Novel Top-Down Approaches for Hierarchical Classification and Their Application to Automatic Music Genre Classification

Carlos N. Silla Jr.
University of Kent: Computing Laboratory
Kent, UK, CT2 7NZ
cns2@kent.ac.uk

Alex A. Freitas
University of Kent: Computing Laboratory
Kent, UK, CT2 7NZ
A.A.Freitas@kent.ac.uk

Abstract—This paper presents two novel hierarchical classi-fication methods which are extensions of a previously proposed selective classifier top-down approach, which consists of selecting - during the training phase - the best classifier at each node of a classifier tree. More precisely, we propose two novel selective top-down hierarchical methods. First, a method that selects the best feature set instead of the best classifier. Secondly, a method that selects both the best classifier and the best representation simultaneously. These methods are evaluated on the task of hierarchical music genre classification using four different types of feature sets extracted from each song and four classifiers.

Index Terms—Hierarchical Classification; Music Genre Clas-sification.

I. Introduction

In the field of machine learning the task of hierarchical classification appeared as an effort to classify a large number of electronic documents into document categories [1] and later was extended to classify internet pages into directory-like structures used by search engines like Yahoo! [2]. Therefore most of the work related to hierarchical classification has been done in the text categorization domain [3] [4] [5]. However, the text categorization domain is not the only one which can benefit from using a classification method that can cope with a hierarchically-organized class structure. Indeed, hierarchical classification methods have been used in other domains like protein function prediction [6] [7] [8], music genre classification [9] [10] [11] and shape classification [12].

In this paper we present two novel approaches for hier-archical classification. In essence, the first approach consists of selecting the best type of feature representation at each classifier node in the class hierarchy. This seems relevant as in hierarchical classification problems the features that most discriminate among classes tend to be different at each level of the class hierarchy. For example, books about computer science may differ from books about animals because of the common words they have (it is less likely that the words *computer* or *program* will appear often in books about animals). However if we consider books about different computer science fields, these two words will most likely not be useful to distinguish between the different fields.

The second approach proposed in this paper consists of selecting the best combination of feature representation and classifier at each classifier node in the class hierarchy. This approach is an extension of the selective classifier approach proposed in [13], where only a classifier (but not a represen-tation) is selected at each node in the class hierarchy.

The remaining of this paper is organized as follows: Sec-tion II presents the background on hierarchical classification. Section III presents related work on the field of Music Genre Classification. Section IV discusses the two new approaches that aim to improve hierarchical classification methods. Section V presents the experimental setup for the experiments. Section VI reports the results of experiments on the task of hierarchical music genre classification. Conclusions and some perspective about future work are stated in Section VII.

II. Hierarchical Classification

According to [14], [15] hierarchical classification methods differ in a number of different criteria. The first criteria is the type of hierarchical structure used. This structure can be either a tree or a DAG (Direct Acyclic Graph). Figure 1 illustrates the different structures. The main difference between them is that in the DAG a node can have more than one parent node. The second criterion is related to how deep the classification in the hierarchy is performed. I.e., the hierarchical classification method can be implemented in a way that will always assign an example to a leaf class node (which [15] refers to as Mandatory Leaf-Node Prediction and [14] refers to as Virtual Category Tree) or the method can consider stopping the classification at any level of the hierarchy (which [15] refers to as Non-Mandatory Leaf Node Prediction and [14] refers to as Category Tree). The third criterion is related to how the hierarchical structure is explored. To deal with hierarchical structure the existing approaches can be divided into three broad groups: Naïve, Big-Bang and Top-Down approaches.

