Category Trees

Category Trees are like Decision Trees, except that they branch on category and not feature. It uses a supervised clustering technique, where each category is represented by a separate base classifier. Each base classifier then classifies its own subset of data rows and creates a centroid to represent the category. If the classifier is subsequently associated with rows from other categories, it needs to create new classifiers for the incorrect data. The classifier therefore branches to new layers when there is a split in the data, and creates new classifiers there for the incorrectly classified rows. The schematic of Figure 1 shows the classification process, where a new layer has been added to classifier A, so that it can correctly re-classify the category A and B sub-groups that are closest to it. The second level uses a subset of the whole dataset that is specifically only the data rows assigned to the classifier at the parent level.

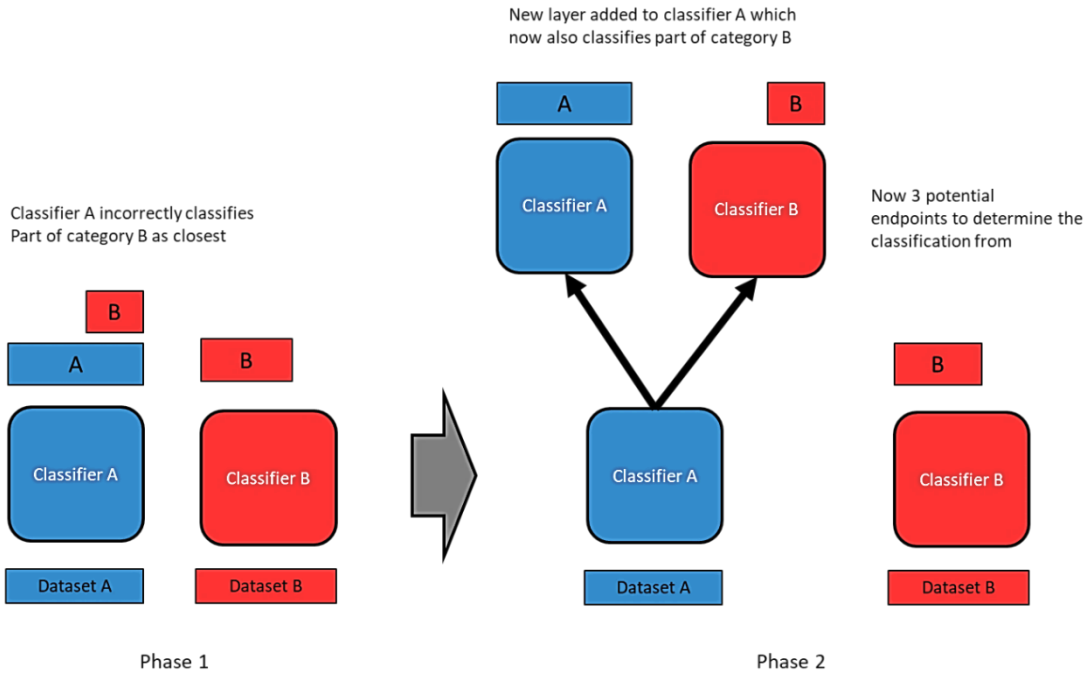

Figure 1. Schematic of the classifier in action. Phase 1 realises that classifier A also classifies part of category B better. Phase 2 then adds a new layer to classifier A, to re-classify this subset only.

Results

Tests show that the method can successfully classify a diverse set of benchmark datasets and many to 100% accuracy. It seems to work well with medical or biological data, for example: Liver, Heart Disease, Heart Images, Breast Cancer, Diabetes and Parkinson Disease, all classified 100% correctly.

Reference

Greer, K. (2021). Category Trees – Classifiers that Branch on Category, International Journal of Artificial Intelligence & Applications (IJAIA), Vol.12, No.6, pp. 65 - 76.