Comparing the Apparent Error Rate of Gender Specifying from Human Skeletal Remains by Using Classification and Cluster Methods

Authors : Jularat Chumnaul

Abstract : In forensic science, corpses from various homicides are different; there are both complete and incomplete, depending on causes of death or forms of homicide. For example, some corpses are cut into pieces, some are camouflaged by dumping into the river, some are buried, some are burned to destroy the evidence, and others. If the corpses are incomplete, it can lead to the difficulty of personally identifying because some tissues and bones are destroyed. To specify gender of the corpses from skeletal remains, the most precise method is DNA identification. However, this method is costly and takes longer so that other identification techniques are used instead. The first technique that is widely used is considering the features of bones. In general, an evidence from the corpses such as some pieces of bones, especially the skull and pelvis can be used to identify their gender. To use this technique, forensic scientists are required observation skills in order to classify the difference between male and female bones. Although this technique is uncomplicated, saving time and cost, and the forensic scientists can fairly accurately determine gender by using this technique (apparently an accuracy rate of 90% or more), the crucial disadvantage is there are only some positions of skeleton that can be used to specify gender such as supraorbital ridge, nuchal crest, temporal lobe, mandible, and chin. Therefore, the skeletal remains that will be used have to be complete. The other technique that is widely used for gender specifying in forensic science and archeology is skeletal measurements. The advantage of this method is it can be used in several positions in one piece of bones, and it can be used even if the bones are not complete. In this study, the classification and cluster analysis are applied to this technique, including the Kth Nearest Neighbor Classification, Classification Tree, Ward Linkage Cluster, K-mean Cluster, and Two Step Cluster. The data contains 507 particular individuals and 9 skeletal measurements (diameter measurements), and the performance of five methods are investigated by considering the apparent error rate (APER). The results from this study indicate that the Two Step Cluster and Kth Nearest Neighbor method seem to be suitable to specify gender from human skeletal remains because both yield small apparent error rate of 0.20% and 4.14%, respectively. On the other hand, the Classification Tree, Ward Linkage Cluster, and K-mean Cluster method are not appropriate since they yield large apparent error rate of 10.65%, 10.65%, and 16.37%, respectively. However, there are other ways to evaluate the performance of classification such as an estimate of the error rate using the holdout procedure or misclassification costs, and the difference methods can make the different conclusions.

Keywords : skeletal measurements, classification, cluster, apparent error rate