Detecting Conflicts in Collaborative Learning through the Valence Change of Atomic Interactions

Germán Lescano[a], Jose Torres-Jimenez[b], Rosanna Costaguta[a], Analía Amandi[c], Carlos Lara[d,*]

[a] *Instituto de Investigación en Informática y Sistemas de Información*
Universidad Nacional de Santiago del Estero. Argentina

[b] *Cinvestav Unidad Tamaulipas, Tamaulipas, México*

[c] *Núcleo de Inteligencia Comportamental Empresarial, CONICET-UNICEN, Argentina*

[d] *Centro de Investigación en Matemáticas, A.C., México*

Abstract

Naturally, every collaboration will bring conflicts that can affect the performance of a team. The earlier a conflict is detected and managed in a collaborative group, the better. Detecting and tracking conflicts in Computer-Supported Collaborative Learning (CSCL) is laborious work. If the teacher does it, the intervention may be out of time. Although written dialogues in groups having a conflict reveal the increment of negative emotions in comparison to non-conflict dialogues, a classifier that only uses statistics of the valence of consecutive messages in a window of the talk shows poor performance. This paper proposes to use features based on the valence change between a message and its response. In this way the algorithm focuses in the kind of interaction. We study different implementations of the bootstrap aggregating technique to detect conflicts. Results obtained show the viability of the proposed approach.

Keywords: CSCL, Conflicts, Emotions, Sentiment Analysis, Bootstrap aggregating

1. Introduction

In collaborative learning (CL), students are grouped and paired to achieve an academic goal (Gokhale, 1995). In Computer-Supported Collaborative Learning (CSCL), learners interact with the aid of computers. Arranging students in groups and assigning them a task does not guarantee that students will engage effectively (Soller, 2001). Benefits are mainly observed when the group works correctly.

There are plenty of theories and frameworks that support collaborative learning. Stahl & Hakkarainen (2020) categorize CSCL theories according to the focal unit of collaboration as (i) *subjective theories*, that focus on the individual mind; (ii) *inter-subjective theories*, oriented to social, community, and cultural levels of analysis; and (iii) *inter-objective theories*, that focus on the group itself as the unit of analysis. Even the subjective theories admit that student learning is influenced by the social context. Committed groups are composed of participants that encourage other members by asking questions, explaining and justifying their opinions, articulating their reasoning, elaborating, and reflecting on their knowledge (Soller, 2001). However, interactions that promote learning do not occur spontaneously (Onrubia & Engel, 2012; Schwarz & Asterhan, 2011; Tchounikine et al., 2010), and the use of computers can even inhibit them (Orvis & Lassiter, 2006).

CSCL groups are exposed to socio-emotional challenges, such as understanding others' perspectives, communication styles, and ways of approaching teamwork (Järvenoja & Järvelä, 2013). In the interaction between members of a group, conflicts emerge naturally (Näykki et al., 2014). Conflicts are disagreements between two or more members of the group (Ayoko et al., 2008). Conflicts generate mostly negative emotions and therefore, affects the group work (Dreu & Weingart, 2003; Jehn, 1997; Jiang et al., 2013; Lee et al., 2015). Conflict resolution depends on the emotional intelligence, social relationships, trust, group cohesion, and social skills manifested by the members (Jiang et al., 2013; Kreijns et al., 2003; Kwon et al., 2014; Lee et al., 2015; Rapisarda, 2002; Slof et al., 2016).

The chat is a common communication channel in CSCL, where participants usually communicate in plain text and emoticons. This channel limits the exchange of emotional expressions because participants cannot perceive gestures or tones of voice of their peers (Feidakis et al., 2014). This limitation affects the group's ability to generate cohesion, feelings of trust, belonging, and satisfaction (Nam, 2014). Analogously, it is harder for teachers to recognize the group's emotions, mainly because of the loss of visual contact and acoustic communication.

Sometimes, teachers want to improve the interpersonal skills of students (Rogat & Adams-Wiggins, 2015; Zheng & Huang, 2016); for instance, they want to instruct students on conflict management strategies and how they are linked to group performance and satisfaction (Lee et al., 2015; Slof et al., 2016). In these cases, it is convenient that the

*Corresponding author

Email addresses: gelescano@unse.edu.ar (Germán Lescano), jtj@cinvestav.mx (Jose Torres-Jimenez), costaguta@unse.edu.ar (Rosanna Costaguta), amandi@exa.unicen.edu.ar (Analía Amandi), carlos.lara@cimat.mx (Carlos Lara)

group's activities can generate socio-cognitive or task conflicts (Buchs et al., 2004). Then, the teacher requires to i) identify conflicts, ii) review the group's interactions following some protocol, and iii) provide feedback that allows students to learn interaction styles (Buchs et al., 2004).

This article addresses the problem of conflict recognition in text interactions. We argue that it is possible to recognize the presence of conflicts through the analysis of the implicit emotions of the text messages. The contribution of this paper is twofold; first, we study the differences in the valence of text messages between conflict and non-conflict situations. Second, we propose a model to recognize conflicts by using statistics of the change of valence between messages and their responses.

The rest of this paper is organized as follows. Section 2 sets out fundamental concepts involved with the subject of study. Section 3 reviews some related literature. Section 4 describes the feature extraction from text messages and reviews the bootstrap scheme used for classification. Section 5 reveals the research methodology. Section 6 presents the experimental results. Section 7 presents a discussion of the work made, and finally, Section 8 concludes this work.

2. Background

2.1. Conflicts in groups

Conflicts can be categorized as task and relationship conflicts (Jehn, 1997). The former are disagreements regarding opinions, perspectives, and ways of processing the task information (Ayoko et al., 2008; Garcia-Prieto et al., 2007; Lee et al., 2015). This type of conflict is positive for the group because it facilitates learning and the development of creative thinking (Jiang et al., 2013), the reevaluation of the status quo, and the elimination of complacency (Ayoko et al., 2008). It promotes innovation, cohesion, commitment, and satisfaction. It also helps to increase the group performance, levels of understanding, and trust levels among group members (Dreu & Weingart, 2003; Jehn, 1997; Lee et al., 2015; Garcia-Prieto et al., 2007).

Relationship conflicts are disagreements between group members caused by issues unrelated to the task. These conflicts are often manifested as frictions, frustrations, and personality clashes (Ayoko et al., 2008; Garcia-Prieto et al., 2007; Jehn & Mannix, 2001). This type of conflict generates stress and decreases cohesion, commitment, satisfaction, performance, and quality in decision-making (Näykki et al., 2014; Dreu & Weingart, 2003; Jehn, 1997; Lee et al., 2015). Regardless of the type of conflict, the emotions involved are negative. Negative emotions decrease the capability of working memory, probably by producing an extraneous cognitive load (Dreu & Weingart, 2003). Thus, a substantial increase of negative emotions hamper the group's work (Dreu & Weingart, 2003; Jehn, 1997; Jiang et al., 2013; Lee et al., 2015), mainly because of the dissemination of negative emotions – aka emotional contagion (Barsade, 2009).

