Filomat 32:5 (2018), 1823–1829
https://doi.org/10.2298/FIL1805823T

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Classification of video lecture learners' cognitive and negative emotional states using a Bayesian belief network

Xiaomei Tao[a,b,*], Qinzhou Niu[a,b], Mike Jackson[c], Martyn Ratcliffe[c]

[a]*Guangxi Key Laboratory of Embedded Technology and Intelligent Systems, Guilin, China*
[b]*School of Information Science and Engineering, Guilin University of Technology, Guilin, China*
[c]*Faculty of Computing, Engineering and the Built Environment, Birmingham City University, Birmingham, UK*

Abstract. In general, Intelligent Tutoring Systems (ITS) fail to take account of the emotional and cognitive states of the students who use them. This paper explores the relationship between emotion and cognition when students learn via the medium of video lectures. A cognitive emotional model was constructed to determine the student's cognitive and emotional state while watching an instructional video. This model was a Bayesian belief network (BBN) model. With the method of ten times 10-fold cross-validation, evaluation results showed that the Bayesian network classifies the emotion state with 60% accuracy and classifies both the emotion and cognitive state with 48.82% accuracy. This model provides an emotional and cognitive states recognition solution for video lecture learners in a non-intrusive way with low cost.

1. Introduction

The video-recorded lecture is a primary feature of most online learning platforms and many educational institutions use video lectures to improve the effectiveness of teaching in and out of classrooms and to support distance-learning students, such as Coursera, Khan Academy, and TED. If the learners feel confused and cannot obtain cognitive and affective support while watching video-recorded lecture, their motivation decreases. When their negative emotion accumulates to a serious degree, the outcome of learning will be seriously weakened. An essential prerequisite for e-learning systems that can modify their behavior with respect to the cognitive and emotional states of a learner is an ability to detect cognitive and emotional states.

This paper introduces a non-invasive approach to assess learners' cognitive and affective states in a video learning environment using a Bayesian belief network. The modeling process and the evaluation of the cognitive affective model is presented.

2010 *Mathematics Subject Classification*. 97C30.
Keywords. Affective learning; Fuzzy classification; Bayesian belief network; Cognitive and emotional states.
Received: 24 october 2017; Accepted: 30 January 2018
Communicated by (Hari M. Srivastava)
Research supported by National Natural Science Foundation of China (No: 51365010, 61563012), Guangxi young teachers promotion project (No: 2017KY0266), and Guangxi Key Laboratory Fund of Embedded Technology and Intelligent Systems at Guilin University of Technology.

*Corresponding author.
Email address: xiaomei.tao@glut.edu.cn (Xiaomei Tao)

2. Literature Review

Emotional states can be recognized through facial recognition [1, 2], voice recognition [3], biological signal detection [4–6], posture analysis [7], text based analysis [8]; and qualitative methods such as think aloud, and interviews [9]. Some of these techniques have their limitations when applied in a video learning environment. Qualitative methods, for example questionnaires, are easy to set up, but are intrusive to the learner's learning process. The approaches on the basis of the analysis to the characteristics of speech signals can enable the classification of emotion states, but the learners cannot provide this type of data whilst watching the video. Internal biological signals can be detected by professional and sophisticated biofeedback devices that have high costs and are intrusive to the learners. Observing external expressions and behavior is an intuitive and effective way to recognize the learner's emotional state during classroom teaching, it has a low cost and is less intrusive when incorporated into an affective learning system. However, generally speaking, in the process of watching video, the frequency of obvious facial expression is low.

Previous research has taken into account an analysis of the causes of the emotional states. Boulay [10] distinguished two kinds of causes of a transition towards a negative motivational state, values-based and expectancies-based. Some other systems, such as [11, 12], adopt a subset of the emotion states developed by OCC theory [13], or variations on this, to reason about the causality in learning situations. The OCC model is a psychological model of emotions that provides a clear and convincing structure of the eliciting conditions of emotions and the variables that affect their intensities. These systems lack consideration of the teaching procedure itself and the content of the material, which means that the learners' cognitive states during learning have not been analyzed comprehensively.