One naïve approach, which is the simplest one, consists of completely ignoring the class hierarchy by predicting only classes at the leaf nodes. This approach behaves like a tra-ditional classification algorithm during training and testing. However, it provides an indirect solution to the problem of hi-
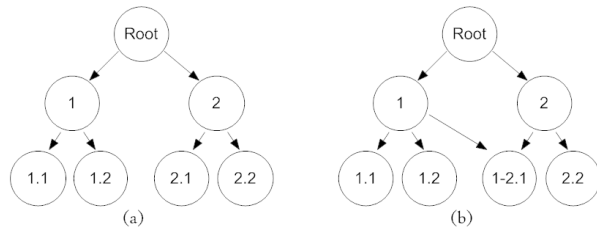
Fig. 1. Hierarchical class structures: (a) Tree (b) DAG

erarchical classification, because, when a leaf class is assigned to an example, one can consider that all its ancestor classes are also implicitly assigned to that instance. However, this very simple approach has the serious disadvantage of having to build a classifier to discriminate among a large number of classes (all leaf classes), without exploring information about parent-child class relationships present in the class hierarchy.

In the big bang approach, a single (relatively complex) classification model is built from the training set, taking into account the class hierarchy as a whole during a single run of the classification algorithm. When used during the test phase, each test example is classified by the induced model, a process that can assign classes at potentially every level of the hierarchy to the test example [15].

The Top-Down approach consists of creating a classifier for every parent node in the class hierarchy (assuming a multi-class classifier is available). In this approach the decision concerning each example's classification depends on the classes previously assigned to the example at higher class levels. Therefore a classifier is trained for each node in the class hierarchy that is not a leaf node, and the classifier's goal is to discriminate among the child classes of the classifier's corresponding node.

For instance, consider the example of Figure 1 (a), and suppose that the classifier at the root node assigns the example to the class 2. The classifier at class node 2, which was only trained with the children of the node 2, in this case 2.1 and 2.2, will make its class assignment – in this case a leaf class. Instead of using a multi-class classifier, a binary classifier can be used. The main differences between the multi-class approach and the binary one are the data preparation step and the underlying classifiers (binary vs. multi-class), as the testing phase is performed in a very similar way. In the binary approach a classifier is trained for each class node (both non-leaf and leaf nodes, with the goal of predicting whether or not an examples has the corresponding class. Note that the multi-class approach has the advantage of requiring a considerable smaller number of classifiers (since it does not need to train a classifier for every node). Hence, the multi-class approach is used in our experiments.

Among the works that use the top-down approach, most of them simply use the standard top-down approach described in this section [3] [16] [17] [10] [11] [4]. However, in [13] an extension of the top-down approach is proposed. We will refer to this method as the Selective Classifier Top-Down approach. Usually in the top-down approach the same classification algorithm is used throughout all the class hierarchy. In [13], the authors hypothesise that it would be possible to improve the predictive accuracy of the top-down approach by using different classification algorithms at different nodes of the class hierarchy. The choice of which classifier to use at a given class node is made on a data-driven manner using the training set. In order to determine which classifier should be used at each node of the class hierarchy, during the training phase, the training set is randomly split into a sub-training and validation sets. Different classifiers are then trained using this sub-training set and are then evaluated on the validation set. The classifier chosen for the current class node is the one with the highest classification accuracy on the validation set. Although the original motivation for the selective top-down approach is appealing, it can be improved, as will be discussed in section IV.

III. MUSIC GENRE CLASSIFICATION

The task of music genre classification is popular in the Music Information Retrieval (MIR) community. The different approaches to the problem can be classified as:

- Content-based (that is, features are extracted directly from the digital signal of digital audio files) [18];
- Symbolic-based (that is, features are extracted from songs in MIDI format. It is symbolic because internally the MIDI format maps which instruments came into play and for how long, which allows the computation of features on a higher level than using the audio signal) [11];
- Lyrics-based (that is the lyrics (or the lack of) are used using text mining techniques to classify a song according to their lyrics;
- Community meta-data based using web-mining techniques to extract information about an author or song from different websites or forums;
- Hybrid approaches, that have been started to be studied recently by combining more than one of the previous approaches. Existing hybrid (or multi-modal) approaches are discussed in [19], which combines the content-based and symbolic-based approaches; in [20], which combines the content, symbolic and community meta-data based approaches; and in [21], which combines the content-based and lyrics-based approaches.

