Adaptive Feature Selection for Hyperspectral Data Analysis Using a Binary Hierarchical Classifier and Tabu Search

Donna Korycinski[1], Melba Crawford[1], J.W. Barnes[2], Joydeep Ghosh[3]
[1]Center for Space Research, University of Texas at Austin
3925 W. Braker Lane, Suite 200, Austin, TX 78759-5321 (512)471-7993
[2]Operations Research and Industrial Engineering Group, Dept. of Mechanical Engineering
[3]Dept. of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX 78712
crawford@csr.utexas.edu

Abstract- **High dimensional inputs coupled with scarcity of labeled data are among the greatest challenges for classification of hyperspectral data. These problems are exacerbated if the number of classes is large. High dimensional output classes may be handled effectively by decomposition into multiple two-class problems, where each sub-problem is solved using a suitable binary classifier, and outputs of this collection of classifiers are combined in a suitable manner to obtain the answer to the original multi-class problem. This approach is taken by the binary hierarchical classifier (BHC). The advantages of the BHC for output decomposition can be further exploited for hyperspectral data analysis by integrating a feature selection methodology with the classifier. Building upon the previously developed best bases BHC algorithm with greedy feature selection, a new method is developed that selects a subset of band groups within metaclasses using reactive tabu search. Experimental results obtained from analysis of Hyperion data acquired over the Okavango Delta in Botswana are superior to those of the greedy feature selection approach and more robust than either the original BHC or the BHC with greedy feature selection.**

INTRODUCTION

For classification of hyperspectral data, there are potentially hundreds of correlated inputs which may result in unstable estimates, particularly when there is a small quantity of training data. Common approaches to overcome this problem are feature extraction and feature subset selection. The computational complexities of optimal feature selection methods have forced acceptance of heuristic techniques that find good, near-optimal subsets in relatively short computational times. A comparative study of several of the well-known optimal and sub-optimal feature selection algorithms [e.g. variations of Sequential Forward/Backward Selection (SFS/SBS), Branch and Bound and relaxed Branch and Bound] is contained in [1]. Other approaches include genetic algorithms [2], simulated annealing [3], and the Tabu Search (TS) *meta*heuristic [4]. In this study, a new model is developed which incorporates the use of TS with the multiclassifier system known as the Best Bases Binary Hierarchical Classifier (BB BHC) [5-8] for analysis of hyperspectral data. The primary goal in development of the BHC was output decomposition for problems with a medium to large number of classes. While the classification accuracies obtained from the BHC are typically good, problems are encountered if the number of inputs is extremely large and the amount of training data is limited.

METHODOLOGY

The BHC was developed for a *C*-class problem and forms a binary tree-type hierarchical classifier (at each node of the tree, only two branches are created) [5]. Sets containing more than one class are known as metaclasses and are the *internal nodes* of the tree structure; sets containing individual classes are the *leaf nodes* of the tree which are the final nodes of the branches. The metaclass at the top of the tree structure includes all original classes. The internal nodes of the tree, to include the top node, depict a two-metaclass problem that partitions the classes at each internal node, O_n, into two *child* nodes, O_{2n} and O_{2n+1}, using the Generalized Associative Modular Learning System (GAMLS) [9], a deterministic annealing-type algorithm, where $O_{2n}+O_{2n+1}=O_n$; this is accomplished recursively at each internal node until the leaves of the tree structure contain the individual classes and no more partitioning can be executed. Ultimately, this framework yields a hierarchical tree structure of *C*-1 internal nodes (two-metaclass problems) and *C* leaf nodes. A feature selection (FS) option is implemented to reduce the input space. It involves a greedy search using the incremental increase in classification accuracy at each node as the selection criterion and is performed as a post processing step after the tree is constructed. Unclassified observations are ultimately labeled using the resulting binary hierarchical tree structure.

The concept of BB was developed [7] and later integrated in the BHC [8] to reduce the input candidates in remotely sensed data. Using best bases feature extraction, the features which are contiguous in the spectrum and are highly correlated are combined to form a single feature. The resulting BB features then replace the original features, thereby reducing the dimensionality of the input space while exploiting the correlation structures inherent in the data.

TS is a *meta*heuristic method for solving combinatorial optimization problems [10]. TS explores from its incumbent solution, examines neighboring solutions, i.e. those solutions that can be reached by a single move within the specified *move-neighborhood*, and moves to the neighboring solution with the best non-tabu solution. It avoids cycling and escapes from local optima by using a *tabu list*, which incorporates solution attributes of recent solutions that are forbidden for *tabu tenure future* moves. An *aspiration criterion* may be introduced to allow TS to make a tabu move if stipulated conditions are satisfied. TS can include intensification and diversification elements: intensification allows a deeper search into promising areas of the solution space, and diversification encourages movement to yet unexplored or less explored areas of the solution space. Finally, the search halts and returns the best solution found when a *stopping criterion* is met. As with all heuristic methods, the solution returned is not guaranteed to be optimal. TS feature selection (TS-FS) is performed on the resulting BB BHC decision tree using the features output by the greedy FS as its incumbent solution. This approach has not previously been utilized for feature selection in hyperspectral data analysis and was investigated in this research as a means of improving classification accuracies within the BHC and BB BHC frameworks. A flowchart of the algorithm is displayed in Fig. 1.

