Emotion Recognition for Education using Sentiment Analysis

María Lucia Barron-Estrada, Ramón Zatarain-Cabada, Raúl Oramas-Bustillos

Tecnológico Nacional de México / Instituto Tecnológico de Culiacán,
Culiacán, Sinaloa, Mexico
{lbarron, rzatarain}@itculiacan.edu.mx

Abstract. Most systems that recognize emotions that are used in learning systems are oriented towards the detection of basic emotions such as happy, sad or angry. However, an intelligent tutor system must be able to detect, in a student, secondary emotions that have to do with learning as boring or frustrated. In the case of sentiment analysis in text, most of the research focuses on recognizing polarities (positive and negative). In this paper we present a sentiment analyzer for the recognition of polarities and emotions centered on learning using textual phrases written in Spanish in the field of computer science that can be used in an ITS (or any learning environment) to detect sentiments and/or emotions of the students. We also present the results obtained from the sentiment analyzer, which are very promising with 88.26% of accuracy.

Keywords: sentiment analysis, machine learning, deep learning, intelligent learning environments, learning centered emotions.

1 Introduction

The application of the AI in educational environments (intelligent environments) aims to create intelligent teaching software that adapts to the pace of student learning. The objective is always the same, to create a scenario that allows the student to be immersed in an educational environment to help the student learn.

A learning environment is a setting, traditionally a classroom, where students learn [21]. Although there are diverse virtual learning environments, such as intelligent learning environments (ILE) or intelligent tutoring systems (ITS), the main objective of a learning system is to facilitate different learning activities [5]. In this area of research related to intelligent learning systems, various fields are combined such as pedagogy, psychology, cognitive sciences, artificial intelligence, among others, where each of them contributes their vision to the development of the discipline [23]. In the case of modern ILE and ITS, they already incorporate the automatic recognition of emotions [2,6] through the human face, brain signals, and voice. In the case of text recognition or sentiment analysis (SA), most of the work focus on recognizing polarities [13,25] and a few of them recognize basic emotions [19,22].

This work presents a system that carries out a sentimental analysis based on polarities and learning-centered emotions. The emotions that are detected are boring,

frustration, neutral, excitement, and engagement. The recognition of emotions can be used in any ITS or ILE to provide personalized feedback to the student after solving an exercise and to adapt the exercises of the learning system based on the emotions detected.

The main contribution of this work is to use emotional recognition techniques, based on text, and their integration in intelligent learning environments such as an ITS or an ILE and combining traditional algorithms of emotion classification with current deep learning (DL) techniques. Another contribution is the design, construction, and testing of a corpus of texts labeled with emotions centered on learning and a corpus labeled with polarity, both datasets in the domain of computer science.

This paper is organized as follows: Section 2 describes the related work. Section 3 presents the methodology to build the sentiment analyzer. Section 4 shows the results and discussions. Finally, in Section 5 the conclusions and future work are described.

2 Related Work

This section presents a review of sentiment analysis with deep learning techniques and sentiment analysis in learning environments.

2.1 Sentiment Analysis Using Deep Learning

Deep learning has revolutionized the way of solving problems that were previously done with traditional Machine Learning (ML) techniques. For example, in [9] the authors describe a Dynamic Convolutional Neural Network (DCNN) architecture for sentiment modeling of sentences. The network supports sentences of variable length and is able to capture relationships between words. The authors conducted experiments with Twitter phrases with binary and multiclass prediction.

In [11] the author uses a Convolutional Neural Network (CNN) with a convolution layer and pre-trained word vectors over 100 billion words from Google News for sentence-level classification tasks. His model, despite being simple, achieved excellent results for different corpus used. The researchers also concluded that pre-trained vectors are extractors of universal characteristics that can be used for various classification tasks.

In [20] the authors propose a CNN-LSTM model consisting of two parts: a regional CNN and an Long Short Term Memory (LSTM), to predict the qualifications of texts with a valence scale. The regional CNN unlike a conventional CNN that considers a complete text as input uses individual sentences as regions so that it can be weighted according to the valence scale. Finally, this regional information is sequentially integrated into all regions using LSTM for prediction with valence scale. The experimental results show that the proposed method outperforms the lexical, regression-based and ANN-based methods proposed in previous studies.

