Multi-Label Hierarchical Classification for Protein Function Prediction

Helyane B. Borges and Julio Cesar Nievola

***Abstract*—**Hierarchical classification is a problem with applications in many areas as protein function prediction where the dates are hierarchically structured. Therefore, it is necessary the development of algorithms able to induce hierarchical classification models. This paper presents experimenters using the algorithm for hierarchical classification called Multi-label Hierarchical Classification using a Competitive Neural Network (MHC-CNN). It was tested in ten datasets the Gene Ontology (GO) Cellular Component Domain. The results are compared with the Clus-HMC and Clus-HSC using the hF-Measure.

***Keywords*—**Hierarchical Classification, Competitive Neural Network, Global Classifier.

I. Introduction

HIERARCHICAL classification is a task of data mining that has been applied in diverse areas such as the music prediction [28], [29], [4], images [30], text (work place) among others. In bioinformatics, it has been used for functional prediction of proteins, since this is not an easy task to accomplish without the help of efficient techniques.

The prediction of protein functions can be treated as a classification problem in data mining, in which proteins attributes are considered a sample in the database and its biological functions as classes (multi-class classifiers) [2].

Most algorithms for multi-label hierarchical classification of proteins have been developed to support class hierarchies with a tree structure, but the use of ontology in predicting protein functions has been used as in the case of Gene Ontology (GO) [13], [16], [20]. The GO terms are hierarchies structured as a directed acyclic graph (DAG), in which a "child" term may be connected to one or more "parents" terms. The classification algorithms developed to support this type of structure typically do not assess the hierarchical model as a whole (global or big-bang approach), which may change the predictive results of the samples.

In this paper an algorithm for hierarchical classification of data for structures such as DAG, developed by Borges and Nievola [26] denominated of MHC-CNN (Multi-label Hierarchical Classification using a Competitive Neural Network) is applied. The experiments are focuses on hierarchical protein function prediction using GO Cellular Component Domain as the aim to verify the comportment of the classifier.

Helyane B. Borges is with the Universidade Tecnológica Federal do Paraná. Brasil (e-mail: helyane@utfpr.edu.br).

Julio Cesar Nievola is with the Pontifícia Universidade Católica do Paraná. Brasil (e-mail: nievola@ppgia.pucpr.br).

II. Hierarchical Classification

Hierarchical classification differs from flat classification because the classes are organized in a hierarchy structured as a tree or a DAG where the nodes of this hierarchy represent the classes that are involved in the classification process.

The main difference between the tree structure and the DAG structure is that in the tree structure each node (each class), except the root node, has only one ancestor (parent), while in the DAG structure each node (class) can have one or more ancestors nodes.

Another characteristic that differs flat classification from hierarchical classification refers to the type of prediction of classes in the hierarchy, which can be distinguished into two categories: mandatory leaf node (possible in flat or hierarchical classification) prediction and non-mandatory leaf node (possible only in hierarchical classification).

In mandatory leaf node prediction all examples should be associated with classes represented by leaf nodes. In the non-mandatory leaf node prediction there is no requirement that the prediction occurs at leaf nodes. Thus, the examples may be associated with classes that are represented by any internal node of the class hierarchy along with their ancestors.

To explore hierarchical classification problems some solutions have been proposed, which can be divided into three main approaches: flat hierarchical classification, local hierarchical classification and global hierarchical classification [4]. These approaches describe how the classifiers are built and not a classification method, such as top-down approach that is often cited in literature as being one of the approaches.

A. Flat Hierarchical Classification

The flat hierarchical classification has the same behavior of a conventional classification algorithm in the training and testing phases. This approach considers that a hierarchical classification problem can be transformed into a flat classification problem disregarding the concept of ancestor and descendant, i.e., it ignores the class hierarchy, predicting only the leaf nodes. This approach is similar to conventional flat classification and can be applied to tree and DAG structures.

B. Local Hierarchical Classification

The local hierarchical classification consists of using M independent local classifiers, each one dealing with the prediction of only one of the classes (M is the total number of nodes in the class hierarchy) [13]. Hence, the number of classifiers that should be trained could be huge in situations where there are a lot of classes.

This approach also has the advantage that each classification model is built using a process of modularization as the local classifier per node approach. The same problem can occur if a node class has been wrongly down-propagated to the following levels of hierarchy.

In general, for local hierarchical classifiers, evaluation uses the top-down method. It starts predicting the class of the first level, and then the class predicted at the next level in the hierarchy is chosen only among the classes which have the previous one as its parent class. This process is repeated for classes at deeper levels.