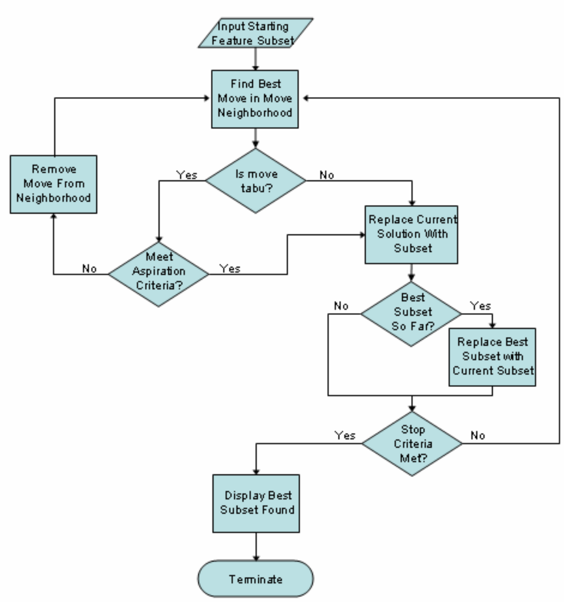

Fig. 1. Flowchart of TS-FS algorithm.

RESULTS

The TS-FS algorithm was applied to BB BHC trees obtained from Hyperion hyperspectral remotely sensed data acquired over Botswana. The data cover a subset of the Okavango Delta of Botswana that is undergoing change due to anthropogenic and natural processes such as seasonal flooding. The Hyperion dataset consists of observations from 14 identified classes representing the land cover types in the area studied, each with 242 candidate features. Uncalibrated and noisy bands that cover water absorption features are removed and the remaining 145 bands are included as candidate features: [10-55, 82-97, 102-119, 134-164, 187-220]. Class information is presented in Table I. The University of Texas Center for Space Research provided ten randomly sampled partitions of the data: 50% for training and 50% for testing the classifiers.

TABLE I
CLASS INFORMATION FOR BOTSWANA HYPERION DATA

Class #	Class Name	Sample Size
1	water	270
2	hippo grass	101
3	floodplain grasses1	251
4	floodplain grasses2	215
5	reeds1	269
6	riparian	269
7	firescar2	259
8	island interior	203
9	acacia woodlands	314
10	acacia shrublands	248
11	acacia grasslands	305
12	short mopane	181
13	mixed mopane	268
14	exposed soils	95

The TS parameters were tuned using the first experiment and were used for the remainder of the experiments. Since the features are combined differently and thereby BB features subsequently selected differently for each metaclass on each tree, only general results are presented. The overall classification accuracies for the BB BHC, BB BHC with greedy FS and the BB BHC with TS-FS are displayed in Table II. A dynamic tabu tenure initialized at 5 was allowed to vary from 3 to 8. The stopping criterion was 30 iterations and the maximum number of iterations to continue with no improvements was set at 10.

TABLE II
OVERALL EXPERIMENT CLASSIFICATION ACCURACIES

Experiment	BB BHC	BB BHC FS	BB BHC TS-FS
HYP1	89.38%	88.94%	91.60%
HYP2	91.54%	86.66%	90.92%
HYP3	91.17%	85.36%	93.27%
HYP4	92.16%	87.77%	93.02%
HYP5	92.28%	86.35%	93.70%
HYP6	91.54%	88.14%	93.39%
HYP7	91.72%	89.31%	91.41%
HYP8	92.46%	86.41%	93.21%
HYP9	90.30%	89.19%	90.80%
HYP10	92.22%	85.55%	91.97%
Average	**91.48%**	**87.37%**	**92.33%**

The BB BHC reduced the 1885 original features per tree to an average of 850.7 BB features per tree (averaging 65.44 BB features per metaclass). The greedy feature selection chose an average of 40.10 BB features per tree while the TS-FS chose an average of 91.70 BB features. The TS-FS retained an average of 25.30 BB greedy features while maintaining an average of 7.70 of the first-chosen BB features per tree. In every experiment using the test data, the tree structure utilizing the TS-FS with BB resulted in higher overall classification accuracies than the BB BHC with the greedy FS by an average of 4.96% per experiment. In 7 of the 10 experiments it achieved higher overall classification accuracies than the BB BHC, and it resulted in a higher overall average accuracy. The average classification accuracies for each class for each BB algorithm are displayed in Table II highlighting the highest average accuracy per class. The BB BHC with TS-FS clearly outperforms the BB BHC with greedy FS, and it exhibits the ability to classify a majority of the classes more consistently. While not improving the classification accuracies for all individual classes, the average overall classification accuracies were improved.