2.2. Emotions in collaborative learning

Groups that learn in CSCL environments are exposed to multiple situations that characterize the social dynamic and generate emotions influencing collaborative learning (Järvelä & Hadwin, 2013). A typical emotion is frustration (Capdeferro & Romero, 2012). Emotions in academic environments attract the interest of many researchers; for instance, Reis et al. (2018) consider that the more common emotions are frustration, fear, anger, relaxed, bored, anxious, joy, interested, among others. Pekrun et al. (2002) group academic emotions into four categories: achievement, epistemic, topic, and social emotions.

Dimensional models of emotions try to describe emotions by their position in two or three dimensions. The bidimensional model describes emotions by their valence (ranging from feeling pleasant to unpleasant) and arousal (ranging from feeling quiet to active) (Coppin & Sander, 2013). Through these two dimensions, it is possible to describe subjective experiences. Pekrun's emotion taxonomy (Pekrun et al., 2011) is consistent with circumplex models; that is, every affect can be the two-dimensional (valence/activation) space.

2.3. *Sentiment Analysis in Text*

Sentiment analysis tries to find sentiments and emotions that people express toward products, organizations, people, and other topics (Appel et al., 2016; Keshavarz & Abadeh, 2017; Sanglerdsinlapachai et al., 2014). It is a challenging task because of the ambiguity of words, the complexity of semantic, and the influence of irony, courtesy, writing style, and the language variation among people and culture (Poria et al., 2014). Coarse-grained sentiment analysis estimates the valence of a text (positive, negative, or neutral). On the other hand, the fine-grained analysis tries to find the specific emotions. (Giatsoglou et al., 2017).

There are two approaches to detect sentiment in text: (i) the lexical approach uses a 'lexicon' to associate words to a given variable; e.g., the intensity of each word (Keshavarz & Abadeh, 2017). The result will be the sentiment expressed by the majority of the text words, (ii) the machine learning approach builds a model that recognizes sentiment via the identification of textual patterns (Sanglerdsinlapachai et al., 2014). This model is built by using supervised machine learning, unsupervised machine learning, or a mixed approach. The machine learning approach requires large training data (Madhoushi et al., 2015; Wang et al., 2016).

3. Related Work

Researchers have proposed several features to detect conflicts. For instance, Millar et al. (1984) study the power dimension of interactions tagged by three alternatives: control gain (↑), give up control (↓), or neutralize control (→). They propose that a sequence of three alternative attempts to gain control is a pattern of conflicts. The *Interaction Process Analysis* (Bales, 1950) assigns a tag to each message in a conversation to describe the interlocutors' behavior. Bales (1950) states that the tag frequency can be used to infer six types of problems.

The SYMLOG methodology (Bales & Steven, 1979) maps groups and their members into a tridimensional space: dominant/submissive, friendly/unfriendly, and task-oriented/emotional-expressive dimensions. To reveal how frequent specific behaviors happen, group members should respond to a questionnaire of 26 Likert-questions. Wall & Galanes (1986) found that conflicts are not significantly related to the dominant/submissive and the task-oriented/ emotional-expressive dimensions; on the contrary, the friendly/unfriendly dimension is related to conflicts.

To some extent, the friendly/unfriendly dimension is characterized by the valence dimension of emotions. Several works (Lescano et al., 2020; Wall & Galanes, 1986; Zachary, 1977; Zakaria et al., 2009) suggest that the valence of emotions can be used to detect conflicts. Another approach is to detect conflicts via specific emotions; for instance, Zakaria et al. (2009) suggest that some negative emotions generate conflicts (e.g., hate and disgust) and some others (e.g., indifference, misconception, among others) can indirectly predict conflicts. In this work, we show that negative emotions of interactions can be used to recognize the presence of conflicts in CSCL environments.

The graph-based approaches to conflict detection build a graph that represents relationships between users. Zachary (1977) uses social network techniques to identify fission in a group. Fission is a phenomenon about subgroup formation due to the differences in the way sentiments are shared. Lescano et al. (2020) propose to calculate the commute times of a normalized graph in which edges represent the members' emotional interactions.

There are other approaches to recognize conflicts, but they impose restrictions on the interface. For instance, Tedesco (2003) proposes a framework based on Belief–Desire–Intention logic (BDI logic) that allows recognizing and mediating conflicts in group planning interactions. Monteserin et al. (2010) define an argumentation plan as a process during a given negotiation; hence, they model the plan as a partial order sequence of arguments that allows students to reach an expected agreement. Their approach offers students assistance inspired by BDI agents (Rao et al., 1995). In contrast, the proposed approach does not impose restrictions on the interface, because it can be used to detect automatically conflicts in CSCL chats.

4. An Algorithm to Detect Conflicts in Collaborative Groups

This section describes the *Bagging Algorithm to Recognize Conflicts in Groups using Sentiments* (BARCIGUS), a technique based on supervised machine learning to detect conflicts in CSCL group interactions by considering the change of valence between messages.

4.1. Preliminaries

A *sequence*, S, is an enumerated collection of objects in which repetitions are allowed and order matters. A *subsequence* S' of S is defined as any series of elements that can be obtained from a given sequence S by deleting some of its elements. Let $D = S \setminus S'$ be the subsequence of elements removed from S to obtain S'. A sequence T is a supersequence of S, if T can be obtained by inserting elements into S. In a *contiguous subsequence* or *interval*, the elements (taken in order) are consecutive in the original sequence. A contiguous supersequence of S can be obtained by inserting elements at the beginning or end of S.

Let X be a discrete random variable with a range $x_0, x_1, ..., x_{J-1}$. The Probability Mass Function (PMF) gives the probability that the discrete random variable is exactly equal to some value. We write $p_X(x_i)$ to represent the probability that the random variable X equals x_i, $P(X = x_i)$. The joint probability distribution of two discrete random variables X, Y is denoted as $p_{XY}(x, y) = P(X = x, Y = y)$. For two independent variables $p_{XY}(x, y) = p(x)p(y)$.

