is multivocal [6, 7], it appears as a strong interaction between different perspectives, which are pictured into words and which reflects the participants. Usually, some words are repeated during a dialogue in order to emphasize a particular idea. As a result, starting from Bakhtin's ideas and following the polyphonic method of analysis [6] the main topics or concepts considered in the dialogue can be tracked by finding the most repeated words [6].

Through a dialogical point of view, everything in a chat is connected. Every participant has free will in what he/she says and an individual path in the discussion. However, as Bakhtin wrote, every utterance has echoes of the preceding ones, so they are related all together [7]. An utterance is in stronger relations with recent past utterances to which responds and with future ones whose response it anticipates. Threads of ideas are generated that exist in parallel and that inter-animate according to divergent and convergent patterns [9]. In other words, in chats appear phenomena that resemble with the polyphonic model, in which several voices with individual personalities inter-animate according to dissonant and consonant chords, the first ones driving to the latter.

Convergence and Divergence

In the CSCL dialogues, students will try to persuade the other students that his ideas and opinions are the most suitable for the given situation. Each one of them will materialize their way of thinking into words and focus on the topic received. Therefore, this interaction will create a transfer of ideas between students, and points of convergence and divergence will appear when participants try to prove their views [9].

The points of convergence depict moments in the conversation where a fact is approved or a better solution is achieved by analyzing the available information that was provided by emitters [9]. The points of divergence within educational dialogues depict moments where creativity appears, because new ideas are introduced by emitters, ideas that can offer a new perspective which can lead to a new path and make to take into considerations other aspects [9]. It is not considered that the points of divergence emphasize moments where one of the participants are wrong, because the truth is not just only one view but a combination of views (convergent or divergent), it may appear as a result of interanimation [9]. Also, the points of divergence can be used as a measure of compatibility between students. When two students have several moments of disagreement along the chat discussion, it means that they have very different views, and each one of them can be a factor of horizon broadening for the other.

In terms of knowledge construction, convergences and divergences are different. Initially, in the chats we analyzed [4] each student is focused on his topic and generates a thread of discussion. The convergences of views suggest that links are generated between these threads, or can be seen as an approval on these discussion threads. The divergence of views suggests that the participants are focused on their own generated thread by being in contrariety with the other threads.

The purpose of the chat is to make its participants to engage in a discussion where they can share their personal knowledge and build new understanding of specific topics using the information shared by the other participants. The exchange of information is realized in more steps. The first step consists in bringing by the student the information that he thinks is appropriate to highlight the advantages of his topic. In the second step, the points of convergence and divergence will arise, because the other students will accept or refuse the ideas explored in the first step.

The approval among participants and the degree of positive interaction in the knowledge transfer are identified by finding the points of convergence and divergence between their views. The environment in which the discussion is taking place should be an encouraging one, where students can express themselves freely, so disagreements should not be made in such a way that a conflict may start or may inhibit other students in sharing new ideas. Disagreements should be a way in which the diversity of information must be

expanded, in other words, the new ideas shared may help others to rectify their beliefs or learn something new.

IMPLEMENTATION

In this section, it is explained the system and the way in which the points of convergence and divergence are detected in chats. These chats have a special structure, which is different from the structure of chats that take place in social networks. This structure is represented in XML format and consists in: nickname of the utterer, the unique id of the utterance emitted by the utterer, the unique id of the explicitly referenced utterance (or -1 if there is not such a link was provided by the user), the time when the utterance is entered, and the text of the utterance [4]. The referenced utterance means that the actual utterance is a response to the referenced one, the echo of the previous one affects the actual one.

In Figure 1 are showed the main steps that the system performs.

Step 1
Chat parsing and explicit links detection - in this phase the chat is parsed from the XML format, and the information contained in that format is saved in data structures that will facilitate the processing. Also, in this phase it is realized the detection of the explicit links, that are pointed by the chat participants when entering a new utterance. An explicit link is a link between two utterances, where usually the later utterance is in response to the earlier one.

