Can Affect Be Detected from Intelligent Tutoring System Interaction Data? – A Preliminary Study

Elizabeth A. Anglo and Ma. Mercedes T. Rodrigo

Department of Information Systems and Computer Sceience,
Ateneo de Manila University, Loyola Heights, Quezon City, Philippines
{eanglo,mrodrigo}@ateneo.edu

Abstract. This study attempted to determine if it is possible to create an automatic affect detector using a combination of semantic and keystroke data. While the resulting models attained detection accuracies comparable with other studies, their reliabilities were not ideal. One model however shows that interaction logs may have a potential as a detector for confusion.

Keywords: Aplusix, affect, interaction data, learner modeling.

1 Introduction

There is growing evidence that learning and emotion are tightly related, e.g., in [1]. Hence affect-sensitive intelligent tutoring systems (ITSs) may become even more effective. Automatic affect detection in ITSs is a challenging area of research. Finding a technology free of the problems associated with the current detectors is still largely an open question [2]. This paper explores the effectiveness of system interaction and semantic features from system logs to implement a real-time affect detector for an ITS.

2 Methodology

Data on the affective states that occur during learning were collected in a separate study [3] involving first and second year high school students as they interacted with Aplusix II: Algebra Learning Assistant [4], an ITS for mathematics. The system interaction (e.g., usage counter for arrow keys) and semantic data (e.g., number of problems attempted in an observation and their difficulty) were gathered from the system logs. The system interaction and semantic data were then synchronized with the affective observations using an unsupervised action filter based on a variable time window, as discussed in [5]. The synchronized logs contained 3, 000 records with each record having 58 features. Three separate datasets were created, with each dataset corresponding to an affective state, i.e., *boredom*, *confusion*, and *frustration*. This paper focuses on these affective states since they have been associated with learning, e.g., in [5]. We next applied 10 data mining approaches, including Bayesian, functions, rules and decision trees from RapidMiner [6] after which the results were examined for classification accuracy, reliability as measured using Cohen's kappa [7], and agreement with results established from previous studies.

V. Aleven, J. Kay, and J. Mostow (Eds.): ITS 2010, Part II, LNCS 6095, pp. 260–262, 2010.

3 Results, Discussion and Conclusion

Table 1 presents the average accuracy, kappa values, and highest number of correct affect guesses obtained from the program runs. The classification accuracies were comparable to each other, that is, no data mining approach was significantly better than the other in classifying the records. However, there are differences in the kappa values and number of correct affect guesses, e.g., the Functional Tree model produced the highest kappa value and made the most number of correct confused state guesses. The accuracy values are relatively good compared to accuracies using other technologies. The small number of records for each observed affect may make the effectiveness of the models debatable. An examination of the confusion matrices revealed how the models work. For boredom, the models detected 99% of the non-bored instances and 4% of the bored instances; for confusion, the models detected 97% of the non-confused records and 13% of the confused instances; and for frustration, the models detected 99% of the non-frustrated instances and only 1% of the frustrated instances. From these results, we conclude that while still not impressively accurate, at least in the case of the model for confusion, given the detrimental effects of this affect, this model offers us some detection capability better than chance alone. While the reliability value was far from being impressive, one may note that it is also the confusion model that gave the highest kappa value. This may mean that the interaction logs indeed have potential for detecting at least the state of confusion.

Table 1. Average Accuracy, Kappa Values, and Highest Number of Correct Guesses

	Boredom	Confusion	Frustration
Number of Records for Affect (% to Total Number of Records in Dataset)	69 (3%)	393 (14%)	73 (3%)
Average Classification Accuracy	97.05%	84.08%	97.02%
Best Kappa Value	0	0.09	0.02
Highest No. of Correct Affect Guesses	3	50	1

Next we observe how the resulting confusion model agrees with other studies. The Functional Tree model characterized a student in the confused state as someone who attempts a smaller number of problems and who works on a bigger number of easy problems compared to the other students. Since confusion precedes gaming and co-occurs gaming [8], we can say that the latter rule in a way indirectly concurs with previous work, e.g., [5] which showed problem difficulty may lead to gaming.

Our ultimate goal is to determine whether system interaction logs are useful for affect detection. In this preliminary analysis, we showed that the system logs offer detection capability for confusion better than chance alone. The model that showed most promise of detection capability was that for confusion. We observe this affect has the most number of occurrences among the three states studied in this paper. This might suggest that by lessening the imbalance in the dataset, we might be able to create better detectors. This is one of the directions that this research will consider next.

Acknowledgments. We thank Jean-Francois Nicaud of the Laboratoire d'Informatique de Grenoble for the use of Aplusix. We thank Sheryl Ann Lim, Alexis Macapanpan, Sheila Pascua, Jerry Santillano, Jessica Sugay, Sinath Tep, and Norma Jean Viehland, and Dr. Ma. Celeste T. Gonzalez for their assistance in organizing and conducting the studies reported here. We also thank the Ateneo de Manila High School, Kostka School of Quezon City, School of the Holy Spirit of Quezon City, St. Alphonsus Liguori Integrated School and St. Paul's College Pasig for their participation in the studies conducted. This research undertaking was made possible by the Philippines Department of Science and Technology Engineering Research and Development for Technology Consortium under the project "Multidimensional Analysis of User-Machine Interactions Towards the Development of Models of Affect". We thank Dr. Ryan Baker of the Department of Social Science and Policy Studies, Worcester Polytechnic Institute, for his invaluable support and guidance.

References

1. Craig, S., Graesser, A., Sullins, J., Gholson, B.: Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. Journal of Educational Media 29(3), 241–250 (2004)
2. Anglo, E.A.: Review – Automatic Affect Detection Technologies. Loyola Schools Review, School of Science and Engineering VIII (2009) (in press)
3. Rodrigo, M. M.T., Baker, R.S.J.d., D'Mello, S.K., Gonzalez, M. C.T., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sugay, J.O., Tep, S., Viehland, N.J.B.: Comparing learners' affect while using an intelligent tutoring system and a simulation problem solving game. In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) ITS 2008. LNCS, vol. 5091, pp. 40–49. Springer, Heidelberg (2008)
4. Nicaud, J.-F., Bouhineau, D., Chaachoua, H.: Mixing microworld and CAS features in building computer systems that help students learn algebra. International Journal of Computers for Mathematical Learning 9, 169–211 (2004)
5. Walonoski, J.A., Heffernan, N.T.: Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 382–391. Springer, Heidelberg (2006)
6. Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T.: Yale (now: RapidMiner): Rapid Prototyping for Complex Data Mining Tasks. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2006 (2006)
7. Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20(1), 37–46 (1960)
8. Rodrigo, M.M.T., Baker, R.S.J.d., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sevilla, L.R.S., Sugay, J.O., Tep, S., Viehland, N.J.B.: Affect and usage choices in simulation problem-solving environments. In: Luckin, R., Koedinger, K.R., Greer, J. (eds.) Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work. IOS Press, Amsterdam (2007)