Data Imputation in EEG Signals for Brainprint Identification

Siaw-Hong Liew[1], Yun-Huoy Choo[1(✉)], and Yin Fen Low[2]

[1] Computational Intelligence and Technologies (CIT) Research Group, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), 76100 Durian Tunggal, Melaka, Malaysia
siawhong.liew@gmail.com, huoy@utem.edu.my

[2] Machine Learning and Signal Processing (MLSP) Research Group, Faculty of Electronics and Computer Engineering, Universiti Teknikal Malaysia Melaka (UTeM), 76100 Durian Tunggal, Melaka, Malaysia
yinfen@utem.edu.my

Abstract. Electroencephalograms (EEG) signals have very low signal-to-noise ratio, thus can be easily affected and changes over milliseconds. Normally, trials with excessive body movements or other types of artefacts with amplitude more than 100 μV should be discarded to reduce the noise stains. Scrapping the affected features is not advisable. Therefore, missing values imputation is essential to avoid incomplete data that may deteriorate the computational modelling performance. Hence, this paper proposes a similarity matching method to replace the missing values in the EEG trials. The main idea of the missing values imputation is based on the similarity measure between the trials. The trials with the highest similarity is taken to replace the missing values for the related EEG channels. Statistical evaluation and classification evaluation are used to evaluate the reliability of the proposed similarity matching method. The mean, variance and standard deviation are used for statistical evaluation. For the classification evaluation, the dataset is classified for brainprint identification by using the Incremental Fuzzy-Rough Nearest Neighbour (IncFRNN). The statistical evaluation proved that the proposed similarity matching imputation method is promising when the missing values are not come from the same channels. The classification results achieved the excellent performance with 98.19% in accuracy and 0.998 in AUC.

Keywords: Missing values · EEG signals · Brainprint identification

1 Introduction

Biometric identification is one-to-N matching, where an individual identity is determined from a group of people that are being evaluated [1]. Both the biometric identification and the authentication share the same pre-processing steps, i.e. the feature extraction and the feature selection. However, the output for biometric identification is based on the number of subjects to be identified, while the output for biometric

J. C. Hung et al. (Eds.): FC 2018, LNEE 542, pp. 278–286, 2019.
https://doi.org/10.1007/978-981-13-3648-5_31