The cognitive affective model described in this paper is a part of the author's PhD research work. A video study was carried out to gather data in order to construct the emotional models in this research. The methodology adopted was "Quick and Dirty Ethnography" [14]. This 'quick and dirty' approach is capable of providing valuable knowledge of the social organisation of work of a large scale work setting in a relatively short space of time. There is a trade-off between the efficiency and the completeness in this methodology. In this research, instead of a large scale study, a total of 15 students, 2 tutors, 4 sessions were used to explore how emotion works in learning generally. The results of the video study indicated that blink frequencies can reflect the learner's emotional states and it is necessary to intervene during students are in self-learning through watching an instructional video. In order to determine the learners' cognitive and emotional states in a video lecture learning environment with a non-intrusive and low cost way, a cognitive affective model was designed. The construction and the evaluation of the cognitive affective model are underpinned by the results and data collected in the video study.

3. The cognitive affective model

Considering the complexity in teaching and learning process, the uncertainty of emotion during learning, and the analysis of the cause of the emotional state, a Bayesian network [15], which has causal and uncertainty representation ability, is an ideal tool to model the emotion problem in learning. A Bayesian network is constructed as modules with the environment information from the sensors in order to predict children's emotion [16]. The results show more than 84% accuracy in the evaluation using data collected from kindergarten classes. Bayesian Networks have been used as causal modeling and reasoning tools extensively in different fields [17–19]. In this paper, only the cognitive states are taken into account as the cause of the emotional state, because this aspect is the main factor that affects the learner's emotional state during the learning process.

3.1. Emotional state and corresponding cognitive state

In this research, the definition of "emotion" [20] is "a relatively short-term, evaluative state focused on a particular intentional object (a person, an event, or a state of affairs)". The particular intentional object means an instructional event in a video lecture. The positive emotional state set is defined as P = {happiness, interest, flow}, and the negative emotional state set is defined as N = {confusion, frustration, boredom}.

Happiness, interest, confusion, frustration, and bored in this context retain their everyday conventional meanings. Flow represents the feeling of complete and energized focus in an activity, with a high level of enjoyment and fulfillment [21]. In the flow zone, the abilities of the student match the difficulty level of the learning material, for example, they can understand the materials delivered by the tutor well and they can give the correct answer to a problem.

In [22], the authors proposed some emotional conditions of learning that should exist corresponding to each cognitive process in order to improve learning, such as in the cognitive process of attention, emotional conditions are: avoiding negative emotions, avoiding emotions like joy or sadness that are not related to the learning activity, and inducing the emotion of curiosity by highlighting an element in the interface suddenly. We assume the learner's emotional states are all caused by the changes of the cognitive states. During the learning process, on the basis of Gagne's instructional theory [23], there are nine instructional events and corresponding cognitive processes. Cognitive state means the state of a person's cognitive processes during learning through watching video lecture. The cognitive state set C = {Receiving, Anticipating, Retrieving, Perceiving, Encoding, Responding, Reinforcing, Generalising}. The mapping relationships between the emotional states and the cognitive states in each steps are produced on the basis of the video study, the student's stimulated recall report. The mapping relationships in the step of presenting stimuli with distinctive features are presented in Table 1 as an example.

Table 1: The mapping relationships in the step of presenting stimuli with distinctive features

Possible emotional states		**Cognitive states**
Positive emotional states	Happy, Flow, Interested	Perceiving = ″successful″.
Negative emotional states	Confused, Frustrated	Perceiving = ″failed″. Students do not understand current knowledge point well.
		Perceiving = ″failed″. Students do not master the prerequisite knowledge point well.
	Bored	Perceiving = ″failed″. Students do not understand current knowledge point well.
		Perceiving = ″successful″. Students understand current knowledge point well.

On the basis of all the mapping relationships in each instructional step, it can be seen that the cognitive state that causes the positive emotional state is unique, while the cognitive state that causes the negative emotional state may differ. The causes of the negative emotional states are complicated, for example, in Table1, the cause of the negative emotional state could be “failed perception” or “successful perception”, and could be related to “current knowledge point” or “prerequisite knowledge point”, which should be related to completely different feedback tactics. Therefore, the negative emotional states are the emphasis of analysis and response.

