The value of multimodal data in classification models to interpret social and emotional aspects of tutoring

First Author, Second Author, and Third Author

No Institute Given

Abstract. There are many aspects of tutoring that are associated with social and emotional learning (SEL). These aspects are complex processes that involve dynamic combinations of competencies, abilities, and knowledge. Therefore, their investigations require approaches that can provide insights into processes across multiple social planes, multiple resources, both human and tool, and data from physical and digital spaces. Hence, they might benefit from approaches that involve multimodal data. In this paper, we present our classification models to predict the tutor evaluators′ scores on the social and emotional aspects of tutor candidates. Moreover, we compare the accuracy of unimodal and multimodal data in models and show that multimodality in models leads to more accurate classifications of the candidates. Based on these results, we argue that, albeit being costly and often not very practical, multimodal models can lead to more accurate predictions of candidates when the intended classifications are on the complex and dynamic constructs of education.

Keywords: social and emotional learning · multimodal data· tutoring

1 Introduction

Tutoring involves a lot more than supporting students with their academic capabilities. Today, as educators and as learners, we are faced with many challenges. For example, an increasingly automated and AI augmented world in which children will experience a very different life to that of their parents [8]. We must prepare for the much-anticipated upheaval by ensuring that our education and training is tuned to the new demands of the workplace and society [10]. In order to achieve this ambitious goal, young learners across the globe require a broad array of social and emotional skills, attitudes, and values to succeed in school, careers, and in life [1]. Effective tutors are those who can support students on those social and emotional aspects as well as their academic capabilities. However, the evaluation of the social and emotional aspects of tutoring is a challenging task. In this paper, we are investigating the particular personal characteristics of tutors that are likely to make them successful at social and emotional aspects of tutoring. More specifically, we present our classification models of trainee tutors in terms of their success at social and emotional aspects of tutoring based on a tailored personality questionnaire and their audio analysis data. Furthermore, we compare the accuracy of models based on unimodal and multimodal data to observe the impact of different modalities of data.

2 Methodology

2.1 The context of the study

In order to create the baseline data, we first, asked expert tutors to score the social and emotional aspects of 47 tutor candidates with a performance-based activity. In the activity, the candidates were given a tutoring task and expert tutors observed the candidates′ activity and gave them a score from 1 to 5. In these scores, 1 and 2 represent an excellent candidate, who can generally be placed at any school to deliver tutoring, 3 is used for those who might need some further training on these aspects, and 4 and 5 are not desirable candidates. In addition to these scores, we interviewed expert tutors to get insights on what kind of social and emotional aspects they were observing during their evaluations. Frequently emerging themes for the expected candidates were social interactivity, engagement, emotional intelligence, and appropriate encouragement/praise of others.

2.2 Data Collection Methods

There is a long-standing wealth of literature characterizing effective tutoring and many of them emphasise the value of social and emotional aspects of tutoring for effective learning [i.e. 2,6]. In order to be able collect meaningful and relevant data on social and emotional aspects of tutoring, we collected data on various psychometric measures. More specifically, temperament was represented by two dimensions- social closeness and social anger, that are assessed by the Adult Temperament Questionnaire (ATQ; [4]). Then, the empathy which we consider as a potential representative of the emotional intelligence, we used the Trait Emotional Intelligence Questionnaire (TEIQue- SF; [7]). These personal characteristics aim to reflect the candidates′ social and emotional capabilities, specifically, testing their ability to develop social connections, be orientated to communicate and readiness to be exposed to a large volume of social interaction, along with abilities to empathy and emotion control that are all crucial in the context of tutoring. Furthermore, to identify the self-reflection of the candidates on their charismatic abilities, namely, be pervasive, confident and have abilities to make other people comfortable during mutual interactions, which are argued as significant for the effective debate tutoring [3], we used the General Charisma Inventory (GCI;[9]). Finally, the questionnaire capturing the reflection of the tutors on their abilities to follow plans and commitments and assessed by items utilised from the Big Five Inventory (BFI; [5]). In addition to these, we added two items about candidates′ previous experience in debating and tutoring. Further to the tailored questionnaire and experience data, we collected 90sec audio recordings of the candidates while they answer the question ”why do they want to become a tutor?” to have input information on their emotional traits. To analyse the audio data we use OpenSMILE open source software package. To clean the data, we omitted windows that are smaller than 1600 ms, computed SD for each variable/candidate, omitted samples outliers, and computed mean values without outliers.