C. Global Hierarchical Classification

The global or big-bang hierarchical classification approach builds a single classification model considering the class hierarchy, based on the training set. In this approach, the prediction can occur at any level of hierarchy and the algorithm respect the hierarchical structure of the classes through the relationship between ancestors and descendants classes. Thus, none of the approaches used for flat classification can be used, without changing the classifier.

The main advantages of this approach are that there is no need to train a large number of classifiers and the automatic manipulation of inconsistency in the prediction of classes. Its main disadvantage is the increased complexity of the single global classifier.

III. Related Works

The hierarchical classification has been widely used in text mining since the 90s. Among the work in this context can be cited Koller and Sahami [21], Sun and Lim [1], Kiritchenko et al [10].

The field of bioinformatics presents several problems to be solved by hierarchical classification, but it is, unfortunately, still little explored. Some works have been published using this approach, specifically in the protein function prediction, but using a hierarchical tree structure [18], [17], [5], [19], [6].

There are very few works using structures DAG in the protein function prediction that consider the class hierarchy (global approach).

Vens et al. [3] developed a hierarchical classification model for the DAG structure using the global or big-bang approach. In this work the authors discuss three kinds of classification: single-label classification (SC), hierarchical single-label classification (HSC) and multi-label hierarchical classification (HMC). For the development of these classifiers the authors used the induction of decision trees and showed how this model can be modified for use in hierarchical DAG structures.

These approaches are implemented in the Clus and consist of generating a single decision tree for the whole hierarchy. This induction algorithm of decision tree is based on the framework Predictive Clustering Trees (PCT) [5].

Aleksovski et al. [22] extended the Clus HMLC developed by Vens et al [3], using other distance measures. The measures used by the authors were Jaccard distance, and SimGIC ImageCLEF. Such measures have been implemented in CLUS. The Clus as far as we know, was the first global classifiers algorithm for DAG structures, developed specifically to resolve problems in the bioinformatics area. Because it was developed based on a decision tree it has the advantage of producing models that can be somehow interpreted by humans.

Otero et al [23] develop a new Ant Colony Optimization algorithm, named hAnt-Miner, for the hierarchical classification problem of predicting protein functions using the GO. The algorithm proposed discovers a single global classification model in the form of an ordered list of if-then classification rules which can predict GO terms at all levels of the GO hierarchy, satisfying the parent-child relationships between GO terms.

Schietgat et al. [32] extended the work of Vens et al. [3] building a multi-label hierarchical classifier called Clus-HMC-ENS that generates a set of trees. The algorithm developed by Vens et al. [3] generates a single tree that provides, for a given gene, its biological functions from a ranking function.

Otero et al. [31] proposed an extension of the classifier hant-Miner to apply in multi-label problems. The classifier is called hmAnt-Miner (Multi-label Hierarchical Classification Ant-Miner). This algorithm finds a single classification model, by means of a list of if-then rules which can predict all classes of the hierarchy. The hmAnt-Miner employs a distance measurement based on the procedure of discretization dynamic of the continuous attributes and heuristic information in the construction of ACO graph. The entropy used in hant-Miner is replaced by the distance measure in hmAnt-Miner, which is a more appropriate measure for multi-label hierarchical classification.

Alves et al. [7] constructed a hierarchical classification model called Hierarchical Multi-Label Classification with an Artificial Immune System (MHCAIS), which uses concepts of an Artificial Immune System (AIS). This hierarchical classifier aims to discover knowledge represented as rules if-then. The author presents two versions of MHCAIS: global and local. The local version builds a classifier for each class, while in the global version a single classifier is generated to distinguish all classes of the application.

Borges and Nievola [24], [25] developed an algorithm for hierarchical classification using the global approach, called Hierarchical Classification using a Competitive Neural Network (HC-CNN) for Protein Function Prediction. This algorithm is based on a Competitive Neural Network. The system was tested in eight datasets based on Funcat using different evaluation measures: distance, precision, recall and HF-measure. The disadvantage of this algorithm is that only makes predictions' of a single label.

Borges and Nievola [26] made changes to the algorithm HC-CNN allowing the algorithm is capable of multi-label prediction samples. The algorithm is called Multi-label Hierarchical Classification using a Competitive Neural Network (MHC-CNN). The experimenters are realized in five GO databases.

