Supervised Taxonomies—algorithms And Applications

ABSTRACT

- This paper focuses on a new type of taxonomy called supervised taxonomy (ST).
- Supervised taxonomies are generated considering background information concerning class labels in addition to distance metrics, and are capable of capturing class-uniform regions in a dataset.
- A hierarchical, agglomerative clustering algorithm, called STAXAC that generates STs is proposed and its properties are analyzed.
- Experimental results are presented that show that STAXAC produces purer taxonomies than the neighbor-joining (NJ) algorithm—a very popular taxonomy generation algorithm.

EXISTING SYSTEM

- A hierarchical, agglomerative clustering algorithm, called staxac that generates sts is proposed and its properties are analyzed.
- Experimental results are presented that show that staxac produces purer taxonomies than the neighbour-joining (nj) algorithm—a very popular taxonomy generation algorithm.
- We introduced novel measures and algorithms that assess classification complexity, class modality, and show that sts can be used as the main input of an effective data-editing tool to enhance the accuracy of k-nearest neighbor classifiers.
- We demonstrated in our experimental evaluation that assessing the classification complexity of a st provides us with a good estimate of the difficulty of the classification problem at hand.

DISADAVANTAGE

- It has been observed that k-nearest neighbor classifiers (k-nn) are sensitive to atypical examples whose presence in the training set may lead to poor accuracy and unnecessary storage

- The purities at intermediate nodes are higher compared with the nj-generated taxonomy.

- Unlike SC and semi-supervised clustering which outputs flat clusters, supervised taxonomy generation outputs hierarchically organized clusters which can be extracted or analyzed for various degree of impurities.

PROPOSE SYSTEM

- A hierarchical, agglomerative clustering algorithm, called STAXAC that generates sts is proposed and its properties are analyzed.
- Propose a svm-based approach to sc. Bagherjeiran et al. [3] introduce supervised similarity assessment.
- The goal of supervised similarity assessment is to obtain a distance function that separates well examples belonging to different classes.
- Vilalta et al. [29] propose a class decomposition method that identifies class-uniform clusters before a classifier is applied;
- However, in this approach the clusters are obtained by clustering the examples of each class separately and not jointly, as SC does.

ADVANTAGES

- The classification complexities were compared against each other and against the average accuracy rates of 24 classifiers on the same 8 benchmark datasets
- With an hypothetical two-class dataset (the hallow disks, and the dark disks).
- The benefits of using sts for dataset editing are briefly discussed.

SOFTWARE REQUIREMENT

- **HARDWARE REQUIREMENT:**
- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 1.44 Mb.
- Monitor : 15 VGA Colour.
- Mouse : Sony.
- Ram : 512 Mb.

CONT..

- **SOFTWARE REQUIREMENT:**
- Operating system : Windows XP.
- Coding Language : ASP. Net with C#
- Data Base : SQL Server 2005.

Conclusion

- The discussions in this papers centers on supervised taxonomies. Supervised taxonomies are generated considering background information in form of class labels in addition to distance metrics, and are capable of capturing class-uniform regions in a dataset.

- By analyzing the generated tree structure, a biologist can interpret the clustering result and gain insights into how the biological groupings are related, taking more a supervised point of view.

- We also introduce, STAXAC, a supervised, hierarchical clustering algorithm to generate supervised taxonomies; it conducts a “wider” search for the best pair of clusters to merge while maximizing purity, enabling the algorithm to obtain sub-trees of higher purities than traditional agglomerative methods.