The *averaging operation* combines m discrete random variables into a single discrete random variable. Let us denote as $C = \text{avg}(X_0, X_1, \ldots, X_{m-1})$ the average of random variables $X_0, X_1, \ldots, X_{m-1}$. The PMF of C is simply the normalized sum of the input densities, that is

$$p_C(x_i) = \frac{p_{X_0}(x_i) + p_{X_1}(x_i) + \ldots + p_{X_{m-1}}(x_i)}{\sum_{\forall x_j} \left[p_{X_0}(x_j) + p_{X_1}(x_j) + \ldots + p_{X_{n-1}}(x_j)\right]} = \frac{\sum_j p_{X_j}(x_i)}{m}. \tag{1}$$

.

4.2. Features Extraction

Our main hypothesis is that text messages between participants in a chat could give enough information to detect conflicts. Specifically, we analyze the emotional content of text messages. Thus, the interaction of participants can be characterized by the valence of their messages.

Consider a sequence of n text messages among participants U within a temporal window,

$$M = [\mu_0^{s_0,r_0}, \mu_1^{s_1,r_1}, \mu_2^{s_2,r_2}, \ldots, \mu_{n-1}^{s_{n-1},r_{n-1}}] = \left[\mu_i^{s_i,r_i}\right]_{i\in\{0,\ldots,n-1\}} \tag{2}$$

where, $\mu_i^{s_i,r_i}$ is a single text message, i is the number of the message, $s_i \in U$ is the sender, $r_i \in U \cup \{*\}$ is the receiver, $s_i \neq r_i$; here, a message to all participants is represented by $*$.

Let $I^{p,q} = I^{q,p}$ be the interaction set between p and q, where $\mu_i^{s_i,r_i} \in I^{p,q}$ is a single message between participants p and q; that is, $s_i \in \{p, q\}$ and $r_i \in \{p, q, *\}$. The *interaction list* $M^{p,q}$ between two participants p and q is defined as a subsequence of M that contains all messages exchanged between p and q in a given window; formally,

$$(\mu \in I^{p,q},\ \forall \mu \in M^{p,q}) \text{ and } (\mu \notin I^{p,q},\ \forall \mu \in M \setminus M^{p,q}). \tag{3}$$

A *two-way interval*, $T = S^p|S^q$, is a contiguous subsequence (interval) of an interaction list between any two participants p and q, that can be represented as a concatenation of two intervals S^p and S^q composed of messages sent by the same participant; i.e., messages of S^r are all sent by the participant $r \in \{p, q\}$ and the receiver is $r \setminus \{p, q\} \cup \{*\}$. An *atomic interaction*, A, is a two-way interval that does not have a contiguous supersequence forming a two-way interval.

Let $B = \{-, 0, +\}$ be the set of sentiment classes (negative, neutral, and positive), and C_T be a message-level classifier (e.g., an automated classifier or evaluation by experts) that assigns a PMF on B to each message μ.

Consider the interval $S^r = [\mu_{r_0}, \mu_{r_1}, \ldots \mu_{r_{n-1}}]$ of messages sent by participant $r \in \{p, q\}$. Let X_{r_i} be the random variable obtained by the classifier C_T from μ_{r_i}. To obtain the PMF of S^r we use the averaging operation, $S^r \sim \text{avg}(X_{r_0}, X_{r_1}, \ldots, X_{r_{n-1}})$; that is,

$$p_{S_r}(b_i) = \frac{\sum_j p_{X_j}(b_i)}{n}, \tag{4}$$

where n is the number of messages in the interval S^r.

msg ID	i	s_i	r_i	Interaction Text
$\mu_0^{1,2}$	0	u_1	u_2	bueno y ahora ¿qué leo?
$\mu_1^{3,*}$	1	u_3	*	nosotras tenemos que leer Gardner, inteligencia múltiples
$\mu_2^{1,3}$	2	u_1	u_3	eso estaba leyendo y me dijeron que no
$\mu_3^{2,1}$	3	u_2	u_1	Si, no sé, léamos éso porque la verdad no entiendo a la profe
$\mu_4^{1,2}$	4	u_1	u_2	yo tampoco entendí
$\mu_5^{3,*}$	5	u_3	*	si, sé qué tenemos que leer, pero del texto que nos dió la profe...
$\mu_6^{3,*}$	6	u_3	*	habla de muchos autores

(a)

$\Rightarrow$

msg ID	p(+)	p(-)	p(0)
$\mu_0^{1,2}$	0.02	0.01	0.97
$\mu_1^{3,*}$	0.01	0.04	0.95
$\mu_2^{1,3}$	0.00	0.97	0.03
$\mu_3^{2,1}$	0.00	0.99	0.01
$\mu_4^{1,2}$	0.01	0.22	0.77
$\mu_5^{3,*}$	0.01	0.44	0.55
$\mu_6^{3,*}$	0.04	0.04	0.92
$\mu_7^{1,*}$	0.00	1.00	0.00

(b)

$\Rightarrow$

users	interaction list
u_1, u_2	$M^{1,2} : [\mu_0^{1,2},\ \mu_3^{2,1},\ \mu_4^{1,2},\ \mu_7^{1,*}]$
u_1, u_3	$M^{1,3} : [\mu_1^{3,*},\ \mu_{1,3}^{2},\ \mu_5^{3,*},\ \mu_6^{3,*}]$
u_2, u_3	$M^{2,3} : [\mu_1^{3,*},\ \mu_5^{3,*},\ \mu_6^{3,*}]$

(c)

$\Rightarrow$

users	atomic int.
u_1, u_2	$A_1 : \mu_0^{1,2} \mid \mu_3^{2,1}$
	$A_2 : \mu_3^{2,1} \mid \mu_4^{1,2}, \mu_{1,*}^{7}$
u_1, u_3	$A_3 : \mu_1^{3,*} \mid \mu_2^{1,3}$
	$A_4 : \mu_2^{1,3} \mid \mu_5^{3,*}, \mu_6^{3,*}$

(d)

$\Rightarrow$

	S^2		
S^1	+	−	0
+	0.00	0.02	0.00
−	0.00	0.01	0.00
0	0.00	0.96	0.01

$A_1 : \mu_0^{1,2} \mid \mu_3^{2,1}$

	S^1		
S^2	+	−	0
+	0.00	0.00	0.00
−	0.00	0.60	0.39
0	0.00	0.01	0.00

$A_2 : \mu_3^{2,1} \mid \mu_4^{1,2}, \mu_7^{1,*}$

	S^1		
S^3	+	−	0
+	0.00	0.01	0.00
−	0.00	0.04	0.00
0	0.00	0.92	0.03

$A_3 : \mu_1^{3,*} \mid \mu_2^{1,3}$

	S^3		
S^1	+	−	0
+	0.00	0.00	0.00
−	0.02	0.23	0.72
0	0.00	0.01	0.02

$A_4 : \mu_2^{1,3} \mid \mu_5^{3,*}, \mu_6^{3,*}$

(e)

$\Rightarrow$

	S^{i+1}		
S^i	+	−	0
+	0.00	0.01	0.00
−	0.01	0.22	0.28
0	0.00	0.48	0.02

$f(M)$

(f)

Figure 1: **Feature extraction**. (a) Messages M of a collaborative group $U = \{u_1, u_2, u_3\}$, (b) PMF on sentiment classes obtained by the automatic classifier C_T for each message, (c) interaction list for every combination of participants, (d) the set of atomic interactions $\mathbf{A}_m = \{A_1, \ldots, A_4\}$, (e) probability table of changes in each atomic interaction, and (f) features of M ($f(M)$).

