Abstract

Online learning, which gained traction in the 1980s, has expanded access to education by offering flexible, interactive, and effective learning opportunities through web-based platforms and learning management systems. The success of online learning is heavily influenced by factors such as self-motivation, student engagement, and interaction between students and instructors. Increased engagement can boost self-motivation and foster more effective interactions, ultimately enhancing the quality and experience of online education. Therefore, improving learner engagement is crucial in overcoming challenges like the digital divide, limited face-to-face interaction, and issues related to self-motivation.

However, in engagement estimation research, due to factors like the Hawthorne effect, existing public datasets often suffer from class imbalance, with relatively few data points representing low engagement levels. This imbalance presents a significant challenge in accurately training and validating machine learning models for engagement estimation. We introduce an original preprocessing approach called "Skipped Moving Average," which not only preserves the integrity of the original video data but also captures its temporal dynamics and variations to address the imbalance issue.

First, to enrich the existing computer vision features and better interpret learners' facial and body language during online learning, we adopted a series of features that can represent facial and body information. Additionally, to further enhance our input features, we experimented with features such as standard deviation and extreme values. We then introduced our proposed Skipped Moving Average data processing method, which includes selecting an appropriate skipping window based on the current data distribution, as well as how to reasonably choose oversampled data sequences using cosine similarity. We also experimented with different normalization methods to evaluate their effectiveness in processing video sequence data.

In the experimental phase, we divided the work into two major parts. Experiment 1 used LSTM and LSTM-FCN models to verify whether the proposed SMA preprocessing method could address the issue of imbalance in the current video sequence data. Ultimately, the combination of Skipped Moving Average Oversampling and Standard Deviation for training and validation produced the best outcomes. For engagement estimation with different labels, it achieved Recall/Precision/F1 scores of 0.462/0.157/0.234 for the low label, 0.449/0.504/0.475 for the high label, and 0.456/0.501/0.477 for the very high label. To further validate our proposed method, we also compared it with the SMOTE oversampling method, which further demonstrated the superiority of our approach.

In Experiment 2, we used transfer learning to verify that the proposed SMA data processing method could be applied to different datasets. The three datasets used in the experiment had varying sample time spans and were quite irregular. Our proposed method achieved Recall/Precision/F1 scores of 0.635/0.720/0.675 for the low engagement label, which is an improvement of nearly 0.25 in the F1 score compared to the results before applying transfer learning.

In this study, we tackled the challenge of class-imbalanced time-series video data in the context of engagement estimation and detection by introducing a novel approach: Skipped Moving Average oversampling. This approach not only mitigates the effects of class imbalance but also preserves the continuity and authenticity of the time-series data, leading to more precise and consistent results in engagement detection.

Keyword: emotional engagement estimation, time-series data, oversampling, online learning, class imbalances data