Step 2
Detection of repetition links – in this phase all utterances are divided into groups of a given length (ex: 20, 40, 100). The higher the length of a group is, the fewer groups will result after division. For every group are found the most frequent words that are used by participants using the NLTK library (which allows to access natural language processings such as stop words removing, tokenization, and lemmatization [5]). Unfortunatelly, most of the frequent words are not relevant, and will be of no help in finding agreements and disagreements. Because of that, after the previous step, a window with all of the most frequent words for every group is displayed, where the user can choose the most relevant words for every group.

Then, repetition links are identified, where a repetition link is a link between two utterances from the same group and in the text of the utterances is found the same word from those selected in the previous step.
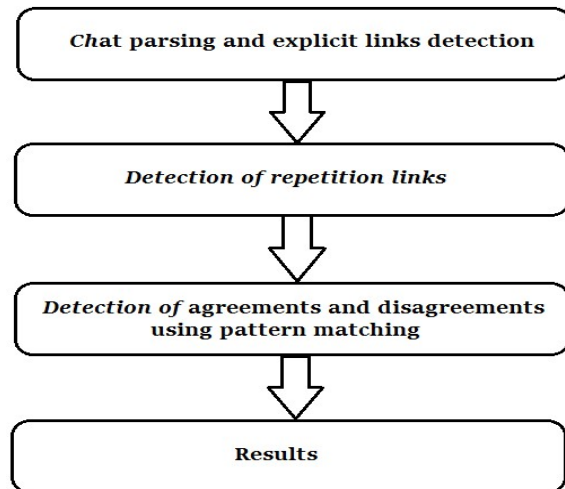

Figure 1. Implementation steps of the system

Step 3
Detection of agreements and disagreements using pattern matching - in this phase a pattern matching module is used to detect points of convergence and divergence in chat. Explicit links are between two utterances, and to detect an agreement or disagreement in these links, the pattern matching module (see below) is used on the utterance with higher id (this utterance is considered a response to the utterance with lower id) and see if matches on any regular expression.

Step 4
Results – in this phase are shown in text format the points of convergence and divergence and a graphic representation of how this moments can be interpreted

The Pattern Matching Module
The pattern matching module, inspired from that one developed in the project LTfLL [10], can be used for finding occurrences in a chat that match a given expression. The module receives as input the text of the utterance and the expression that should be matched. The output is the matching solution if it exists.

This tool is suitable and very helpful for achieving the purpose of this task because in inter-human communication agreements and disagreements are expressed by specific patterns. These patterns will be the matching expressions used by the pattern searching module that will find the moments of interests in a chat, discussed in this paper.

In Table 1 are presented the elementary expressions that can be used to create complex expressions that the pattern matcher can use to detect points of convergence and divergence in utterances.

Elementary expressions	Matching
word	any occurence of this word
„text"	any occurrence of this text (more words)
*	any string (even null string)
<D „stem">	any word which has the stem „stem"
<S „word">	any word that is a synonym of the word "word"
<NN>, <PRP>, <VBD>	matches any word that was annotated by the part of speech tagger with that label

Table 1. Elementary expressions

Operators help to create more complex expressions by combining more elementary expressions. Examples are shown in Table 2 and there is also presented their meaning and the way in which these operators can be used.

Operator	Symbol	Matching of new expression
Concatenation	Exp1 Exp2	The resulting expression matches a text if there is text1 that matches Exp1 and text2 that matches Exp2 and text = text1 + whitespaces + text2.
AND	Exp1 & Exp2	The resulting expressions matches text only if Exp1 matches text and Exp2 matches text.
OR	Exp1 | Exp2	The resulting expressions matches text only if Exp1 matches text or Exp2 matches text.

Table 2. Pattern matcher's operators for creating complex expressions

For grouping the expressions normal parenthesis can be used. By default, the evaluation of expressions is made by the pattern searching module from left to right. For example, Exp1 Exp2 & Exp3 is interpreted as (Exp1 Exp2) & Exp3 which is far different from Exp1 (Exp2 & Exp3).