By considering independence of intervals S^p and S^q, the joint probability distribution of the atomic interaction $A = S^p|S^q$ is

$$p_A(b_1, b_2) = p_{S^p}(b_1) p_{S^q}(b_2) \quad \text{for all } b_1, b_2 \in B \tag{5}$$

To obtain the features of a temporal window M, the proposed approach finds the set $\mathbf{A}_M$ of all atomic interactions of M. The features of the window, $f(M) \in \mathrm{R}^9$, are found by averaging the PMF of atomic interactions in $\mathbf{A}_M$; that is, $f(M) = \mathrm{avg}(\mathbf{A}_M)$, the features of the window are used as inputs for the classifier. The complete feature extraction process is exemplified in Fig. 1.

4.3. Bootstrap aggregating

There are several types of supervised machine learning techniques. This paper focuses on those that use *Bootstrap aggregating*, aka bagging, which is commonly used to improve the stability and accuracy of the classification.

Given a learning set of N training samples, $\{(x_1, y_1), \ldots, (x_N,\ y_N)\}$ such that x_i is a feature vector and y_i is its label, a supervised machine learning algorithm seeks a function $g : X \mapsto Y$, where X is the input space and Y is the output

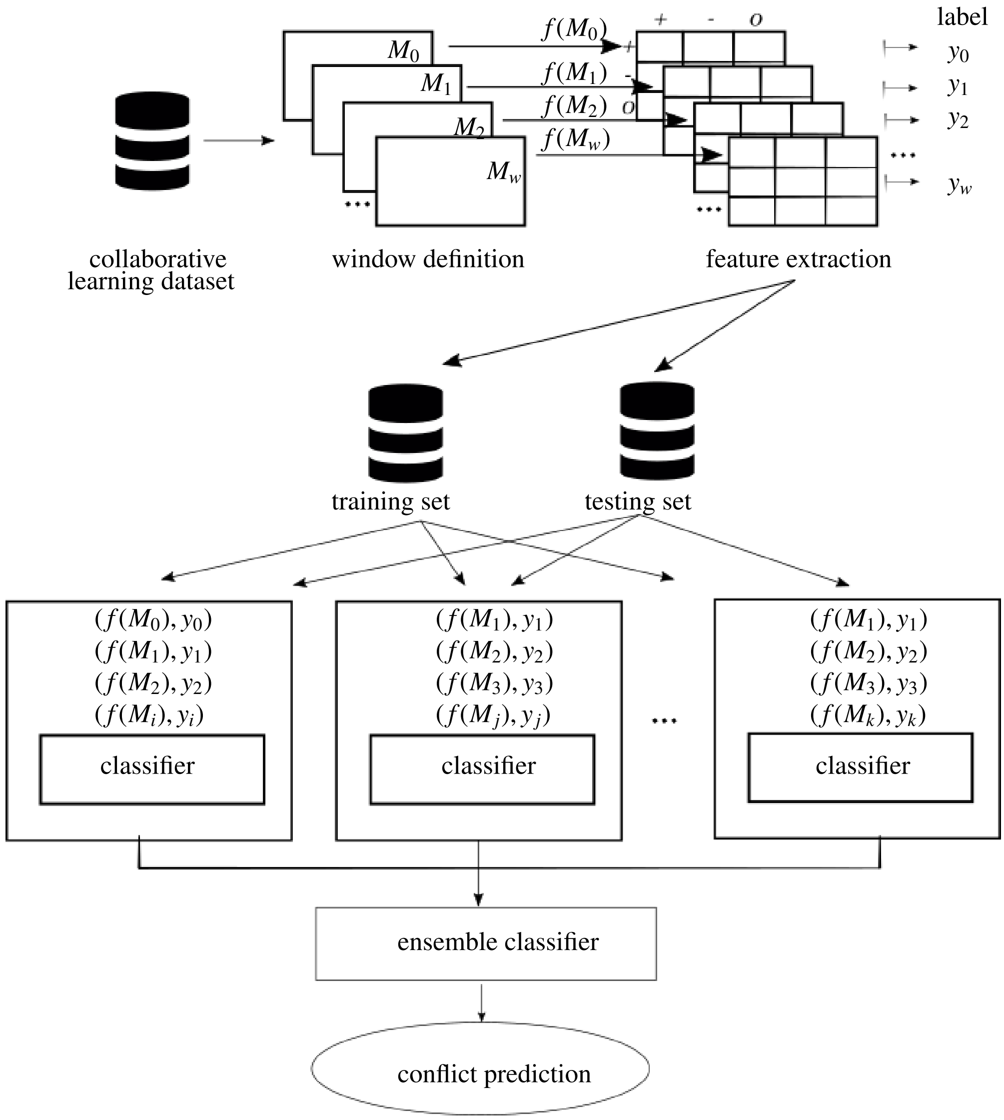

Figure 2: **The windows $M_0, \ldots, M_w$ are defined from the collaborative dataset. Then, a set of features $x_i = f(M_i)$ is estimated for each window M_i and associated to the boolean tag y_i (presence/absence of conflict). Then, these samples are split into training and testing sets. The bagging approach trains each model in the ensemble using a randomly drawn subset of the training set. The ensemble classifier combines several base models to improve the prediction**.

space. In the proposed approach, $x_i \in \mathrm{R}^9$ are the features of M, and $y_i \in \{0, 1\}$ is an indicator of whether or not exists a conflict in the window M.

As shown in Fig. 2, the bagging approach generates new training sets (bootstrap samples) through sampling with replacement, a base model is trained for each bootstrap sample. Finally, to get the ensemble classifier, the outputs of base classifiers are combined using a majority vote scheme (Li et al., 2015).

When bagging is used, it is important to carefully select the base classifiers because bootstrap samples overlap about 63.2% with the original data set (Li et al., 2015). If the learning algorithm is insensitive to the change in training data, all the base classifiers will output similar predictions; in consequence, the combination of these base classifiers cannot improve the performance of ensemble (Li et al., 2015). For this reason, for base classifiers it is recommended to select unstable learning algorithms such as neural networks, or decision trees.