Note however, that few researches have considered the use of hierarchies. Some of the works that considered the use of hierarchies were:

Burred and Lerch [22] used a musicological consistent taxonomy which has 4 levels of depth and contains 17 leaf node classes being 3 speech classes, 13 music classes and 1 background sounds class. To cope with the hierarchy they use a top-down approach using a feature selection algorithm at each split and a 3-component Gaussian Mixture Model as a base classifier at each split. In the music application domain, as well as the other domains, some features will be more suitable than others at different splits of the top-down approach.

McKay and Fujinaga [11] used a symbolic dataset considering two taxonomies: the first taxonomy has 2 levels of depth and contains 9 leaf node classes and the second taxonomy has 38 leaf node classes and has 3 levels of depth. To cope with the hierarchy they use a top-down approach using an ensemble of Feed Forward Neural Networks (FFNNs) and k-NN. The motivation for the ensemble is that they have features which are one-dimensional, e.g. average duration of melodic arcs, and multi-dimensional, e.g. the bins of a histogram consisting of the relative frequency of different melodic intervals. The set of one-dimensional features was handled by the k-NN classifier, while a FFNN classifier was employed for each set of multi-dimensional features. The final classification at each split was given by the combination of this set of classifiers. To the k-NN classifier, at each split, they applied a genetic algorithm feature selection mechanism.

Li and Ogihara [10] used two datasets with content-based features and a standard top-down approach using SVM classifiers. The dataset A is the GTZAN dataset [18] where the authors manually designed a hierarchy of depth 2 with 10 leaf-node classes and 4 intermediate nodes. The second dataset contains another manually designed hierarchy of depth 2 with 5 leaf-node classes and 2 intermediate nodes.

Brecheisen *et al.* [23] used a hierarchy of 3 levels depth, with 11 leaf node classes. Their method is an enhanced top-down approach that uses feature selection, multiple representations from the same object and also allow hierarchically multi-label classifications by using a two-layer classification process (2LCP). On the first layer they employ standard SVMs with the One-Against-One problem decomposition approach to cope with the multi-class problem. On the second layer another SVM classifier is trained in order to handle the multiple-label assignment and also to aggregate the voting vectors from the output of the classifiers on the first layer.

IV. NOVEL SELECTIVE TOP-DOWN APPROACHES

One interesting aspect of research that has been little investigated in the literature is the development and evaluation of new strategies to handle different feature representations in hierarchical classification. The motivation behind this idea arises from the questions: "Do the features used to distinguish between different classes have the same importance at different levels of the hierarchy?" Moreover, "would different types of features at different class nodes improve the classification accuracy / interpretability of the results"?

The first novel method proposed in this paper is the Selective Feature Representation Top-Down Method (by contrast to the Selective Classifier Top-Down Method proposed by [13]). The former uses a strategy similar to the one used by the latter, but instead of selecting the best classifier at each parent node in the class tree (like in [13]), it selects the best feature representation at each parent node of the class hierarchy.

The motivation behind this approach can be explained by doing an analogy with the biological taxonomy of animals. The latter is a hierarchy that consists of eight levels. At each level of the hierarchy, groups of animals are distinguished from one another by their dissimilarities. To illustrate this argument, Table I contains four animals (two kinds of cats and two kinds of horses) and their respective classifications using the biological taxonomy. The important question here is: "Are the features that distinguish between Carnivore and Perissodactyla (e.g. shape of the hoofs, type of alimentation, type of teeth), the same as the features that distinguish between a Persian Cat and a Siamese Cat (e.g. length of the fur, thickness of the fur, color of the fur, shape of the skull)?". Moreover, are the features that distinguish between the different cats the same as the ones that distinguish between the different horses (height, weight, thickness of the hair (which is referred as fur in small animals))? From this analysis, it is clear that the classification of different objects (in this example, the animals), benefits from different representations at different levels of the hierarchy. For the task of music genre classification, these differences are also present as pointed out in [22] features related to beat strength are more likely to perform better in separating classical from pop music than in classifying into chamber music sub-genres.