IV. Multi-Label Hierarchical Using a Competitive Neural network (MHC-CNN)

The MHC-CNN algorithm used in this paper is based on a Competitive Artificial Neural Network [8], [11]. Fig. 1 shows a neural network model with one layer where each output neuron represents a class. The term "Input neurons" defined in the figure represents the attributes of the instances, according to the interpretation of the data set. In the processing layer or output layer, which in a competitive network is the output mapping, is represented the hierarchy classes where each neuron is connected to its ancestor(s) and possibly descendants.

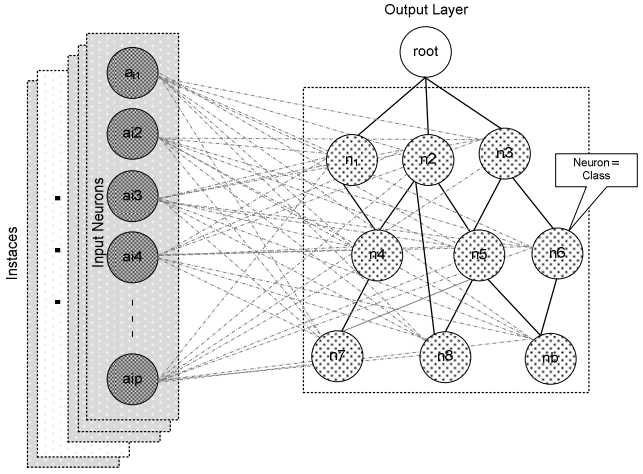

Fig. 1 Example of the MHC-CNN

In the traditional competitive network, for example, the Kohonen network, the neurons of the output layer are arranged in a grid network [8], which can be rectangular, hexagonal, among others, and they represent the network topology. In MHC-CNN algorithm the topology is a tree, where each neuron is connected with ancestors (parents) and descendent (children) neurons. These neurons (output layer) are created according to the number of classes in the hierarchy, and each neuron in the output layer is connected to all neurons of the input layer.

Neurons are stimulated by the input examples during the competitive process. In this way, it will be considered the "winner" the neuron who is more similar to the input instance selected. The comparison is made through the use of distance measures.

Prior to the training, some parameters should be defined, for instance, the amount of epochs to train the neural network and its learning rate (initial and final) which will decrease exponentially during the training, and the synaptic weights are randomly initialized. The training process of the network is divided into three phases, as in a traditional competitive network: Competition, Cooperation and Adaptation. Details of the operation of the algorithm are shown in [26].

V. Evaluation Measures

A. Distance-Based Depth-Dependent Measures

When evaluating the result of a hierarchical prediction three situations may occur: correct prediction, partially correct prediction and incorrect prediction [26].

B. Hierarchy Based Measure

This kind of measure was developed by Kiritchenko et al. [9] and uses concepts of ancestral and descendant classes. The author proposes two evaluation measures: hierarchical precision and hierarchical recall which take into account the hierarchical relationships [11], [12]. These measures are based on conventional measures of precision and recall.

VI. Experimenters and Results

The databases used were the same used by Vens et al. [3] in their experiments. In the present experiments the domain cellular component was used. Table I shows the characteristic of the selected database. For the all experiments 2/3 of the examples were used for training and 1/3 for testing (hold-out procedure) as used by Vens et al. [3]. In addition, all sets were normalized using the Min-Max approach.

TABLE I
Characteristic the Databases on the Domain Cellular Component

Database	Amount Samples	Amount Atributes	Amount Class	Amount Min/Max Class per Sample
Cellcycle	3751	77	547	1/9
Church	3749	27	547	1/9
Derisi	3719	63	547	1/9
Einsen	2418	79	547	1/9
Expr	3773	551	547	1/9
Gasch1	93	173	547	1/9
Gasch2	3773	52	547	1/9
Pheno	1586	69	462	1/7
Seq	3900	478	547	1/9
Spo	3697	80	547	1/9

In addition, the attributes missing in the database were imputed. The criterion used for imputation of missing attribute values was to calculate the arithmetic average of the closer ancestor classes to which belongs the sample that has the missing attribute.

The initial learning rate and final learning rate used in the experiments were 0.1 and 0.01, respectively. The neural network synaptic weights were generated randomly, according to a uniform distribution. The evaluation of the classification was made taking into account all levels of the hierarchy.

The results at the MHC-CNN were compared with the Clus-HMC and Clus-HSC using distance-based depth dependent measure and hierarchy-based measures.

For the Clus experiments fifty-one thresholds between 0 and 1 were used. To compare the performance of the algorithms two thresholds were selected: 0.5 and 0.1. Fig. 2 shows the results obtained with 50 epochs for training the neural network and the thresholds selected to Clus-HMC and Clus-HSC.