5. Method

5.1. Participants

Four universities participated in this study, two from Argentine and two from Colombia. First- and fourth-year undergraduate computer-sciences degree program students participate in this study. The students were grouped into 74 groups according to different criteria (random, learning style, or students' needs). The distribution of groups according to the number of participants is: 12(16.22%) groups had 2 participants, 52(70.27%) had 3 participants, 8(10.81%) had 4 participants, and only 2(2.7%) had 5 participants. Each homework assignment was provided to the students over a week's duration.

5.2. Dataset

To support collaborative learning, the groups used Collab (Lescano & Costaguta, 2018). Collab is a web application that provides a chat for collaborative working. It enables course and student management, group creation (random, learning style, or manual), and homework assignment. Data collection was made between July 2018 and July 2019. Dialog messages from the Collab's chat were segmented into windows. For this aim, a window was defined as a set of messages with three interactions generated by the same emitter, u_i. Collaborative conflicts occur between two or more participants; hence, we only considered valid those windows with at least one message of a member $u_j \in U \setminus \{ u_i \}$. A total of 2717 windows were detected.

To analyze the presence of conflict in a window, we applied content analysis, a research technique that enables us to make valid and reproducible inferences from text to the context of its use. The content analysis was made through Krippendorf's methodology (Krippendorff, 2004). Two experts assigned a binary label that indicates the presence (or absence) of conflict to each window. For this aim, experts consider frictions, personality clashes, frustrations, and other indicators that can be associated with conflicts. The alpha reliability was 0.91.

5.3. Valence estimation

We used two approaches to estimate the valence of text messages:

- *Manual Estimation.* Each interaction in the dataset was tagged by two trained persons who assign a categorical emotion according to Pekrun's taxonomy (Pekrun, 2014). The alpha reliability for emotion labeling was 0.8. As shown in Table 1, the valence of each text message was obtained by associating emotion tag to their valence using Senticnet (Cambria et al., 2010).
- *Automatic Estimation.* We use the *pySentimiento*[1], a text valence classifier available as a library in Python. This classifier was trained with the TASS 2020 corpus. The TASS is a workshop that presents every year different challenges related to sentiment analysis in Spanish (Luque & Pérez, 2018; Martínez Cámara et al., 2018). 'pySentimiento' gives the probabilities that a sentence expresses a positive, negative and neutral sentiment.

5.4. Experiments

Experiment 1. This experiment aimed to assess the relationship between conflict and valence of text messages in collaborative learning. First, each message of the dataset was categorized as positive, negative, or neutral according the sign of its valence (estimated manually). Similarly, each simple interaction (consecutive messages between a pair of participants) of a dialog window was labeled according to the valence sign of its two components; i.e., there are $3^2 = 9$ labels for simple interactions (positive to positive, positive to neutral, positive to negative, and so on). Finally, we compared conflict and non-conflict situations in terms of the labels assigned to single messages and simple interactions.

[1] pySentimiento - https://github.com/finiteautomata/pysentimiento/

Table 1: Pekrun's emotions and their valence.

Emotion	v	Emotion	v
Anger	−0.66	Anxiety	−0.33
Compassion	0.843	Confusion	−0.76
Contempt	−0.83	Disappointment	−0.88
Dislike	−0.91	Enjoyment	0.949
Frustration	−0.33	Hope	0.892
Like	0.802	Love	0.83
Shame	−0.19	Surprise	0.66
Sympathy	0.758	Delight	0.827
Pride	0.802	Curiosity	0.911
Admiration	1.00	Envy	−0.33
Social Anxiety	−0.33	Neutral	0.0

Experiment 2. This experiment aimed to test the performance of BARCIGUS using several base classifiers, including:

- Multilayer Perceptron (MLP) (Gupta & Sinha, 2000),
- J4.8 decision tree, an implementation of the well known C4.5 algorithm (Hssina et al., 2014),
- Logistic Model Tree (LMT) (Landwehr et al., 2005),
- Reduced Error Pruning Tree (REPTree) (Mohamed et al., 2012), and
- Random Tree (Breiman, 2001).

In this experiment, 10-fold cross-validation was applied. That is, the 2717 windows were randomly divided into ten subsets of similar size. At each run, BARCIGUS was trained on nine subsets data while the remaining subset was used for testing.

The automatic labeling gives us the PMF of each dialog's sentence. To obtain the PMF for the manual case, we choose the class b^* according to the sign of the valence found in the manual labeling. Then, we set

$$p(b_i) = \begin{cases} 1 & \text{if } b_i = b^* \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

5.5. Metrics

To evaluate the performance of BARCIGUS, We used the following metrics (Verbiest et al., 2015):

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{F1-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{9}$$

where TP, FP and FN are the number of true-positive, false-positive and false-negative instances, respectively.

The Area-Under the Curve (AUC) is a common measure to assess the performance of a classifier. The most important advantage is that it is independent of the class distribution, which makes it interesting to evaluate imbalanced problems (Verbiest et al., 2015). ROC curves that are more to the northwest are better and have a bigger surface (Verbiest et al., 2015).

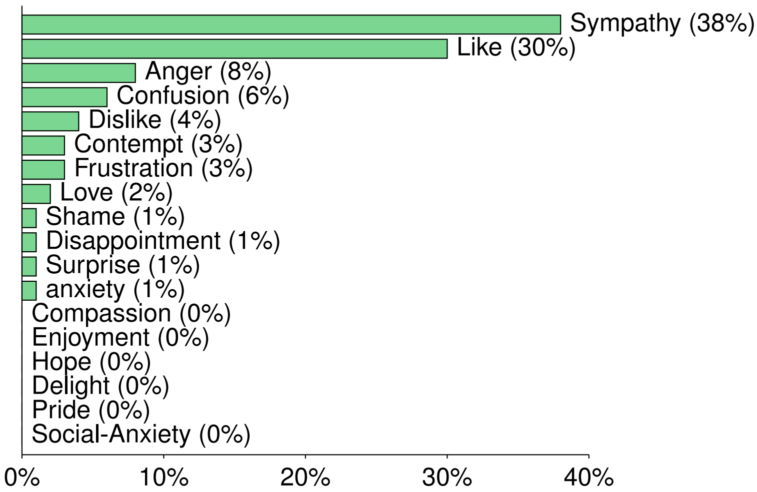

Figure 3: Distribution of non-neutral emotions in the text messages in the colab dataset.

5.6. Statistical Analysis

The SPSS statistical package was used for purposes of statistical analysis. The chi-square test was used for testing independence between two factors with nominal levels. The level of significance was set to $p = 0.05$. To evaluate the strength of association we use Cramer's V.