In Table 3 are shown some expressions that the pattern searching module is using to detect the moments of interest from chat.

RESULTS
The analysis of points of convergences and divergences were made on a chat which is in xml format. In the first part of this section will be shown the Convergences/Divergences between participants. In Figure 2 are shown this relations between participants represented graphically.

Expressions for detecting agreements		Expressions for detecting disagreements	
Expression	**Examples of matches**	**Expression**	**Examples of matches**
So (<VB> | <VBP> | <MD>) <PRP>	So do I	(<PRP> (<VB> | <VBP> | <MD>) not) | (<PRP> <VBP> <VB>)	I don't think so I do not I do not think so
<PRP> too	me, too	<PRP> (<VB> | <VBP> | <MD>) either	I don't either! I can't either! I won't either! I didn't either! I wouldn't either!
Neither (<VB> | <VBP> | <MD>) <PRP>	neither do I	I don't agree	I don't agree
(<DT> | <PRP>) (<VBZ> | <VB>) true	that is true it is true	Not at all	Not at all
<PRP> (<MD> | <VBP> | <VBD>) too	I will too, I can too, I did too, etc.	I don't agree	I don't agree

Table 3. Expressions for detecting points of convergence and divergence
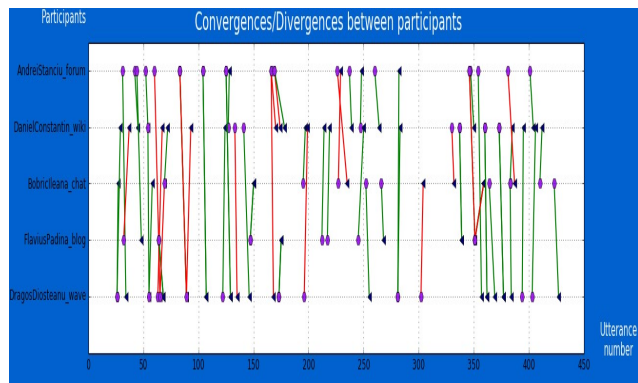

Figure 2. Convergences (in green) and Divergences (in red) between participants

Because there are too many moments to analyze, for exemplification we will focus on what is identified by the system between utterances 150 and 200. In Figure 2 is shown that in this interval were found seven moments of interest, where four of them are divergent (marked with red line) and three of them are convergent (marked with green line). From the graph is hard to detect the exact utterances that are implied in a given moment. To identify the utterances more precisely, a textual representation is provided, which shows that the first moment identified after the utterance 150 begins at utterance 166.

"DIV Matched expression: <PRP> (<VB> | <VBP> | <MD>) wrong
The referred utterance:
User: AndreiStanciu_forum
ID: 166
Text: wikis, by deffinition, are made for public wide use
Utterance which express disagreement:
User: DragosDiosteanu_wave
ID:168
Text: You are wrong here, Costa can tell us more"

For the next moments is hard to identify on the graph which one was first. From the textual representation, the moments are extracted :

"DIV Matched expression: Not necessarily
The referred utterance:
 User: AndreiStanciu_forum
 ID: 166
 Text: wikis, by deffinition, are made for public wide use
Utterance which express disagreement:
 User: DanielConstantin_wiki
 ID:170
 Text: not necessarily. we can restrict access to certain articles. that way" our wiki will have some sections for the general public who want to find out some details about our technologies

"CONV Matched expression: agree
The referred utterance:
 User: AndreiStanciu_forum
 ID: 169
 Text: yes, they have wikis that are publicly available, with public information, for the everyday user that takes an intrest in that company's products
Utterance which express agreement:
 User: DanielConstantin_wiki
 ID:178
 Text: We can include that information in a wiki too. but I agree with Dragos's suggestion. Besides, the main purpose of our wiki would be to include technical details"