TABLE III
OVERALL CLASS PERCENTAGE ACCURACIES

Class #	BB BHC	BB BHC FS	BB BHC TS-FS
1	100.00	99.40	99.78
2	94.00	95.60	96.00
3	95.36	89.12	94.40
4	96.92	94.77	95.44
5	89.19	75.15	88.97
6	80.16	60.81	82.99
7	98.80	91.77	95.81
8	96.52	96.41	96.21
9	86.18	89.09	84.83
10	90.06	90.72	90.00
11	92.45	89.08	93.70
12	89.09	91.90	92.46
13	88.36	78.27	92.38
14	81.06	98.44	99.16
Average	**91.30**	**88.61**	**93.01**

For the 10 experiments, the BB BHC constructed 7 different tree structures, and no tree was duplicated more than twice. A representative tree structure is displayed in Fig. 2. All of the trees do not share the same partition of the root node. Experiments HYP1, HYP2, HYP7 and HYP9 place the acacia woodlands (class 9) with the right branch while all other experiments place it with the left branch. These four experiments yield the four lowest overall BB BHC with TS-FS classification accuracies while yielding four of the six lowest accuracies for the BB BHC.

CONCLUSIONS AND FUTURE WORK

The impact of TS-FS upon the BHC classification accuracies was demonstrated to be positive. When FS was conducted, TS's ability to find improved feature subsets significantly improved the overall classification accuracies. Furthermore, these improved feature subsets provide more domain knowledge, overall classifier interpretability and possible transportability of the classifiers. The TS implementations are sensitive to the resulting decision tree structures; therefore, if better decision trees can be constructed, the TS implementations will be enhanced and ultimately more useful for increasing classification accuracies. Using the TS-FS in the construction of the decision tree is one method to accomplish this. Additionally, alternative classifiers and/or measures of goodness pertaining to the subset selection may improve the ability of the decision tree to classify with greater accuracies. A distance measure could be incorporated into the BHC framework in both the classifier and feature subset selection. Typical distance measures include the Mahalanobis distance, the Battacharyya distance, the Jeffries-Matusita distance and the Patrick-Fisher distance. TS-FS is also being investigated with other classifiers, including the Bayesian Pairwise Classifier [11].

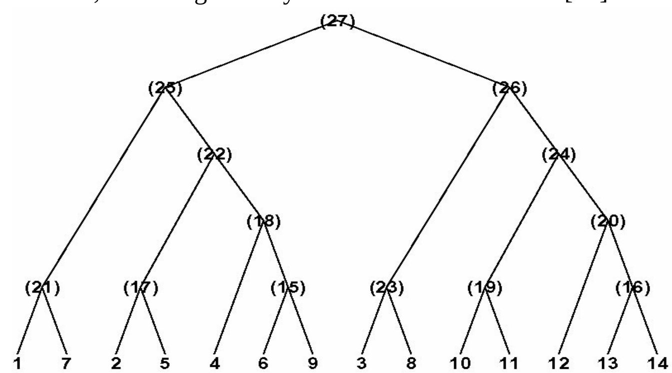

Fig 2. Representative decision tree structure.

REFERENCES

[1] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognition*, vol. 33, issue 1, pp. 25-41, Jan. 2000.

[2] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *Pattern Recognition Lett.*, vol. 10, pp. 335-347, Nov. 1989.

[3] W. Siedlecki and J. Sklansky, "On automatic feature selection," *Int. J. of Pattern Recognition and Artif. Intell.*, vol. 2, issue 2, pp. 197-220, 1988.

[4] H. Zhang and G. Sun, "Feature Selection Using Tabu Search Method," *Pattern Recognition*, vol. 35, issue 3, pp. 701-711, Mar. 2002.

[5] S. Kumar, *Modular learning through output space decomposition*, Ph.D. dissertation, The University of Texas at Austin, 2000.

[6] S. Kumar, J. Ghosh, and M. M. Crawford, "A Hierarchical Multiclassifier System for Hyperspectral Data Analysis," *Lecture Notes in Computer Science*, F. Roli and J. Kittler, Eds., vol. 1857, pp. 270-279, 2000.

[7] S. Kumar, J. Ghosh, and M. Crawford, "Best-bases feature extraction algorithms for classification of hyperspectral data," *IEEE Trans. Geosci. Rem. Sens.*, vol. 39, issue 7, pp. 1368-1379, Jul. 2001.

[8] A. Henneguelle, *Feature Extraction for Hyperspectral Data Analysis*, Masters thesis, The University of Texas at Austin, 2002.

[9] S. Kumar and J. Ghosh, "GAMLS: A generalized framework for associative modular learning systems," in *Proc. of the Appl. and Sci. of Comp. Intel. II, Vol. 3722*, 1999, pp. 24-34.

[10] F. Glover and M. Laguna, *Tabu Search*. Boston: Kluwer Academic Publishers, 2001.

[11] M.M. Crawford, S. Kumar, M.R. Ricard, J.C. Gibeaut, and A.L. Neuenschwander, "Fusion of airborne polarimetric and interferometric SAR data for classification of coastal environments," *IEEE Trans. Geosci. Rem. Sens.*, vol. 37, issue 3, pp. 1306-1315, May 1999.