6. Results

As shown in Figure 3, some emotions were rarely found in manual tags; such as compassion, enjoyment, hope, delight, pride, social anxiety. By considering the manual labeling as ground truth, and using the more probable class obtained by automatic classification (positive, negative, or neutral). As shown in Table 2, the 'pySentimiento' produced a weighted average F1 score of 0.71. But, the F1 score for the positive and negative classes were much lower, 0.43 and 0.33, respectively.

Table 2: Performance of the 'pySentimiento' classifier on our dataset of chats interactions.

Class	Precision	Recall	F1	Support
+	0.59	0.34	0.43	1312
−	0.25	0.48	0.33	485
0	0.81	0.82	0.82	5368
Weighted Avg.	0.73	0.71	0.71	7165

Results of Experiment 1.

The relation between emotional category and the presence of conflict was significant, $\chi^2(2, N = 39636) = 1997.178$, $p < .01$, with a medium (Cramer's V=.2) effect size according to Cohen's conventions for Cramer's V (Cohen. J., 1988). As shown in Table 3, conflict situations have a higher relative frequency of negative messages than non-conflict situations.

A chi-square test of independence showed that there was a significant association between change of emotional category and the presence of conflict, $\chi^2(8, N = 31943) = 4616.262$, $p < .01$. The effect size for this finding was medium (Cramer's V=.38). Table 4 shows that there are more transitions to negative interactions in conflicting situations than non-conflicting situations.

Table 3: Summary of positive, negative and neutral messages in conflicting/non-conflicting situations.

valence	conflict		non-conflict	
	count	%	count	%
+	136	19.43	8718	22.39
−	333	47.57	2244	5.76
0	231	33.00	27974	71.85
Total	700	100.00	38936	100.00

Table 4: Summary of valence change of any two consecutive interactions in conflicting/non conflicting situations.

valence change	conflict		non-conflict	
	count	%	count	%
+ to +	14	2.38	2456	7.83
+ to -	71	12.05	425	1.36
+ to 0	38	6.45	5446	17.37
- to +	44	7.47	417	1.33
- to -	164	27.84	216	0.69
- to 0	101	17.15	1478	4.71
0 to +	66	11.21	3133	9.99
0 to -	29	4.92	930	2.97
0 to 0	62	10.53	16853	53.75
Total	589	100.0	31354	100.00

Table 5: Results obtained for the evaluation of conflict recognition by the bagging model built with different base classifiers and using manual estimation of valence. Number of iterations set up 10.

Base Classifier	**Precision**	**Recall**	**F1**	**ROC**
MLP	0.875	0.514	0.647	0.893
J4.8	0.855	0.596	0.703	0.889
LMT	0.886	0.642	0.745	0.942
REPTree	0.886	0.569	0.693	0.922
RandomTree	**0.939**	**0.706**	**0.806**	**0.941**

Table 6: Results obtained for the evaluation of conflict recognition by the bagging model built with different base classifiers and using an automatic classifier to measure the valence. Number of iterations set up 10.

Base Classifier	**Precision**	**Recall**	**F1**	**ROC**
MLP	0.872	0.376	0.526	0.628
J4.8	0.969	0.578	0.724	0.875
LMT	0.967	0.541	0.694	0.915
REPTree	0.931	0.495	0.647	0.897
RandomTree	**0.929**	**0.716**	**0.808**	**0.944**

Results of Experiment 2.

Results of different base classifiers for manual labeling are shown in Table 5. The best performance was obtained with the Random Forest (that uses the Random Tree classifier). Table 6 shows similar information but using 'pySentimiento' to measure the messages' valence.

7. Discussion

The experiments show differences in how emotions are manifested in students' interactions when conflicts happened. Changes to negative states are substantially increased in conflicting stages in comparison to non-conflicting stages. This result is aligned with findings reported in the literature (Jiang et al., 2013; Lee et al., 2015; Zakaria et al., 2009; Lescano et al., 2020). This happens because the emotional state of people conditions their behavior and their way of interacting (Heerdink et al., 2013; Garcia-Prieto et al., 2003). Wall & Galanes (1986) state that negative emotions expressed between group members are directly related to the number of conflicts.

Despite these differences, it is hard to detect conflict situations because the frequency of windows with conflicts is much lower than windows without conflicts. For instance, in the Collab dataset, conflicting situations only occur in 1.77% of the windows. Then, classifiers that use simple valence statistics (as frequency) to detect conflicts do not perform well; for instance, a random forest classifier with simple valence does not detect any conflict (a similar result was obtained by considering the simple value change of consecutive messages).

As the results of Experiment 2 show, the performance improves significantly by using features based on the valence of atomic interactions and RandomTree as the base classifier for the bagging. The atomic interactions reflect how two participants interact. As suggested by Millar et al. (1984), it is important the responses of persons to detect a conflict –they propose that three consecutive attempts to gaining control between two persons could be considered as a conflict pattern. In the proposed approach, valence changes give valuable information on the existence of a conflict. For instance, a negative to negative interaction could represent the first step toward an escalating, runaway spiral of negative interchanges while a negative to positive (or neutral) interaction could reflect the resolution of personal differences.

Results obtained in Experiment 2 also show an interesting effect of BARCIGUS: the performance is almost equal for manual and automatic sentiment classifiers (despite the accuracy of the automatic classifier is lower). We believe the context is transferred to a given sentence from its adjacent sentences, this effect compensates errors of sentiment classifiers. For instance, a person could effectively classify a message and its response being negative, by considering the context. For the same hypothetical case, the classifier could incorrectly classify the first negative message to be positive (to some extent negative). And it could correctly classify the negative response. The missing context problem is diminished by the joint probabilities used in the proposed feature detection.

Another advantage observed in Experiment 2 is that BARCIGUS is tolerant to errors in valence estimation. Additionally, as we consider how likely a sentence is positive, negative, and neutral, the individual differences in communication styles can be considered into account.

8. Conclusions and future work

In this study, we proposed BARCIGUS, a robust technique to recognize conflicts in chat interactions taking into account the valence change between a message and its response. Results obtained show that the technique can reach an F1 of 0.81 and it can be tolerant to errors in valence measurement. Despite this, it may be valuable to explore other strategies to improve valence estimation. One strategy is to define the normalized emotion as a function of the message valence and the writing style of the sender.

Experimental results show that it is feasible to detect task and relationship conflicts; but, the proposed approach does not differentiate between them. Once the conflict is detected, teachers entail the responsibility for conflict management. Then, teachers require pedagogical training, resources, and knowledge to intervene. Developing an intelligent agent that promotes conflict resolution competencies in learners is another research line of high interest. Knowing the type of conflict, such an agent can suggest the best action plan.