"DIV Matched expression: No way
The referred utterance:
 User: AndreiStanciu_forum
 ID: 169
 Text: yes, they have wikis that are publicly available, with public information, for the everyday user that takes an intrest in that company's products
Utterance which express disagreement:
                    User: DanielConstantin_wiki
 ID:174
 Text: besides there is no way that you can structure a database to make information easier to access that a wiki"

"CONV Matched expression: agree with <PRP>
The referred utterance:
 User: DragosDiosteanu_wave
 ID: 173
 Text: We can use a wweb site or a blog for doing that Andrei.
Utterance which express agreement:
 User: FlaviusPadina_blog
 ID:175
 Text: i agree with you and we also can post news or updates for our proucts"

From those textual representations and the graphic representation from Figure 2, another moment of divergence

was found between utterances (166, 170), and two moments of convergence were found between utterances (169, 178) and (173, 175). Also, another moment was identified between utterances (169, 174) which doesn't really express a disagreement, but is considered a point were new information is shared with the other participants.

CONCLUSIONS

The proposed system aims to implement a method that is able to analyze chats in xml format and to detect points of convergence and divergence. The accuracy of the results depends on the agreement and disagreement patterns that are selected which must be as precise as possible. The accuracy of the system can be increased if more patterns for detection agreements and disagreements are added. Another way to improve the results is to use better patterns that are suitable to detect points of convergence and divergence and no others matchings that would not be considered adequate. This system is useful because makes possible to find the degree in which the students participate to the process of knowledge building.

REFERENCES

1. Trăuşan-Matu, S., Chiru, C., Bogdan, R., Identificarea actelor de vorbire in dialogurile purtate pe chat, in Ştefan Trăuşan-Matu, Costin Pribeanu (Eds.) *Interacţiune Om-Calculator*, Editura Printech, Bucureşti, (2004) pp. 206-214

2. Stahl, G. *Group cognition: Computer support for building collaborative knowledge.* Cambridge, MA: MIT Press, (2006).

3. Trăuşan-Matu, S., Dascalu, M., Rebedea, T., PolyCAFe—automatic support for the polyphonic analysis of CSCL chats, *International Journal of Computer-Supported Collaborative Learning*, Volume 9(2), Springer, (2014), pp. 127-156

4. Trăuşan-Matu, S., Dascalu, M., Rebedea, T., Gartner, A., *Corpus de conversaţii multi-participant şi editor pentru adnotarea lui,* Revista Română de Interacţiune Om-Calculator 3(1), (2010), pp. 53-64

5. http://www.nltk.org/ (last accessed 30th June 2017)

6. https://ceasefiremagazine.co.uk/in-theory-bakhtin-1/ (last accessed 30th June 2017)

7. Bakhtin, M.M. The dialogic imagination: Four essays (trans: Emerson, C. & Holquist, M.). Austin and London: The University of Texas Press, (1981)

8. Trausan-Matu, S., The Polyphonic Model of Hybrid and Collaborative Learning, In Wang, F.,L., Fong., J., Kwan, R.C., *Handbook of Research on Hybrid Learning*

*Models: Advanced   Tools,   Technologies,   and
Applications*, Information Science Publishing, Hershey,
New York, (2010), pp 466-486

9. Trausan-Matu, S., Stahl, G., Sarmiento, J., Supporting
Polyphonic Collaborative Learning, E-service Journal,
vol. 6, nr. 1, Indiana University Press, (2007), pp. 58-74.

10. Trausan-Matu, Stefan; Dessus, Philippe; Rebedea,
Traian; Mandin, Sonia; Villiot-Leclercq, Emmanuelle;
Dascalu, Mihai; Gartner, Alexandru; Chiru, Costin;
Banica, Dan; Mihaila, Dan; Lemaire, Benoît; Zampa,
Virginie; Graziani, Eugène, LTfLL - D5.2 Learning
support and feedback, (2009)