TABLE I
FOUR ANIMALS ACCORDING TO THEIR BIOLOGICAL TAXONOMY

	Animal			
	Persian Cat	Siamese Cat	Breton Horse	Arab Horse
Kingdom	Animalia	Animalia	Animalia	Animalia
Phylum	Chordata	Chordata	Chordata	Chordata
Class	Mammalia	Mammalia	Mammalia	Mammalia
Order	Carnivore	Carnivore	Perissodactyla	Perissodactyla
Family	Felidae	Felidae	Equidae	Equidae
Genus	*Felix*	*Felix*	*Equus*	*Equus*
Species	*F. domesticus*	*F. domesticus*	*E. caballus*	*E. caballus*
Breed	Persian	Siamese	Breton	Arab

The second method proposed in this paper is a novel combination of two methods. More precisely, instead of only using the top-down classifier selection method [13] or the top-down feature representation selection method, they are used together. One drawback of such approach is the combinatorial explosion if the number of candidate feature representations and/or the number of classifiers is significantly increased, because, during the training phase, at each class node the system will be trained with all the available classifiers considering all the different feature representations. This drawback is partially mitigated by the use of multi-class classifiers, which avoid the need for training one classifier for every node (note that in a tree the majority of the nodes are leaves), by comparison with binary classifiers – as explained earlier.

V. EXPERIMENTAL SETUP

A. Creation of the dataset

In order to create the dataset used in this experiments, we have joined two databases: the Latin Music Database [24] (LMD) and a subset of the magnatune database which was used in the ISMIR (International Conference on Music Information Retrieval) 2004 music genre classification contest. The LMD contains over 3,000 songs from 10 Latin genres (Axé: 313 songs, Bachata: 313 songs, Bolero: 315 songs, Forró: 313 songs, Gaúcha: 311 songs, Merengue: 315 songs, Pagode: 307 songs, Salsa: 311 songs, Sertaneja: 321 songs, Tango: 408
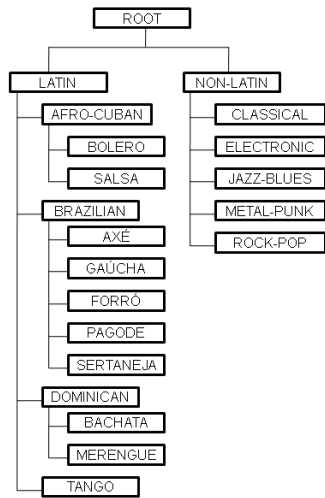
Fig. 2. Music Genre Database Class Hierarchy

songs). The magnatune database contains 1,432 songs from 6 genres (Classical: 615 songs, Electronic: 229 songs, Jazz-Blues: 52 songs, Rock-Pop: 202 songs, World: 244 songs, Metal-Punk: 90 songs). However, since the *World* genre of the magnatune database is too broad, i.e. it might even contain Latin music genres in it, we removed it from our experiments. The final hierarchical database used in the experiments has 15 leaf classes and its hierarchy is shown in Fig. 2.

B. Feature Extraction

In this work the problem of automatic music genre classification is viewed as a data mining problem where a song is represented in terms of feature vectors. The aim of feature extraction is to represent a song into a compact and descriptive way that is suitable to be dealt by machine learning algorithms.

As seen in previous flat classification experiments [25] [26] experiments, the use of different parts of the music signal affects the classification accuracy. Therefore, instead of using only one version of the dataset, we generated three versions of the dataset by extracting features from three 30-second music segments. The music segments have the same duration, which is equivalent to 1,153 audio samples in MP3 format. It is important to notice that regardless of the bitrate of the file, when dealing with MP3 files, the number of audio samples (which denotes the duration of the music) is always the same [27]. For this reason we use the following strategy to extract features from three music segments of a song:

- The first segment (Seg_{beg}) is extracted from the beginning of the song, from audio sample $s(0)$ to audio sample $s(1153)$;
- Let N denote the total number of audio samples of a song, the second segment (Seg_{mid}) is extracted from the middle of the song, from audio sample $s(N/3 + 500)$ to audio sample $s(N/3 + 1653)$;

- The third segment (Seg_{end}) is extracted from the end part of the song, but a particular strategy is adopted to avoid getting noisy or silenced endings that are common in some MP3 files. Then, the third segment is extracted from audio sample $s(N - 1453)$ to audio sample $s(N - 300)$.