3.2. Using a Bayesian belief network to model the learner's emotion

Mathematically, the cause analysis problem may be viewed as a problem integrated with diagnosis. The cognitive affective model is constructed by a Bayesian belief network. The Bayesian belief network structure for the negative emotional state is shown in Figure 1. The top layer presents the learning contextual information. Node 1, 2, . . . , to 5 represent *student's capability, learned or not, knowledge point difficulty level, prerequisite knowledge point mastered or not, learning duration* respectively. The middle layer represents the cognitive state. Node 6, 7, . . . , 13 represent cognitive state in set C = {*Receiving, Anticipating, Retrieving, Perceiving, Encoding, Responding, Reinforcing, Generalising*} respectively. The bottom layer is the specific negative emotional state in set N = {*confusion, frustration, boredom*}. The information in the top layer could be obtained from the learner's profile and video clip information. In the middle layer, the nodes represent a learner's cognitive state which is related to a given instructional step.

Given a group of input (x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11, x12, x13), (x1, x2, x3, x4, x5) is about the learning context and (x6, x7, x8, x9, x10, x11, x12, x13) is about the instructional step. Node xj, one of (x6, x7, x8, x9, x10, x11, x12, x13), is absent on the basis of the current instructional step. The probability of

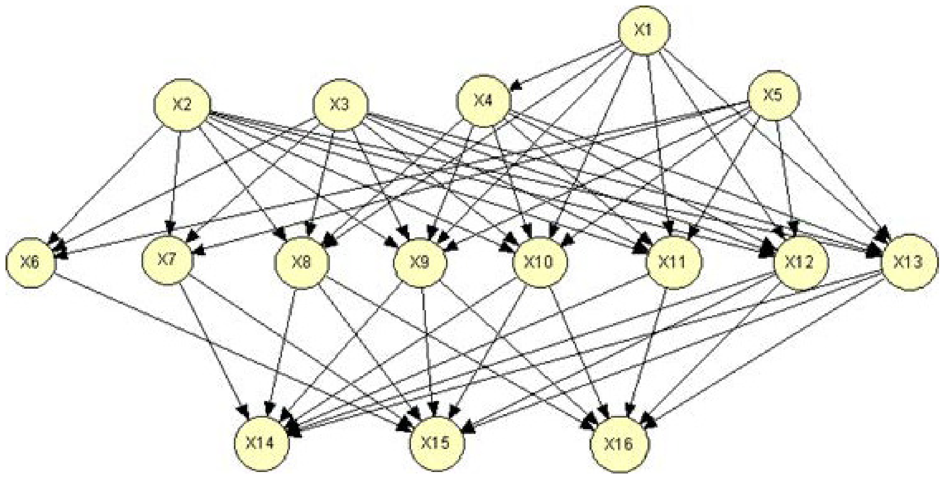

Figure 1: The Bayesian belief network structure for the negative emotional state

each negative emotional state could be calculated using the BBN, and the node X that has the maximum probability to the state of "yes" could be selected using equation 1.

$$X = \arg\max_{i=14,15,16} \mathrm{P}(x_i \,|\, x_1, x_2, \ldots, x_{13}) \tag{1}$$

The learner's cognitive state is "successful" or "failed" could be inferred by equation 2, the state that has higher probability p is the learner's cognitive state.

$$p = \max_{x_j \notin (x_6, x_7, \ldots, x_{13})} (p(x_j = successful \,|\, x_1, x_2, \ldots, x_{13}), p(x_j = failed \,|\, x_1, x_2, \ldots, x_{13})) \quad (j = 6, 7, \ldots, 13) \tag{2}$$

x_j is one of $(x_6, x_7, \ldots, x_{13})$, but when calculating $p(x_j = successful|x_1, x_2, \ldots, x_{13})$ or $p(x_j = failed \,|\, x_1, x_2, \ldots, x_{13})$, x_j is NOT in $(x_1, x_2, \ldots, x_{13})$.

Inference in Bayesian networks is performed by the Junction Tree algorithm [24]. The conditional probability table is determined by the data in the video study and Expectation Maximization(EM) parameter learning algorithm [25].