In [18] they propose simple, trained models with annotations where each sentence is annotated with two classes as negative, positive at the sentence level, and they also model the linguistic role of the lexical feelings with LSTM, the words of denial and the words of intensity. The results show that the models are capable of capturing the

linguistic role of sentimental words, words of denial and words of intensity in sentimental expression.

In [15] they used a lexical-based approach to apply semantic rules and then used a DeepCNN with pre-trained vectors at the character level to capture morphological information of each word to determine how those words are formed and their relationship to others. They then used a Bi-LSTM bidirectional network that produces the representation of the characteristics of the whole sentence from pre-trained vectors at the word level. For the training of the model, 100 billion words of Google News called Word2Vec [14] and TwitterGlove [17] of Stanford University were used. In this work, they report an accuracy of 86.63 %.

2.2   Sentiment Analysis in Learning Environments

In [10] a research paper is presented that proposes to take into account the opinions of the users in order to elaborate an evaluation in this respect using hybrid learning methods based on SVM and Hidden Markov Model (HMM) in an e-learning environment. The authors indicate that sentiment analysis becomes more difficult and challenging when it is done for e-learning blogs.

In [8] a model of sentiment analysis is presented to extract the opinions of the students to evaluate the quality of a course in two steps. In the classification of opinion, machine learning methods have been applied to classify an opinion as positive or negative for the publications of each student. Then, the extraction of opinion is used to extract characteristics, such as teachers, exams and resources, from the content generated by the user for a specific course.

In SA-E [1] the authors analyze the feedback of students who take a class in real time via Twitter messages with two machine learning algorithms: Naïve Bayes and SVM to identify the positive or negative sentiments of the students in real time. The objective was to collect the comments and analyze these data to help improve teaching.

SentBuk [16] is an application that runs on the social network Facebook that extracts information about the user's sentiment. It uses the opinions that students write and classifies them as positive, negative or neutral to detect significant emotional changes. This information is used to enrich the SentBuk system, so that they adopt the tasks that will be proposed to the student, based on their emotional state. These emotions are also used as useful feedback for the teacher of the course. The classification method implemented in SentBuk follows a hybrid approach: it combines lexical and traditional machine learning techniques.

In [4] present an algorithm based on an ANN to predict the student's dropout in MOOC courses using the analysis of sentiment and show the importance of the sentiments of students in this task.

In [12] student opinions are collected in the form of running text and sentiment analysis is performed to identify important aspects together with the orientations using supervised and semi-supervised ML techniques.

In the work of [7], they use sentiment analysis with supervised machine learning algorithms of type SVM, Naïve Bayes, nearest neighbors (KNN) and ANN to find the polarity of the comments of students based on the predefined characteristics of teaching and learning. The study involves the application of a combination of machine learning and natural language processing (NLP) techniques in student feedback data. The results

were compared to find the best performance with respect to several evaluation criteria for the different algorithms.

In [19] they explore NLP and machine learning techniques based on student comments to help university administrators and teachers address problem areas in teaching and learning. The system they developed analyzes student comments in online course surveys and real sources to identify the polarity of feeling, expressed emotions and student satisfaction. They also made a comparison with the results of the direct evaluation to determine the reliability of the system.

3 Methodology for Building and Integration of the Sentiment Analyzer

The proposed methodology for the development of our work is based on the inductive scientific method that is organized into four stages: analysis, design, implementation, and testing. The monitoring of each stage will allow reaching the solution of the development of the system that performs the analysis of feelings.
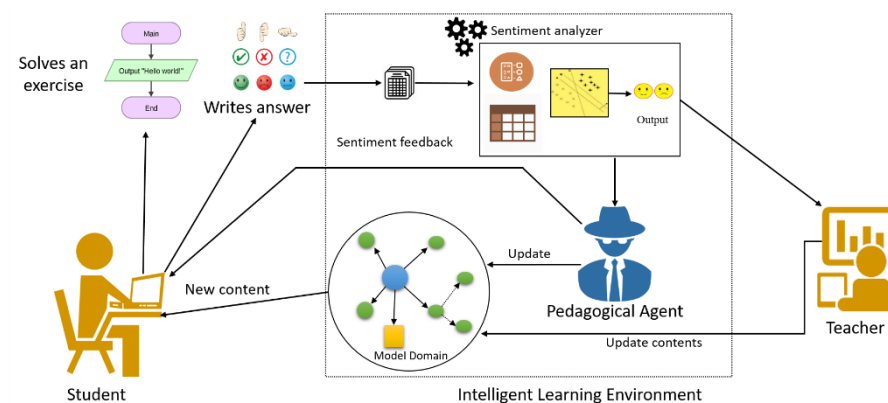

Fig. 1. The general scheme of the system to perform sentiment analysis with application to ILE and ITS.