3 Results

In order to reduce the large set of variables into a smaller set of components, that account for most of the variance in the tailored questionnaire variables, a principal components analysis (PCA) was run on the 22-question questionnaire created. Inspection of the correlation matrix showed that two variables (social closeness and rare social anger) had no correlation coefficient greater than 0.3, thus they were removed from the PCA. All the rest 20 variables had at least one correlation coefficient greater than 0.3 (KMO= .791, Barlett′s sphericity was significant p¡.0005). PCA revealed five components that had eigenvalues greater than one and which explained 29.728%, 10.847%, 9.159%, 7.458% and 5.683% of the total variance, respectively. Visual inspection of the scree plot indicated that four components should be retained. The four-component solution explained 57.193% of the total variance. A Varimax orthogonal rotation was employed to aid interpretability. The rotated solution exhibited 'simple structure' (Thurstone, 1947) with strong loadings of extraversion, outgoingness, and leadership items on component 1, charisma, enthusiasm, and the tendency to make people comfortable items on component 2, assertiveness, organization and the tendency of being influential items on component 3 and neuroticism, non-assertiveness items on component 4. There was not enough data to implement a similar PCA approach for the audio data. However, we omitted highly correlated variables from the data.

3.1 Classifications of Tutor Candidates from Various Data Inputs

Multinomial logistic regression to classify the candidates into three groups: those scored 1 or 2, those scored 3, an those scored 4 or 5 was found to be the best classification tool for all data modalities and variables investigated here. We tested all models′ fitting information with pearson chi-square tests. Table 1 below shows the results of the model built with just the two experience variables, (df=14) = 10.73, p=0.707. However, the model′s fitting information shows that the full model does not significantly predict the scores.

Table 2 shows the results of the classification using only the audio variables. The model′s goodness of fit shows that the model fits the data well, (df=56) = 61.63, p=0.282. Moreover, the model′s fitting information shows that the full model significantly predicts the score (df=24) =41.72, p=0.014. In the likelihood ratio tests, Interest - passive (df = 2) = 15.33, p=0.000), Emotion anger (df = 2) = 9.06, p=0.011), Affect nervous (df = 2) = 20.19, p=0.000), and Affect

Table 1: classification based on experience variables.

Observed score	Estimated score 1/2	3	4/5	Percent correct
1/2	0	9	0	0.0%
3	0	22	0	100.0%
4/5	0	10	0	0.0%
Overall	0.0%	100.0%	0.0%	53.7%

Table 2: classification based on audio variables.

Observed score	Estimated score 1/2	3	4/5	Percent correct
1/2	8	1	0	88.9%
3	2	15	5	68.2%
4/5	0	3	7	70.0%
Overall	24.4%	46.3%	29.3%	73.2%

Table 3: classification based on experience and the survey variables.

Observed score	Estimated score 1/2	3	4/5	Percent correct
1/2	4	5	0	44.4%
3	3	17	2	77.3%
4/5	1	8	1	10.0%
Overall	19.5%	73.2%	7.3%	53.7%

Table 4: multimodal classification, the experience, survey, and two of the stronger audio variables.