Note that the distance measure obtained by the MHC-CNN obtained the best results comparing to others measures. One explanation for this behavior is that the distance measure assigns weights to the classes in the hierarchy, i.e., classes that were predicted at deeper levels tend to receive a value smaller than those predicted at levels closer to the root.

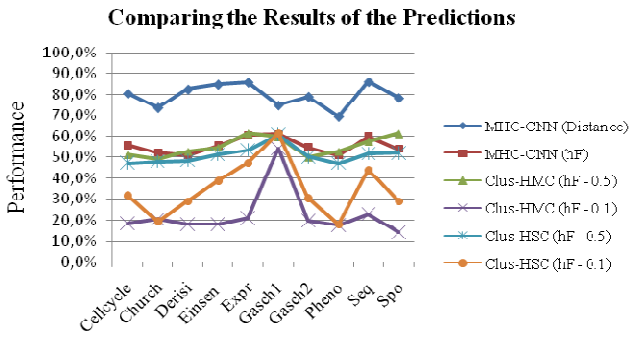

Fig. 2 Results of the Predictions

An observation to be made is that in all the databases classes with very few instances were not removed from the class hierarchy. This may explain the low results presented, since it is hard for a classifier to be able to predict a lot of classes and get good results.

The results were statistically compared using the Friedman [14], [15] and Nemenyi [27] test to verify whether there is statistical significance between the differences the performances of the algorithms. The results, according to these tests, were that there is not a significant difference between the algorithms.

VII. Conclusion

This paper has presented the MHC-CNN algorithm for the hierarchical classification problem of predicting protein functions using the GO. The algorithm proposed is based in a competitive neural network.

Classification global approach used by the classifier has the advantage of reporting a single result.

The MHC-CNN was applied in ten datasets in the cellular component domain. The results of the predictions were assessed using two approaches to hierarchical classification measures: distance-based depth-dependent measure and hierarchical measured.

Furthermore, the algorithm is able to predict samples with many classes, which is also one of the major problems is respect the proteins prediction.

References

[1] A. Sum; E. Lim. Hierarchical Text Classification and Evaluation. In Proceedings of the International Conference on Data Mining (ICDM 2001), California, USA, Nov, 2001. p. 521-528.

[2] A. Freitas; A. C. Carvalho. “A Tutorial on Hierarchical Classification with Applications in Bioinformatics”. In Research and Trends in Data Mining Technologies and Applications, chapter VII, pp. 175–208. Idea Group, 2007.

[3] C. Vens, J. Struyf, L. Schietgat, S. D˘zeroski e H. Blockeel. Decision trees for hierarchical multi-label classification. Machine Learning, vol. 73, pp. 185–214, 2008.

[4] C. Silla and A. A. Freitas. A survey of hierarchical classification across different application domains. Data Mining and Knowledge Discovery. Abr, 2010.

[5] H. Blockeel, M. Bruynooghe, S. Dzeroski, J. Ramon e J. Struyf. Hierarchical multi-classification. In Workshop on Multi-Relational Data Mining, pp. 21–35, 2002.

[6] N. Holden; A. A. Freitas. A hybrid particle swarm/ant colony algorithm for the classification of hierarchical biological data. In Swarm Inteligence Symposium, 2005 Proceedings 2005 IEEE, p. 100-107.

[7] R. T. Alves; M. R. Delgado; A. A. Freitas. A. A. Multi-label hierarchical classification of protein functions with artificial immune systems. In Proc. Advances in Bioinformatics and Computational Biology, 2008 v. 5167, p.1-12.

[8] S. Haykin. Redes neurais: princípios e prática. 2.ed. Tradução de, Paulo Martins Engel. Porto Alegre: Bookman, 2001.

[9] S. Kiritchenko; S. Matwin; F. Famili. Hierarchical Text Categorization as a Tool of Associating Genes with Gene Ontology Codes. In European Workshop on Data Mining and Text Mining in Bioinformatics, pp. 30–34, 2004.

[10] S. Kiritchenko; S. Matwin; F. Famili; R. Nock. Learning and evaluation in the presence of class hierarchies: Application to text categorization. In Proc. of the 19th Canadian Conf. on Artificial Intelligence, Lecture Notes in Artificial Intelligence, 2006. v. 4013, p. 395-406.

[11] T. Kohonen. The Self-Organizing Map. Proceedings of IEEE. v.78, n.9. p-1464-1480. 1990.