For the extraction of features from the music segments, we have used four types of feature representation that are the state of art in music genre classification: The Inset-Onset Interval Histogram Coefficient (IOIHC) descriptors (40 features) [28]; Rhythm Histogram (RH) features[1] [29] (60 features); Statistical Spectrum Descriptors[1] (SSD) [29] (168 features); and the MARSYAS[2] [18] framework (30 features).

C. Classifiers used

In this work the following four classification algorithms were used: k Nearest Neighbors (k-NN) with k = 3; Naive Bayes (NB); Multi-Layer Neural Network (MLP) with the back propagation momentum algorithm; and Support Vector Machines (SVM) with pairwise classification in order to implement a multi-class SVM classifier. All the experiments were conducted using the WEKA Data mining Tool [30] with default parameters.

VI. COMPUTATIONAL RESULTS

Unfortunately, in the task of hierarchical classification there are no standard measures to evaluate the results. A comprehensive review of hierarchical classification measures can be found in [31]. In this work we have used the standard classification accuracy for flat classification adapted in a straightforward manner to the problem of hierarchical classification, by measuring the classification accuracy at the leaf level of the class hierarchy. This is acceptable because in our dataset each example is assigned exactly one leaf class (unlike datasets where some examples are not assigned a leaf class or are assigned 2 or more leaf classes). To perform the experiments we used the stratified ten-fold cross-validation procedure. To perform the selection of the best classifier, best feature representation or their combination, we divide the training set into sub-training (80%) and validation sets (20%), using stratified randomly selected examples.

Table II shows the results for the beginning segment of the songs. Let us compare the results for different feature representations (across the columns) for each classifier approach (row). As shown in the table, in the cases where we use the Standard Top-Down (TD) approach with a single fixed classifier throughout the class hierarchy (first four rows in the table), the Selective Feature Representation Top-Down (S.R.) approach (last column) improved the predictive accuracy for all classifiers, and the improvement was statiscally significant in the vast majority of the cases. To show this, in these first four rows, in the first four columns (for each of the four fixed representations), each cell has the symbol "∗" if the accuracy reported in that cell is statiscally significantly smaller than

---

[1] Available at: http://www.ifs.tuwien.ac.at/mir/audiofeatureextraction.html
[2] Available at: http://marsyas.sness.net/

the accuracy reported in the last column (selective representation) for the corresponding row. Statistical significance was measured by the paired two-tailed Student's t-test, using a confidence level of 95%. This statistical test was also used in the analysis of all other results reported in this section.

Let us consider now the last row in Table II. Each of the first four results in that row refers to the use of the Selective Classifier Top-Down (S.C.) approach with a fixed feature representation throughout the class hierarchy. The use of this Selective Classifier approach improved the predictive accuracy for almost all representations (the exception was for MARSYAS, where TD MLP obtained a better accuracy than the selective classifier approach). The improvement was statistically significant in the vast majority of the cases. To show this, in each cell of the first four rows and first four columns the symbol "§" is inserted if the accuracy for the fixed classifier is significantly smaller than the accuracy for the selective classifier approach in the corresponding column.

Finally, a very positive result emerges when we observe the accuracy value at the last cell of the last row, which refers to the simultaneous use of both the Selective Classifier and the Selective Feature Representation Top-Down approaches. This accuracy is higher than all the other 24 accuracies reported in the table, i.e., using both types of Selective Top-Down approaches simultaneously improved the accuracy over all cases of: (a) fixed classifier and fixed representation; (b) fixed classifier and selective representation, and (c) selective classifier and fixed representation. To analyze the statistical significance of these results, the symbol "†" is inserted in each cell where the result of the corresponding approach is significantly lower than the accuracy of selecting both classifiers and representations. As can be seen in Table II, the latter approach obtained significantly better accuracies in 23 out of 24 cases.