3.1 Work Analysis and Design

In the analysis stage, the state of the art review was carried out to know which approaches have been used to deal with the problems related to sentiment analysis. A review and analysis of frameworks that are used in sentiment analysis were also carried out. In relation to the design stage, Fig. 1 illustrates the general scheme of the system to perform sentiment analysis in an ILE or ITS for a computer science domain. From the previous diagram, the student solves an exercise in the area of computer science provided by the ILE or ITS by means of responding with texts to the questions. These responses are the input to the SA system (sentiment analyzer) to perform the detection of emotion. For the detection of emotions, an DL model with its own corpus is used. A pedagogical agent uses the detected emotion to perform two activities: first, to send a

personalized feedback to the student based on the emotion detected; second, to present the contents of the student's domain model based on those emotions.

To design the best system architecture to perform SA, we first need to find a model that is suitable for the detection of learning-centered emotions. To find the best model, different tests were performed with different architectures. Fig. 2 shows the design for a DL model, which is a CNN-type network in combination with another LSTM network.
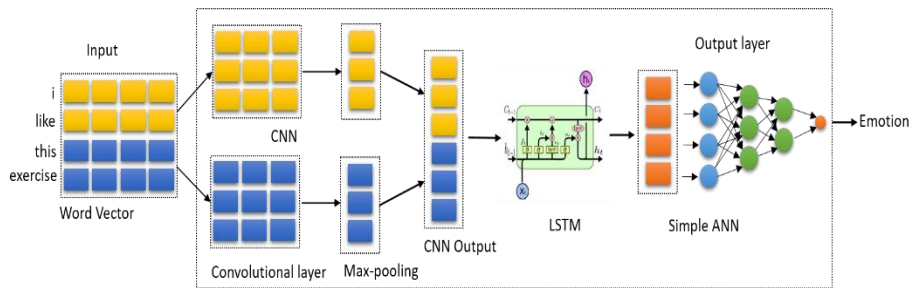

Fig. 2. CNN and LSTM Model.

The model receives as input a vector of words in a format called *one hot encoding* that encodes a text to a numerical format for the DL model to perform its prediction work. Subsequently, a CNN network takes this input vector and divides the information into smaller parts using a convoluted layer to detect patterns or characteristics, and combining this information in the deepest layers of the network. The output of the CNN network represents a feature vector that serves as an input to an LSTM network that processes that information in search of new patterns to pass that information to a simple neural network that is in the output layer and that classifies the text of entry into an emotion.

Once you have the best DL model, the next step is to establish the architecture of the model to perform sentiment analysis (Fig. 3). The emotion recognizer does the following: given an input text, the polarity of the text is first determined using the DL model presented in Fig. 2. Depending on the result of the positive or negative classification, the same model is used to determine whether the Text belongs to a category as engaged or excited or either frustrated or boring.
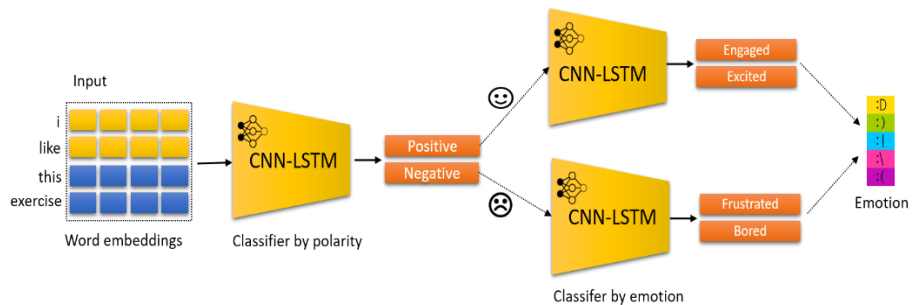

Fig. 3. SA Architecture for emotion recognizing.

