Advanced Science Letters
Vol. , 000–000, 2017

Improved Boosted Decision Tree Algorithms by Adaptive Apriori and Post-pruning for predicting Obstructive Sleep Apnea

Doreen Ying Ying Sim[1*], Chee Siong Teh[1], Ahmad Izuanuddin Ismail[2]

[1]Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, Kuching, Sarawak, Malaysia
[2]Respiratory Medicine Unit, Department of Respiratory Medicine, UiTM Medical Specialist Centre, Faculty of Medicine, Universiti Teknologi MARA, Selangor, Malaysia

The improved version of Boosted Decision Tree algorithm, named as Boosted Adaptive Apriori post-Pruned Decision Tree (Boosted AApoP-DT), was developed by referring to Adaptive Apriori (AA) properties and by using post-pruning technique. The post-pruning technique used is mainly the error-complexity pruning for the decision trees categorized under Classification and Regression Trees. This technique estimates the re-substitution, cross-validation and generalization error rates before and after the post-pruning. The novelty of the post-pruning technique applied is that it is augmented by AA properties and these depend on the data characteristics in the dataset(s) being accessed. This algorithm is then boosted by using AdaBoost ensemble method. After comparing and contrasting this developed algorithm with the algorithm without being augmented by AA, i.e. Boosted post-Pruned Decision Tree (Boosted poP-DT), and the classical boosted decision tree algorithm, i.e. Boosted DT, there is a stepwise improvement shown when comparison proceeds from Boosted DT to Boosted poP-DT and to Boosted AApoP-DT.

Keywords: Boosted Adaptive Apriori post-Pruned Decision Tree; Adaptive Apriori; post-pruning technique; error-complexity pruning; AdaBoost ensemble method.

1. INTRODUCTION

Risk factors of Obstructive Sleep Apnea (OSA) have well been globally researched, but to develop a prediction system for OSA disease based on Adaptive Apriori (AA) and post-pruning approaches by using raw data** collected in Malaysia is a relatively new area yet to be researched upon[1-4]. Novelty of this research is the post-pruning techniques applied are the error complexity pruning techniques which are augmented by the AA properties. Main contribution of this research is that the post-pruning techniques (applied before applying AdaBoost) embark on error-complexity pruning by complexity penalty or pruning of sub-tree(s) in order to reduce the re-substitution error rate, but to finally select the optimally pruned subtree(s) based on AA properties of the datasets. Another contribution is the AA augmented post-pruning applied can further refine the outlier(s) detection, noise(s) detection and subtree and/or node removals. This contribution differs from the existing or past research work, i.e. using post-pruning without AA.

* Email: dsdoreenyy@gmail.com ** see Acknowledgments

Item-sets that have a support above the minimum support are known as frequent itemsets[5-9]. Since frequent item-sets serve as an estimation of joint probabilities of events, mining frequent item-sets become very important in pattern recognition[5-9,12,13]. Eq.1 shows that Apriori is the special case of Adaptive Apriori (AA). Novelty of AA is that it breaks the barrier of uniform minimum support by defining the best minimum support, i.e. P_{minsup}, for each schema individually with respect to the preservation of Apriori. Unlike Apriori, AA does not assure every subset of a frequent itemset to be frequent[5,6,9,10].

$$\text{Apriori} \subseteq \text{Adaptive Apriori} \qquad \text{(Eq.1)}$$

Post-pruning is usually applied for datasets which its characteristics are not well-known, but it is more reliable than pre-pruning. It is applied in a bottom-up fashion[1,7-10]. Main contribution of this part of research involves pruning of subtree(s) or nodes which violate monotonicity or other AA properties. Again, this research contribution differs from existing or past research which does not embark on AA properties or characteristics of datasets.

 1936-6612/2017/4/400/008 doi:10.1166/asl.2011.1261