4. Evaluation

4.1. Data

In the observation of an interactive environment, we obtained 10 student's video files totaling around 375 minutes length. From the stimulated recall of the students in the interaction environment, there were 266 original records in total obtained. The data set that are used to evaluate the emotion analysis model are the same data that was used to learn the parameters in the Bayesian belief network. In this data set, the specified emotional state and cognitive state's value came from the stimulated reports by students themselves in the video study. The cases set for EM learning is obtained through processing the original cases in the video study. This case set with 173 cases related to negative states is used for parameter learning and to evaluate the emotion analysis model. Experienced tutors were invited to label the instructional event in the video, and the corresponding cognitive states are transferred by the Gagne's instructional theory. For example, *Receiving* is a cognitive state related to the instructional step of "Gaining attention", if the current video clip was labeled with "Gaining attention", we can infer that the cognitive state is *Receiving and* other cognitive states are "NA" (not applicable) due to the independence between these cognitive states. The tutors who taught the students were asked to label the learning contextual information, such as student's capability, learned or not, knowledge point difficulty level, etc. The specified emotional state and cognitive state's value came from the stimulated reports by students themselves in the video study.

4.2. Method

The training data set and the validation data set are the same set, and the model was trained and validated using 10-fold cross-validation [26]. With this method, the model is trained on data from 90% of the students and is then evaluated for accuracy on the remaining 10%. The 10-fold cross-validation method is repeated ten times to achieve an average value. In this evaluation, the group id of ten groups are 1, 2, . . . , 10 respectively. A new group of training data set and validation data set are produced each time. In each group, 90% data selected randomly from the whole data set form the training data set and the remaining 10% data are used to as evaluation data.

4.3. Results

The accuracy rate for ten groups in the evaluation respectively to the emotional state and to both emotional state and cognitive state are presented in Table 2.

Table 2: A summary of accuracy rate for ten groups

Group id	**accuracy rate (emotional state)**	**accuracy rate (emotional and cognitive state)**
1	70.59%	64.71%
2	64.71%	52.94%
3	64.71%	64.71%
4	52.94%	47.06%
5	64.71%	41.18%
6	58.82%	47.06%
7	58.82%	35.29%
8	52.94%	41.18%
9	52.94%	52.94%
10	58.82%	41.18%
Average accuracy rate	60.00%	48.82%

With the method of ten times 10-fold cross-validation, evaluation results showed that the Bayesian network classifies the emotion state with 60% accuracy and classifies both the emotion and cognitive state with 48.82% accuracy. There are 3 emotional states and 2 cognitive states' values (i.e. successful or failed)., therefore the accuracy by random selection would be respectively are 33.3% and 16.7% accurate. This supports the hypothesis that the cognitive affective model can be used to classify negative emotion and cognitive state.

5. Discussion

The accuracy rates in related research are summarized in Table 3 for comparison. D'Mello and Graesser [27] detected learners' affect by monitoring their body position and arousal during interactions with an Intelligent Tutoring System. Training and validation data on affective states were collected in a learning session with the ITS. Sabourin et al. [28] developed learner's emotional states predictive models by modeling cognitive appraisal process. Predictive models are empirically learned from data acquired from interacting with the game-based learning environment. The Bayesian network in this paper was designed on the basis of the context of video lecture learners watching instructional video. The results listed in table 3 from D'Mello and Graesser [27] was the maximum classification accuracies obtained across classifiers including Bayesian, Instance based classifiers, Rule, Decision Tree, etc. These three studies achieved similar recognition accuracy rates under the same states number. For example, for two states, the accuracy rates in [28] are 66.8% and 72.6% respectively using BBN and DBN, whilst the accuracy rates in [27] is 71%. For three and four states, the accuracy rates is 55% and 46% in [27], and the accuracy rates for 3 affective states and 3 affective states + 1 cognitive state in our research is 60% and 48.82%.