Table III shows the results for the middle segment of the songs. The Selective Feature Representation Top-Down (last column) approach improved the predictive accuracy for all classifiers with one exception (TD k-NN with SSD feature representation), and the improvement was statiscally significant in the vast majority of the cases (13 out of 16). Again, the symbol "*" is used to indicate that the accuracy with a fixed representation is statistically significantly smaller than the accuracy with the selective representation approach.

Let us consider now the last row in Table III. Each of the first four results in that row refers to the use of the Selective Classifier Top-Down approach with a fixed feature representation throughout the class hierarchy. The use of this Selective Classifier approach improved the predictive accuracy for almost all representations (the exceptions were for IOIHC and MARSYAS, where TD MLP obtained better accuracies than the selective classifier approach). Again, the symbol "§" is used to indicate that the accuracy with a fixed classifier is statistically significantly smaller than the accuracy with the selective classifier approach in the same column.

The simultaneous use of both the Selective Classifier and the Selective Feature Representation Top-Down approaches

(last row, last column) yields a higher accuracy than all the other 24 accuracies reported in the table. Again, the symbol "†" is used to indicate that this approach has significantly better accuracies than the other approaches, which was observed in 21 out of 24 cases.

Table IV shows the results for the end segment of the songs. Unlike the other segments, it seems that the IOIHC features are not very predictive for the end part of the song and this has an impact on the results of the proposed methods.

The use of the Selective Classifier approach is unaffected by this fact (as it does not deal with the use of different representations) and improved the predictive accuracy for almost all representations (the exceptions were for IOIHC and MARSYAS, where TD MLP obtained a better accuracy than the selective classifier approach).

However, the accuracy with the Selective Representation (last column) approach is the same as the accuracy with a fixed representation in two cases: TD k-NN and TD MLP both using the SSD representation. The reason for this is that although in the previous cases the IOIHC features are not the most discriminative features overall, they were often used to discriminate between Salsa and Bolero. Since the performance of the IOIHC representation is degraded at the end part of the song, the selective representation approach ends up always using only one type of feature for both these classifiers. This is an interesting finding as it shows how the selective representation indeed exploits the best feature representation to distinguish between different classes at different levels of the class hierarchy. Moreover, the selective representation approach still obtains better results than a fixed representation

TABLE II
PREDICTIVE ACCURACY (%) FOR Seg_{beg}.

Classifiers	IOIHC	MARSYAS	RH	SSD	S.R.
TD k-NN	43.52*†	48.66*§†	43.23*†	64.52*§†	64.81†
TD NB	28.60*§†	46.46*§†	32.61*§†	39.97*§†	47.54†
TD MLP	45.00*†	61.51*†	41.03*§†	66.14§†	66.91†
TD SVM	37.75*§†	57.06*§†	42.01*§†	67.40§†	67.38†
S.C.	44.55†	59.92†	44.77†	69.20	69.39

TABLE III
PREDICTIVE ACCURACY (%) FOR Seg_{mid}.

Classifiers	IOIHC	MARSYAS	RH	SSD	S.R.
TD k-NN	49.00*§†	54.03*§†	50.71*†	71.68§†	71.64†
TD NB	34.47*§†	52.48*§†	36.93*§†	48.31*§†	54.24†
TD MLP	51.04*†	66.32*§†	48.54*§†	76.96	77.99
TD SVM	43.26*§†	61.91*§†	49.66*§†	76.12§†	76.23†
S.C.	50.97†	64.97†	51.14†	78.70	78.82

TABLE IV
PREDICTIVE ACCURACY (%) FOR Seg_{end}.