Detecting conflicts in CSCL environments is a useful and necessary tool because distance learning has become increasingly important, mainly because of the lockdown on the COVID-19 epidemic. The proposed approach does not impose a rigid format to detect conflicts in CSCL environments. Instead, students can express themselves freely.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT author statement

Germán Lescano: Conceptualization, Methodology, Software, Writing-Original draft preparation. Jose Torres-Jimenez: Formal analysis, validation. Rosanna Costaguta: Resources, Supervision. Analía Amandi: Project administration. Carlos Lara: Methodology, Data curation, Writing - Review & Editing.

References

Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems*, *108*, 110–124.

Ayoko, O. B., Callan, V. J., & Härtel, C. E. J. (2008). Group Research. *Small Group Research*, *39*, 121–149.

Bales, R. (1950). *Interaction Process Analysis: A Method for the Study of Small Groups*. Cambridge, Mass. Addison-Wesley Press.

Bales, R., & Steven, C. (1979). *SYMLOG: A System for the Multiple Level Observation of Groups*. New York: Free.

Barsade, S. G. (2009). The Ripple Effect : Emotional Contagion and Its Influence on Group Behavior. *Administrative Science Quarterly*, *47*, 644–675.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Buchs, C., Butera, F., Mugny, G., & Darnon, C. (2004). Conflict Elaboration and Cognitive Outcomes. *Therory Into Practice*, *43*, 23–30.

Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010). Senticnet: A publicly available semantic resource for opinion mining. In *AAAI fall symposium: commonsense knowledge*. Citeseer volume 10.

Capdeferro, N., & Romero, M. (2012). Are Online Learners Frustrated with Collaborative Learning Experiences? *The International Review of Research in Open and Distance Learning*, *13*.

Cohen. J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Coppin, G., & Sander, D. (2013). Contemporary Theories and Concepts in the Psychology of Emotions. In C. Pelachaud (Ed.), *Emotion-Oriented Systems* (pp. 3–32). New York, USA: John Wiley & Sons, Inc.

Dreu, C. K. W. D., & Weingart, L. R. (2003). Task Versus Relationship Conflict, Team Performance, and Team Member Satisfaction: A Meta-Analysis. *Journal of Applied Psychology*, *88*, 741–749.

Feidakis, M., Daradoumis, T., Caballé, S., & Conesa, J. (2014). Embedding Emotion Awareness into E-Learning Environments. *International Journal of Emerging Technologies in Learning (IJET)*, *9*, 39–46.

Garcia-Prieto, P., Bellard, E., & Schneider, S. C. (2003). Experiencing Diversity, Conflict, and Emotions in Teams. *Applied Psychology: An International Review*, *52*, 413–440.

Garcia-Prieto, P., Mackie, D. M., Tran, V., & Smith, E. R. (2007). Chapter 7 Intergroup Emotions in Workgroups: Some Emotional Antecedents and Consequences of Belonging. *Research on Managing Groups and Teams*, *10*, 145–184.

Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. C. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications*, *69*, 214–224.

Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education*, .

Gupta, P., & Sinha, N. K. (2000). *Neural Networks for Identification of Nonlinear Systems: An Overview*. Academic Press. URL: `http://dx.doiXX.org/10.1016/B978-0-12-646490-0.50017-2`.

Heerdink, M. W., Kleef, G. A. V., Homan, A. C., & Fischer, A. H. (2013). On the Social Influence of Emotions in Groups : Interpersonal Effects of Anger and Happiness on Conformity Versus Deviance. *Journal of Personality and Social Psychology*, *105*, 262–284.

Hssina, B., Merbouha, A., Ezzikouri, H., & Erritali, M. (2014). A comparative study of decision tree ID3 and C4.5. *International Journal of Advanced Computer Science and Applications*, *4*, 13–19.

Järvelä, S., & Hadwin, A. F. (2013). New Frontiers: Regulating Learning in CSCL. *Educational Psychologist*, *48*, 25–39.

Järvenoja, H., & Järvelä, S. (2013). Regulating Emotions Together for Motivated Collaboration. In *Affective Learning Together. Social and Emotional Dimensions of Collaborative Learning.* chapter 8. (pp. 162–181). New York, United States: Routledge.

Jehn, K. A. (1997). A Qualitative Analysis of Conflict Types and Dimensions in Organizational Groups. *Administrative Science Quarterly*, *42*, 530–557.

Jehn, K. A., & Mannix, E. A. (2001). The Dynamic Nature of Conflict: A Longitudinal Study of Intragroup Conflict and Group Performance. *Academy of Management Journal*, *44*, 238–251.

Jiang, J. Y., Zhang, X., & Tjosvold, D. (2013). Emotion Regulation as a Boundary Condition of the Relationship Between Team Conflict and Performance: A Multi-level Examination. *Journal of Organizational Behavior*, *34*, 714–734.

Keshavarz, H., & Abadeh, M. S. (2017). ALGA: Adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs. *Knowledge-Based Systems*, *122*, 1–16.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments : a review of the research. *Computers in Human Behavior*, *19*, 335–353.

Krippendorff, K. (2004). *Content Analysis: An Introduction to its Methodology*. USA: SAGE Publications.

Kwon, K., Liu, Y. H., & Johnson, L. P. (2014). Group regulation and social-emotional interactions observed in computer supported collaborative learning: Comparison between good vs. poor collaborators. *Computers and Education*, *78*, 185–200.

Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning*, *59*, 161–205.

Lee, D., Huh, Y., & Reigeluth, C. M. (2015). Collaboration, intragroup conflict, and social skills in project-based learning. *Instructional Science*, *43*, 561–590.

Lescano, G., & Costaguta, R. (2018). COLLAB: Conflicts and Sentiments in chats. In *XIX International Conference on Human Computer Interaction (Interacción 2018).*

Lescano, G., Lara, C., Collazos, C., & Costaguta, R. (2020). A Technique for Conflict Detection in Collaborative Learning Environment by Using Text Sentiment. In L. Martínez-Villaseñor, O. Herrera Alcántara, H. Ponce, & F. Castro-Espinoza (Eds.), *Advances in Computational Intelligence* (pp. 39–50). Cham, Switzerland: Springer Nature Switzerland AG volume 12469 (II). (Lncs (lnai ed.).

Li, Y., Gao, J., Li, Q., & Fan, W. (2015). Ensemble Learning. In C. Aggarwal (Ed.), *Data Classification. Algorithms and Applications* (p. 483). New York, USA: Chapman & Hall/CRC.