María Lucia Barron-Estrada, Ramón Zatarain-Cabada, Raúl Oramas-Bustillos

3.2 Implementation and Integration with the ILE Java Sensei

Java Sensei [3, 24] is an ILE to teach programming in the Java programming language. The goal of the ILE is to analyze the cognitive and affective states of the student, to personalize their learning rhythm depending on those states. Once the best DL model was obtained to perform SA, the next step is the integration of the module to Java Sensei. The ILE Java Sensei uses a layered design to organize all its modules and components. This design allows the system to be scalable and easy to maintain. Information flows between layers, modules, and components through interfaces. Fig 4. Illustrates the architecture of Java Sensei.

The tutor's layer contains four modules: The ITS module that coordinates the entire tutoring system; the domain module contains the knowledge of the expert; the recommendation module is responsible for adapting the content presented to the student based on the evaluation grades based in Likert-scale provided by the students for each educational resource that they use, and the student module stores all the cognitive and emotional information of the student.

The affective layer contains the affective part of the system. It contains a module for the recognition of facial expressions as well as a module to recognize educational emotions based on the answers of the students' exercises, which is presented in this paper. The fuzzy logic module implements the mechanism of inference of the system based on the emotions detected by the facial expression recognizer and the sentiment analyzer.

Other pedagogical aspects such as the quality of the answers are also taken into account to feed these fuzzy rules. A pedagogical agent uses this information. The web content layer contains the corpus of sentimental text that is used to train the SA module, a repository of exercises and educational resources as well as the rules of the fuzzy logic system. Finally, the data layer stores the information produced by the students.
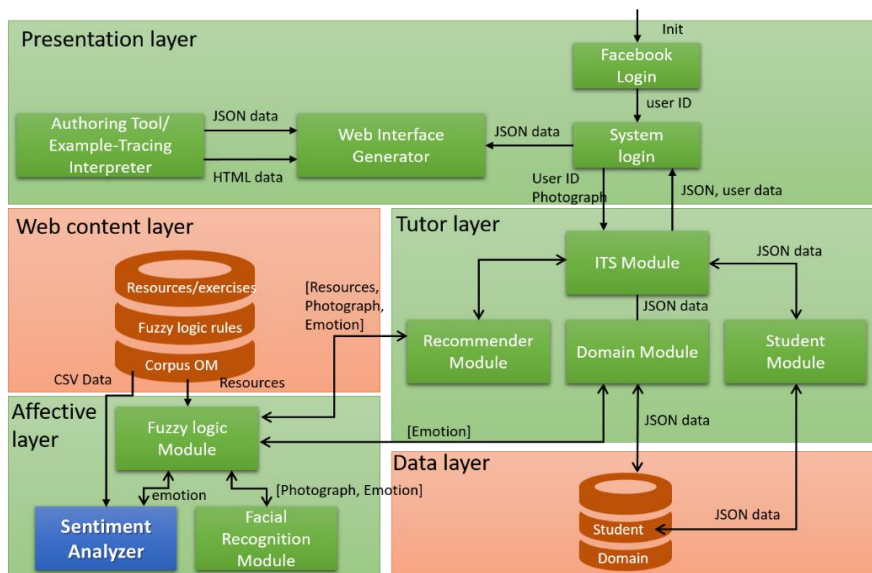

Fig. 4. Java Sensei Architecture with SA integration.

4 Discussion and Future Work

In this section, we present the results obtained from the classification of two datasets: one containing texts labeled with polarities and one containing text labeled with learning-centered emotions. There are a series of metrics that are used to estimate the quality of the classification algorithm that is part of our learning model. The simplest method to calculate the effectiveness of a classifier is the accuracy measure that calculates the average of text documents correctly classified over the total of documents to be classified. To obtain this value we applied a cross-validation technique with a dataset of 15400 texts in Spanish with 90% for the training data and 10% for the data test. Traditional ML classifier algorithms we used for the evaluation were Bernoulli Naïve Bayes, Multinomial Naïve Bayes, Support Vector Machine, Linear Support Vector Machine, Stochastic Gradient Descent, and K-Nearest Neighbors (KNN). For deep learning (see Fig. 3), we created a model based on combining CNN and LSTM neural networks. Table 1 shows the accuracy values obtained according to the size of the corpus. The traditional ML classifier with the highest score was Bernoulli Naïve Bayes, with an accuracy of 76.77%. In general, Bernoulli Naïve Bayes is a robust classifier against irrelevant features and it is suitable for text classifying. On the other hand, the deep learning model reached a precision of 88.26%, surpassing the other classifiers.

Table 1. Accuracy values obtained using different classifiers.