Sabourin et al. [28] achieved better accuracy by considering the emotional states transition using DBN than without considering the emotional states transition using BN in their own work. D'Mello and

Table 3: Accuracy rate comparison

Research work	Accuracy rate	Accuracy rate by random selection	States to identify	Technique
Sabourin, Mott [28]	66.8%	50%	valence (positive or negative states)	BBN
	72.6%			DBN
	25.5%	14.29%	7 affective states	BBN
	32.6%			DBN
D'Mello and Graesser [27]	71%	50%	2 affective states	Bayesian, Instance based classifiers, Rule, Decision Tree
	55%	33.3%	3 affective states	
	46%	25%	4 affective states	
This research	60%	33.3%	3 affective states	BBN
	48.82%	16.7%	3 affective states + 1 cognitive state	BBN

Graesser [29] proposed a hypothesis to illustrate the transition among the states of confusion, frustration and boredom in deep learning. The confusion state occurs due to cognition disequilibrium, and transits to the frustration state when the student experiences failure. Persistent frustration may also transition into boredom. Adding this hypothesis in to the emotion analysis model could help to categorize the negative emotional states. In our research, the modeling of the emotion analysis model mainly considers the causal relationship between the cognitive state and emotional state, and is implemented by Bayesian belief network. The inconsistency of state recognition in our experiment mainly appears in the cases that the student reported they were in frustration while the network inferred that they were in confusion. The model can distinguish frustration from confusion in the event of providing feedback, but cannot achieve this in other situations. The state of boredom may be caused by learning content which is either too complex or too simple. In addition, the boredom state caused by too simple content can be inferred by the Bayesian network in the emotion analysis model with a "successful" cognitive state. But if the emotional state is caused by content which is too complex, it tends to be categorized to "confusion" state with a "failed" cognitive state in the Bayesian network. Although the inferred emotional state is inconsistent, the correct inferred cognitive state can ensure that the cognitive feedback is appropriate. If considering the transition among the states in the emotion analysis model and modeling this by DBN might produce better results. This could be realized by adding a time slot at time t_i and add the links between the nodes in t_i to the nodes in t_{i+1}. Although this would require further research to ascertain if the results could be generalized.

6. Conclusion

This paper introduced the use of a Bayesian belief network model to determine a student's cognitive and emotional state while watching an instructional video. Evaluation results showed that the Bayesian network classifies the emotion state with 60% accuracy for three states and classifies both the emotion and cognitive state with 48.82% accuracy. The classification rate needs to be improved by considering temporal relations among the emotional states. However, this approach is a low cost and efficient solution for emotion and cognition recognition when the users are watching instructional video lectures.

Acknowledgments

As the research of the thesis is sponsored by National Natural Science Foundation of China (No: 51365010, 61563012), Guangxi young teachers promotion project (No: 2017KY0266), Guangxi Key Laboratory Fund of Embedded Technology and Intelligent Systems at Guilin University of Technology, we would like to extend our sincere gratitude to them. And we would also like to express our appreciation to Faculty of Computing, Engineering and the Built Environment at Birmingham City University.

References

[1] Linn, A. Happy? Sad? Angry? This Microsoft tool recognizes emotions in pictures. 2015 [cited 2015 11.20]; Available from: http://blogs.microsoft.com/next/2015/11/11/happy-sad-angry-this-microsoft-tool-recognizes-emotions-in-pictures/.

[2] Shojaeilangari, S., Yau, W.Y., Nandakumar, K., Li, J., Teoh, E. K., Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning, IEEE Transactions on Image Processing 24 (2015) 2140–2152.

[3] Q. Jin, C. Li, S. Chen, H. Wu, Speech emotion recognition with acoustic and lexical features, in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2015, pp. 4749–4753. Australia, Brisbane.

[4] Li, X., Zhang, P., Song, D., Yu, G., Hou, Y., Hu, B., EEG Based Emotion Identification Using Unsupervised Deep Feature Learning. in SIGIR2015 Workshop on NeuroPhysiological Methods in IR Research. 2015. Santiago, Chile.

[5] Chen, J., Hu, B., Xu, L., Moore, P., Su, Y., Feature-level fusion of multimodal physiological signals for emotion recognition. in IEEE International Conference on Bioinformatics and Biomedicine (2015) 395–399.

[6] Chen, J., Hu, B., Moore, P., Zhang, X., Ma, X., Electroencephalogram-based emotion assessment system using ontology and data mining techniques. Applied Soft Computing 30 (2015) 663–674.