Luque, F. M., & Pérez, J. M. (2018). Atalaya at TASS 2018: Sentiment analysis with tweet embeddings and data augmentation. In E. Martínez Cámara, Y. Almeida-Cruz, M. C. Díaz Galiano, S. Estévez-Velarde, M. Á. García

Cumbreras, M. García Vega, Y. Gutiérrez, A. Montejo Ráez, A. Montoyo, R. Muñoz, A. Piad-Morffis, & J. Villena Román (Eds.), *Proceedings of TASS 2018: Workshop on Sentiment Analysis at SEPLN* (pp. 29–35). Sevilla, España: CEUR-WS volume 2172.

Madhoushi, Z., Hamdan, A. R., & Zainudin, S. (2015). Sentiment analysis techniques in recent works. In *Science and Information Conference, SAI 2015* (pp. 288–291). doi:10.1109/SAI.2015.7237157.

Martínez Cámara, E., Almeida-Cruz, Y., Díaz Galiano, M. C., Estévez-Velarde, S., García Cumbreras, M. Á., García Vega, M., Gutiérrez, Y., Montejo Ráez, A., Montoyo, A., Muñoz, R., Piad-Morffis, A., & Villena Román, J. (2018). Overview of TASS 2018: Opinions, health and emotions. In E. Martínez Cámara, Y. Almeida-Cruz, M. C. Díaz Galiano, S. Estévez-Velarde, M. Á. García Cumbreras, M. García Vega, Y. Gutiérrez, A. Montejo Ráez, A. Montoyo, & R. Muñoz (Eds.), *Proceedings of TASS 2018: Workshop on Semantic Analysis at SEPLN (TASS 2018)*. Sevilla, España: CEUR-WS.

Millar, F. E., Rogers, L. E., & Bavelas, J. B. (1984). Identifying patterns of verbal conflict in interpersonal dynamics. *Western Journal of Speech Communication*, *48*, 231–246.

Mohamed, W. N. H. W., Salleh, M. N. M., & Omar, A. H. (2012). A comparative study of Reduced Error Pruning method in decision tree algorithms. *Proceedings - 2012 IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2012*, (pp. 392–397).

Monteserin, A., Schiaffino, S., & Amandi, A. (2010). Assisting students with argumentation plans when solving problems in CSCL. *Computers and Education*, *54*, 416–426.

Nam, C. W. (2014). The effects of trust and constructive controversy on student achievement and attitude in online cooperative learning environments. *Computers in Human Behavior*, *37*, 237–248.

Näykki, P., Järvelä, S., Kirschner, P. A., & Järvenoja, H. (2014). Socio-emotional conflict in collaborative learning-A process-oriented case study in a higher education context. *International Journal of Educational Research*, *68*, 1–14.

Onrubia, J., & Engel, A. (2012). The role of teacher assistance on the effects of a macro-script in collaborative writing tasks. *International Journal of Computer-Supported Collaborative Learning*, *7*, 161–186.

Orvis, K. L., & Lassiter, A. L. (2006). Computer-supported collaborative learning: The role of the instuctor. In *Teaching and learning with virtual teams* (pp. 158–179). IGI Global.

Pekrun, R. (2014). *Emotions and Learning. Practices Series-24*. International Academy of Education & International Bureau of Education.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*, 36–48.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. (2002). Academic emotions in students' self-regulated learning and achievement: a program of qualitative and quantitative research. *Educational Psychologist*, *37*, 91–105.

Poria, S., Cambria, E., Winterstein, G., & Huang, G. B. (2014). Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems*, *69*, 45–63.

Rao, A. S., Georgeff, M. P. et al. (1995). Bdi agents: from theory to practice. In *Icmas* (pp. 312–319). volume 95.

Rapisarda, B. A. (2002). *The impact of emotional intelligence on work team cohesiveness and performance*. Ph.D. thesis Case Western Reserve University.

Reis, R. C. D., Isotani, S., Rodriguez, C. L., Lyra, K. T., Jaques, P. A., & Bittencourt, I. I. (2018). Affective states in computer-supported collaborative learning: Studying the past to drive the future. *Computers and Education*, *120*, 29–50. doi:10.1016/j.compedu.2018.01.015.

Rogat, T. K., & Adams-Wiggins, K. R. (2015). Interrelation between regulatory and socioemotional processes within collaborative groups characterized by facilitative and directive other-regulation. *Computers in Human Behavior*, *52*, 589–600. URL: `http://dx.doiXX.org/10.1016/j.chb.2015.01.026`.

Sanglerdsinlapachai, N., Plangprasopchok, A., & Nantajeewarawat, E. (2014). Exploiting rhetorical structures to improve feature-based sentiment analysis. In *International Computer Science and Engineering Conference, ICSEC 2014* (pp. 180–185).

Schwarz, B. B., & Asterhan, C. S. (2011). E-moderation of synchronous discussions in educational settings: A nascent practice. *Journal of the Learning Sciences*, *20*, 395–442.

Slof, B., Nijdam, D., & Janssen, J. (2016). Do interpersonal skills and interpersonal perceptions predict student learning in CSCL-environments? *Computers and Education*, *97*, 49–60.

Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education (IJAIED)*, *12*, 40–62.

Stahl, G., & Hakkarainen, K. (2020). Theories of CSCL. *International handbook of computer-supported collaborative learning. New York, NY: Springer*, .

Tchounikine, P., Rummel, N., & McLaren, B. M. (2010). Computer supported collaborative learning and intelligent tutoring systems. In *Advances in intelligent tutoring systems* (pp. 447–463). Springer.

Tedesco, P. A. (2003). MArCo: Building an artificial conflict mediator to support group planning interactions. *International Journal of Artificial Intelligence in Education*, *13*, 117–155.

Verbiest, N., Vermeulen, K., & Teredesai, A. (2015). Evaluation of Classification Methods. In C. Aggarwal (Ed.), *Data Classification. Algorithms and Applications* (pp. 633–656). New York, USA: Chapman & Hall/CRC.

Wall, V. D., & Galanes, G. J. (1986). The SYMLOG dimensions and small group conflict. *Central States Speech Journal*, *37*, 61–78.

Wang, Z., Tong, V. J., Ruan, P., & Li, F. (2016). Lexicon Knowledge Extraction Algorithm with Sentiment Polarity Calculation. In *IEEE 16th International Conference on Data Mining Workshops* (pp. 978–983). doi:10.1109/ICDMW.2016.37.

Zachary, W. W. (1977). An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, *33*, 452–473.

Zakaria, C., Curé, O., & Smaïli, K. (2009). Conflict Ontology Enrichment Based on Triggers. In *The 2nd International workshop on Ontologies and Information Systems for the Semanti*. Napa Valley, California, United States.

Zheng, L., & Huang, R. (2016). The effects of sentiments and co-regulation on group performance in computer supported collaborative learning. *Internet and Higher Education*, *28*, 59–67.