Classifiers	IOIHC	MARSYAS	RH	SSD	S.R.
TD k-NN	27.84*†	50.87*§†	47.43*†	69.97§†	69.97†
TD NB	13.65*§†	49.68*§†	35.36*§†	46.10*§†	50.78†
TD MLP	27.16*†	63.04*†	44.65*§†	72.39	72.97
TD SVM	12.42*§†	59.53*†	44.88*§†	72.85	72.85
S.C.	26.59†	61.77†	47.71†	73.50	73.50

in 14 out of 16 (the other 2 case being a draw) of which 13 are significantly better (marked by "*").

Also, the use of both the Selective Classifier and the Selective Representation approaches (last row, last column) had the same result as using only the selective classifier approach with the SSD representation. The reason for this is the same explained for the selective representation approach. In any case, the use of both selective classifier and selective representation approaches still has better results in 23 out of 24 cases of which 19 are significantly better (marked by "†").

VII. CONCLUDING REMARKS

In this work we presented two novel top-down approaches for hierarchical classification. These methods were evaluated on the task of hierarchical music genre classification with a dataset of over 4,000 songs and 15 leaf classes. An analysis of the experimental results shows that the novel approaches significantly improve the classification accuracy when compared to the standard top-down approach in the vast majority of cases. Moreover, the proposed methods benefit from dynamically selecting the best feature representation for each class node in the class hierarchy instead of using only one fixed feature representation throughout the class hierarchy. As future work, we plan to investigate the use of feature selection algorithms for hierarchical classification.

ACKNOWLEDGMENTS

REFERENCES

[1] M. Sasaki and K. Kita, "Rule-based text categorization using hierarchical categories," in *Proc. of IEEE Int. Conf. on Systems, Man, and Cybernetics*, 1998, pp. 2827–2830.

[2] Y. Labrou and T. Finin, "Yahoo! as an ontology – using yahoo! categories to describe documents," in *Proc. of the ACM Conf. on Information and Knowledge Management*, 1999, pp. 180–187.

[3] S. T. Dumais and H. Chen, "Hierarchical classification of Web content," in *Proc. of the 23rd ACM Int. Conf. on Research and Development in Information Retrieval*, N. J. Belkin, P. Ingwersen, and M.-K. Leong, Eds., 2000, pp. 256–263.

[4] D. Tikk, G. Biró, and J. D. Yang, "A hierarchical text categorization approach and its application to frt expansion," *Australian Journal of Intelligent Information Processing Systems*, vol. 8, no. 3, pp. 123–131, 2004.

[5] K. Wang, S. Zhou, and Y. He, "Hierarchical classification of real life documents," in *Proc. of the 1st SIAM Int. Conf. on Data Mining*, Chicago, US, 2001.

[6] Z. Barutcuoglu, R. E. Schapire, and O. G. Troyanskaya, "Hierarchical multi-label prediction of gene function," *Systems Biology*, vol. 22, pp. 830–836, 2006.

[7] A. Clare and R. D. King, "Knowledge discovery in multi-label phenotype data," in *5th European Conf. on Principles of Data Mining and Knowledge Discovery*, ser. Lecture Notes in Computer Science, L. De Raedt and A. Siebes, Eds., vol. 2168.   Springer, 2001, pp. 42–53.

[8] N. Holden and A. A. Freitas, "Improving the performance of hierarchical classification with swarm intelligence," in *Proc. 6th European Conf. on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics (EvoBio-2008)*, ser. Lecture Notes in Computer Science, vol. 4973.   Springer, 2008, pp. 48–60.

[9] C. DeCoro, Z. Barutcuoglu, and R. Fiebrink, "Bayesian aggregation for hierarchical genre classification," in *Proc. of the 8th Int. Conf. on Music Information Retrieval*, Vienna, Austria, 2007, pp. 77–80.

[10] T. Li and M. Ogihara, "Music genre classification with taxonomy," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2005, pp. 197–200.

[11] C. McKay and I. Fujinaga, "Automatic genre classification using large high-level musical feature sets," in *Proc. of the Int. Conf. on Music Information Retrieval*, 2004, pp. 525–530.