[7] Gao, L., Qi, L., Guan, L., Online Behavioral Analysis with Application to Emotion State Identification. IEEE Intelligent Systems 31 (2016) 32–39.

[8] Yasmina, D., Hajar, M., & Hassan, A. M. Using youtube comments for text-based emotion recognition. Procedia Computer Science 83 (2016) 292–299.

[9] D'Mello, S. K., Craig, S. D., Sullins, J., Graesser, A. C., Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. International Journal of Artificial Intelligence in Education 16 (2006) 3–28.

[10] Boulay, B., Towards a Motivationally Intelligent Pedagogy: How Should an Intelligent Tutor Respond to the Unmotivated or the Demotivated? New Perspectives on Affect and Learning Technologies, (2011) 41–52.

[11] Hernández, Yasmín, Sucar, L Enrique. User study to evaluate an affective behavior model in an educational game. Supplementary Proceedings of the 13th International Conference of Artificial Intelligence in Education (AIED 2007), Workshop on modeling and scaffolding affective experiences to impact learning, July 9 th 2007, Marina Del Rey, CA, USA.

[12] Conati, C., H. Maclaren, Data-driven refinement of a probabilistic model of user affect. User modeling 2005 (2005) 40–49.

[13] Ortony A., Gerald L. C., Allan C., The cognitive structure of emotions. Cambridge university press, 1990.

[14] Hughes, J., King, V., Rodden, T., Andersen, H., The role of ethnography in interactive systems design. Interactions 2 (1995) 56–65.

[15] Pearl J., Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning. Society University of California Irvine 1985 (1985) 329–334.

[16] Cho S.B., J.H. Kim. Predicting group emotion in kindergarten classes by modular Bayesian networks. in International Conference of Soft Computing and Pattern Recognition (2015) 298–302.

[17] Kekolahti P., J. Karikoski, A. Riikonen, The effect of an individual's age on the perceived importance and usage intensity of communications services—A Bayesian Network analysis. Information Systems Frontiers 17 (2015) 1313–1333.

[18] Constantinou, A. C., Fenton, N., Marsh, W., Radlinski, L., From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support. Artificial Intelligence in Medicine 67 (2016) 75–93.

[19] Chen, Y., Chen, Y. W., Xu, X. B., Pan, C. C., Yang, J. B., Yang, G. K., A data-driven approximate causal inference model using the evidential reasoning rule. Knowledge-Based Systems 88 (2015) 264–272.

[20] Parkinson, B. and A.M. Colman, Emotion and motivation. Longman Harlow,, United Kingdom, 1995.

[21] Csikszentmihalyi, M. Flow, The Psychology of Optimal Experience. New York, Harper and Row, 1990.

[22] Chaffar, Soumaya, C. Frasson, The Emotional Conditions of Learning. Eighteenth International Florida Artificial Intelligence Research Society Conference, Clearwater Beach, Florida, Usa DBLP, (2005) 201–206.

[23] Gagne, R.M., The conditions of learning. New York, NY: Holt, Rinehart & Winston, 1965.

[24] Lauritzen, S.L. and D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems. Journal of the Royal Statistical Society. Series B (Methodological), 1988 157–224.

[25] Lauritzen, S.L., The EM algorithm for graphical association models with missing data. Computational Statistics & Data Analysis 19 (1995) 191–201.

[26] Kohavi, R., A study of cross-validation and bootstrap for accuracy estimation and model selection. International Joint Conference on Artificial Intelligence. Morgan Kaufmann Publishers Inc. 14 (1995) 1137–1143.

[27] D'Mello, S. and A. Graesser, Automatic Detection of Learner's Affect from Gross Body Language. Applied Artificial Intelligence 23 (2009) 123–150.

[28] Sabourin, J., B. Mott, J.C. Lester, Modeling learner affect with theoretically grounded dynamic bayesian networks, in Affective Computing and Intelligent Interaction (2011) 286–295.

[29] D'Mello, S., A. Graesser, Dynamics of Affective States during Complex Learning. Learning and Instruction 22 (2012) 145–157.