Observed score	Estimated score 1/2	3	4/5	Percent correct
1/2	8	1	0	88.9%
3	1	19	2	86.4%
4/5	0	4	6	60.0%
Overall	22.0%	58.5%	19.5%	80.5%

aggressive (df = 2) = 8.40, p=0.015) were found to be significant in the classification model. Table 3 shows the classification results using the survey and the experience (df=56) = 61.13, p=0.297). However, the model′s fitting information shows that the full model does not significantly predict the score. As can be seen in table 4, when we built the multimodal classification model, using all modalities (the survey variables, the experience variables and two of the most significant audio variables (nervous and anger audio indicators), the model′s goodness of fit shows that the model fits the data well (df=52) = 48.56, p=0.610. Moreover, the model′s fitting information shows that the full model significantly predicts the score, better than the intercept-only model alone (df=28) =47.05, p=0.014). In the likelihood ratio tests, the Extrovert outgoing lead factor (df = 2) = 11.08, p=0.004), the Assertive organized influential factor (df = 2) = 13.03, p=0.001), the neurotic not assertive factor (df = 2) = 8.2, p=0.017), the social closeness survey item (df = 2) = 11.50, p=0.003), the social anger rare survey item, (df = 2) = 7.35, p=0.025), tutoring experience, (df = 6) = 14.38, p=0.026), and debating experience (df = 6) = 21.92, p=0.001) variables were found to be significant in the classification model. The two audio variables, also found to be significant predictors) :Affect nervous, (df = 2) = 19.92, p=0.000), and Emotion Anger, (df = 2) = 19.92, p=0.000).

4 Conclusions

In this paper, we showed that the extrovert leader, assertive organizer, neurotic, and charismatic, personality traits, as well as tutoring and debating experience, are all significant factors that can be considered to identify tutors who are likely to be successful at social and emotional aspects of tutoring. Furthermore, when two audio variables (affect nervous and emotion anger) were added to the classification model, the value of other variables has become significant in the classification outputs. Specifically, it was found that candidates who were given 3 by expert tutors are more likely to show higher level of anger relatively to those who scored 1 and 2 (p= 0.042, B = 78.34) and are less likely to show signs of being nervous (p= 0.014, B = -426.697), where these difference were not significant between those who scored 4 and 5 and those who scored 1 and 2. Although coming from a relatively small sample size of 47 tutor candidates, the results show the potential value of multimodality in classification models that are built to evaluate complex and dynamic educational constructs such as social and emotional aspects of tutoring.

5 References

1. Aspen Institute. (2019). *From a nation at risk to a nation at hope. National Commission Final Report.* Retreived from https://www.aspeninstitute.org/events/national-commission-final-report-release/

2. Drovnikov, A. S., Nikolaev, E. L., Afanasev, A. S., Ivanov, V. N., Petrova, T. N., Tenyukova, G. G., ... & Povshednaya, F. V. (2016). Teachers professional competence assessment technology in qualification improvement process.*International Review of Management and Marketing*,6(1), 111-115.

3. Evagorou, M., & Dillon, J. (2011). *Argumentation in the teaching of science. InThe professional knowledge base of science teaching*(pp. 189-203). Springer, Dordrecht.

4. Evans, D. E., & Rothbart, M. K. (2007). Developing a model for adult temperament.*Journal of Research in Personality*,41(4), 868-888.

5. John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives.*Handbook of personality: Theory and research*,2(1999), 102-138.

6. Moafian, F., & Ghanizadeh, A. (2009). The relationship between Iranian EFL teachers′ emotional intelligence and their self-efficacy in Language Institutes.*System*,37(4), 708-718.

7. Petrides, K. V. (2009). Psychometric properties of the trait emotional intelligence questionnaire (TEIQue). In*Assessing emotional intelligence*(pp. 85-101). Springer, Boston, MA.

8. Siraj, I. (2017). Nurturing 21st century skills in early childhood education and care: the way forward′. In Loble, L., Creenaline, T. & Hayes, J. , (Ed.), *Future Frontiers Education for an AI world* (pp. 141-153). Australia: Melbourne University Press & NSW Department of Education.

9. Tskhay, K. O., Zhu, R., Zou, C., & Rule, N. O. (2018). Charisma in everyday life: Conceptualization and validation of the General Charisma Inventory.*Journal of personality and social psychology*,114(1), 131.

10. Tucker, M. (2017) Educationg for a Digital Future: the Challenge In Loble, L., Creenaline, T. & Hayes, J., (Ed.), *Future Frontiers Education for an AI world* (pp. 21-38). Australia: Melbourne University Press & NSW Department of Education.