[12] Z. Barutcuoglu and C. DeCoro, "Hierarchical shape classification using bayesian aggregation," in *Proc. of the IEEE Conf. on Shape Modeling and Applications*, 2006.

[13] A. Secker, M. Davies, A. Freitas, J. Timmis, M. Mendao, and D. Flower, "An experimental comparison of classification algorithms for the hierarchical prediction of protein function," *Expert Update (the BCS-SGAI Magazine)*, vol. 9, no. 3, pp. 17–22, 2007.

[14] A. Sun and E.-P. Lim, "Hierarchical text classification and evaluation," in *Proc. of the IEEE Int. Conf. on Data Mining*, 2001, pp. 521–528.

[15] A. A. Freitas and A. C. P. L. F. de Carvalho, *Research and Trends in Data Mining Technologies and Applications*.   Idea Group, 2007, ch. A Tutorial on Hierarchical Classification with Applications in Bioinformatics, pp. 175–208.

[16] N. Holden and A. A. Freitas, "A hybrid particle swarm/ant colony algorithm for the classification of hierarchical biological data," in *Proc. of the IEEE Swarm Intelligence Symposium*, 2005, pp. 100–107.

[17] ——, "Hierarchical classification of g-protein-coupled receptors with a pso/aco algorithm," in *Proc. of the IEEE Swarm Intelligence Symposium*, 2006, pp. 77–84.

[18] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[19] T. Lidy, A. Rauber, A. Pertusa, and J. M. Inesta, "Improving genre classification by combination of audio and symbolic descriptors using a transcription system," in *Proc. of the 8th Int. Conf. on Music Information Retrieval*, 2007, pp. 23–27.

[20] C. McKay and I. Fujinaga, "Combining features extracted from audio, symbolic and cultural sources," in *Proc. of the 9th Int. Conf. on Music Information Retrieval*, 2008, pp. 597–602.

[21] R. Mayer, R. Neumayer, and A. Rauber, "Combination of audio and lyrics features for genre classification in digital audio collections," in *Proc. of the ACM 16th Int. Conf. on Multimedia*, 2008, pp. 159–168.

[22] J. J. Burred and A. Lerch, "A hierarchical approach to automatic musical genre classification," in *Proc. of the 6th Int. Conf. on Digital Audio Effects*, 2003, pp. 8–11.

[23] S. Brecheisen, H.-P. Kriegel, P. Kunath, and A. Pryakhin, "Hierarchical genre classification for large music collections," in *Proc. of the IEEE 7th Int. Conf. on Multimedia & Expo*, 2006, pp. 1385–1388.

[24] C. N. Silla Jr., A. L. Koerich, and C. A. A. Kaestner, "The latin music database," in *Proc. of the 9th Int. Conf. on Music Information Retrieval*, 2008, pp. 451–456.

[25] C. N. Silla Jr., C. A. A. Kaestner, and A. L. Koerich, "Automatic music genre classification using ensemble of classifiers," in *Proc. of the IEEE Int. Conf. on Systems, Man, and Cybernetics*, 2007, pp. 1687–1692.

[26] C. N. Silla Jr., A. L. Koerich, and C. A. A. Kaestner, "A machine learning approach to automatic music genre classification," *Journal of the Brazilian Computer Society*, vol. 14, no. 3, pp. 7–18, 2008.

[27] S. Hacker, *MP3: The Definitive Guide*, 1st ed.   O'Reilly, 2000.

[28] F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer, "Evaluating rhythmic descriptors for musical genre classification," in *Proc. of 25th Int. AES Conf.*, 2004, pp. 196–204.

[29] T. Lidy and A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification," in *Proc. of the 6th Int. Conf. on Music Information Retrieval*, 2005, pp. 11–15.

[30] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd ed.   San Francisco: Morgan Kaufmann, 2005.

[31] E. Costa, A. Lorena, A. Carvalho, and A. Freitas., "A review of performance evaluation measures for hierarchical classifiers," in *Evaluation Methods for Machine Learning II: papers from the 2007 AAAI Workshop*.   AAAI Press, 2007, pp. 1–6.