Classifier	Accuracy
Bernoulli NB	76.77%
Multinomial NB	75.31%
SVC	75.79%
Linear SVC	74.69%
SGDC Classifier	76.69%
KNN	68.46%
CNN + LSTM*	88.26%

* Deep learning model.

In a first approach, the system has been tested with 43 students in our research lab. Students must solve several exercises to learn and practice the topics of Java language. Each lesson has 15 exercises. At the end of each exercise, the system asks for student answers. Table 2, shows some examples of recollected answers.

Table 2. Collected opinions with polarity evaluation.

Opinion (Spanish)	Opinion (English)	Evaluation
No entendí este ejercicio	I did not understand this exercise	Negative
Me gusto este ejercicio	I liked this exercise	Positive
No tengo mucho conocimiento sobre este tema	I do not have much knowledge about this topic	Positive
El ejercicio es confuso	This exercise is confusing	Negative
Este ejercicio es sencillo de entender	This exercise is easy to understand	Positive

We recollected 178 student texts. The SA module evaluated 71 texts as positive and 107 as negative. We carried out another experiment with 53 computer-engineering students (45 men and 8 women) at the Instituto Tecnológico de Culiacán. We applied a learning system that covers the topics of Programming Fundamentals. After studying each sub-theme in text and visual or multimedia (video) format, each student registered her/his opinion/answer with a text not greater than 255 characters with respect to the content of the studied subject. Besides that, they labeled his opinion with an emoticon that represented his emotional state: bored, frustrated, neutral, excited or engaged. The total of recorded opinions was 851. Table 3 shows some examples of the opinions recorded by the students who participated in the experiment.

Table 3. Collected opinions with emotion evaluation.

| Opinion (Spanish)	Opinion (English)	Evaluation
Me gusto bastante el vídeo.	I really liked the video.	Engaged
No me gusto que las voces fueran de España, además los gráficos de la animación están algo feos.	I did not like that the voices were from Spain, besides the graphics of the animation are somewhat ugly.	Frustrated
El video es bueno aunque creo que le falto profundizar más.	The video is good although I think it needs to go deeper.	Neutral
Vaya, es algo complejo.	Wow, it's something complex.	Bored
Quizás con un ejemplo quedaría más claro.	Perhaps with an example, it would be clearer.	Excited
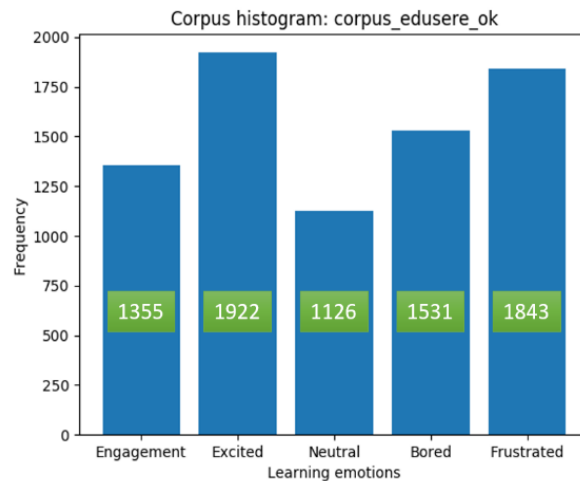
Fig. 5. Distribution of answers (opinions) in learning-centered emotions.

From 851 opinions, we found a tendency to neutral, excited, and engaged emotions, which indicates that new resources must be included that induce students to write

negative comments. Figure 5 shows the distribution of the opinions collected in the first experiment. We trained the previously used classifiers with this new corpus focused on learning, using cross-validation of 90% for training data, and 10% for test data. The algorithm with better precision was Linear SVC with 60%. The level of prediction is low because the size of the corpus is still small and it is not balanced as can be observed in Figure 5.

5 Conclusions and Future Work

Detecting emotions is a complex task even for human beings. Using only text or dialogs to detect emotions is a very hard challenge so that this research needs to be continued. This work presents the building of a sentiment analyzer for detecting emotions related to an educational context. The tests validate that this work had successful results in two type of database: one corpus with texts labeled with polarities and one corpus with texts labeled with learning-centered emotions. The results indicate that convolutional neural networks are a valuable choice in the recognition of polarities. We need more work with respect to classifying secondary emotions. The CNN architecture presented is a good beginning to recognize non-basic emotions and can be applied to recognize emotions from other different sources (e.g. facial expressions). In the future, we want to increase and balance the size of the two databases and we want to test them with more different machine learning methods to increase the number of classifiers. In addition, we will add different types of recognition from other sources such as voice, facial expressions, and brain signals.

References

1. Altrabsheh, N. *et al.*: SA-E: sentiment analysis for education. In: International Conference on Intelligent Decision Technologies. pp. 353–362 (2013)
2. Baker, R.Sj. *et al.*: Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. Int. J. Hum. Comput. Stud. 68, 4, 223–241 (2010)
3. Barrón-Estrada, M.L. *et al.*: Sentiment Analysis in an Affective Intelligent Tutoring System. In: Chang, M. *et al.* (eds.) 17th IEEE International Conference on Advanced Learning Technologies, (ICALT 2017). Timisoara, Romania, pp. 394–397 (2017)
4. Chaplot, D.S. *et al.*: Predicting Student Attrition in MOOCs using Sentiment Analysis and Neural Networks. In: AIED Workshops (2015)
5. Coen, M.H. *et al.*: Design principles for intelligent environments. In: AAAI/IAAI, pp. 547–554 (1998)
6. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learn. Instr. 22, 2, pp. 145–157 (2012)
7. Dhanalakshmi, V. *et al.*: Opinion mining from student feedback data using supervised learning algorithms. In: Big Data and Smart City (ICBDSC), 2016 3rd MEC International Conference on. pp. 1–5 (2016)
8. El-Halees, A.: Mining opinions in user-generated contents to improve course evaluation. In: International Conference on Software Engineering and Computer Systems, pp. 107–115 (2011)
9. Kalchbrenner, N. *et al.*: A Convolutional Neural Network for Modelling Sentences. In: Proc. 52nd Annu. Meet. Assoc. Comput. Linguist. (Volume 1 Long Pap. pp. 655–665

(2014)

10. Kechaou, Z. *et al.*: Improving e-learning with sentiment analysis of users' opinions. In: Global Engineering Education Conference (EDUCON), 2011 IEEE. pp. 1032–1038 (2011)

11. Kim, Y.: Convolutional Neural Networks for Sentence Classification (2014)

12. Kumar, A., Jain, R.: Sentiment analysis and feedback evaluation. In: MOOCs, Innovation and Technology in Education (MITE), 2015 IEEE 3rd International Conference on. pp. 433–436 (2015)

13. Medhat, W. *et al.*: Sentiment analysis algorithms and applications: A survey. Ain Shams Eng. J. 5, 4, pp. 1093–1113 (2014)

14. Mikolov, T. *et al.*: word2vec. Google Sch (2014)

15. Nguyen, H., Nguyen, M.-L.: A Deep Neural Architecture for Sentence-level Sentiment Classification in Twitter Social Networking (2017)

16. Ortigosa, A. *et al.*: Sentiment analysis in Facebook and its application to e-learning. Comput. Human Behav. 31, pp. 527–541 (2014)

17. Pennington, J. *et al.*: Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 1532–1543 (2014)

18. Qian, Q. *et al.*: Linguistically Regularized LSTMs for Sentiment Classification (2016)

19. Rani, S., Kumar, P.: A sentiment analysis system to improve teaching and learning. Computer (Long. Beach. Calif). 50, 5, pp. 36–43 (2017)

20. Wang, J. *et al.*: Dimensional Sentiment Analysis Using a Regional CNN-LSTM Model. In: Proc. 54th Annu. Meet. Assoc. Comput. Linguist. (Volume 2 Short Pap. January, pp. 225–230 (2016)

21. Warger, T., Dobbin, G.: Learning environments: where space, technology, and culture converge. EDUCAUSE Learning Initiative (2009)

22. Wen, M. *et al.*: Sentiment Analysis in MOOC Discussion Forums: What does it tell us? In: Proc. Educ. Data Min. (EDM), pp. 1–8 (2014)

23. Woolf, B.P.: Building Intelligent Interactive Tutors Student-centered strategies for revolutionizing e-learning. Morgan Kaufmann (2010)

24. Zatarain-Cabada, R. *et al.*: An affective and Web 3.0-based learning environment for a programming language. Telemat. Informatics. 35, 3, pp. 611–628 (2018)

25. Zhang, L., Liu, B.: Sentiment Analysis and Opinion Mining. In: Encycl. Mach. Learn. Data Min. May, pp. 1–10 (2016)