[12] Y. Guan et al. Predicting gene function in a hierarchical context with an ensemble of classifiers. Genome Biology, v. 9, p. 1-18. June, 2008.

[13] Barutcuoglu, Z., Schapire, R. E. & Troyanskaya, O. G. Hierarchical multi-label prediction of gene function. Bioinformatics. v. 22 n. 7, p. 830–836. 2006.

[14] M. Friedman. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, v. 32, p. 675–701.

[15] M. Friedman. 1940. A comparison of alternative tests of significance for the problem of m rankings. In: Annals of Mathematical Statistics. v. 11, p. 86–92.

[16] K. Tu; H. Yu; Z. Guo; X. Li. Learnability-based further prediction of gene functions in Gene Ontology. Genomics. v. 84, p. 922-928. 2004.

[17] Jensen, L. J. et al. 2003. Prediction of human protein function according to Gene Ontology categories. Bioinformatics. v. 19, n. 5, p. 635–642.

[18] A. Clare; R. D. King. Knowledge discovery in multi-label phenotype data. In Proceedings of the 5th European Conference on Principles and Practice of Knowledge Discovery and Data Mining (PKDD-2001), Freiburg, Germany. p. 42-53, 2001.

[19] A. Clare; R. D. King. “Predicting gene function in Saccharomyces cerevisiae”. Bioinformatics, vol. 19, pp. 42–49, 2003.

[20] B. Jin; B. Muller; Zhai. C; Lu. X. Multi-label literature classification based on the gene ontology graph. BMC Bioinformatics. v. 9, n. 525, p. 1-15, 2008.

[21] D. Koller; M. Sahami. Hierarchically classifying documents using very few words. In Proc. of the 14th Int. Conf. on Machine Learning (ICML 1997), San Francisco, CA, USA, p. 170-178, 1997.

[22] D, Aleksovki; D, Kocev; S. Dzeroski. Evaluation of distance measures for hierarchical multilabel classification in functional genomics. In Proc. of the 1st Workshop on Learning from Multi-Label Data (MLD), p. 5-16, 2009.

[23] F. Otero; A. Freitas; C. Johnson. A hierarchical classification ant colony algorithm for predicting gene ontology terms. In European Conference on Evolutionary Computation, achine Learning and Data Mining in Bioinformatics, volume LNCS, p. 68–79. Springer, 2009.

[24] H. B. Borges; J, C. Nievola. Hierarchical Classification using a Competitive Neural Network. In: 8th International Conference on Natural Computation (ICNC'12), 2012, Chongqing, China. 8th International Conference on Natural Computation (ICNC'12). Piscataway, NJ : IEEE Press, 2012. v. 1. p. 1-6.

[25] H. B. Borges; J, C. Nievola. Hierarchical Classification Using a Competitive Neural Network for Protein Function Prediction. In: 14th International Conference on Artificial Intelligence (ICAI'12), 2012, Las Vegas. 14th International Conference on Artificial Intelligence (ICAI'12). USA : CSREA Press, 2012. v. 1. p. 1-7.

[26] H. B. Borges; J, C. Nievola. Multi-Label Hierarchical Classification using a Competitive Neural Network for Protein Function Prediction. In:

2012 International Joint Conference on Neural Networks (IJCNN 2012), 2012, Brisbane, Austrália. 2012 International Joint Conference on Neural Networks (IJCNN 2012). Piscataway, NJ : IEEE Press, 2012. v. 1. p. 1-8.
[27] J. Desmar. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Resarch, 7:1-30, 2006.
[28] J. J. Burred; A. Lerch. A hierarchical approach to automatic musical genre classification. In Proc. of the 6th Int. Conf. on Digital Audio Effects, pp. 8-11, 2003.
[29] C. Decoro; Z. Barutcuoglu; R. Fiebrink. Bayesiana aggregation for hierarchical gene classification. In: Proc. of the 8th Int. Conf. on Music Information Retrieval, pp. 77-80, 2007.
[30] I. Dimitrovski; D. Kocev; S. Loskovska; S. Dzeroski. Hierarchical annotation of medical images. In Proc. of the 11th Int. Multiconference Information Society, A:174-177, 2008.
[31] F. Otero; A. Freitas; C. Johnson. A hierarchical multi-label classification ant colony algorithm for protein function prediction, Memetic Computing, vol. 2, pp. 165–181, 2010.
[32] L. Schietgat, C. Vens, J. Struyf, H. Blockeel, D. Kocev, and S. Dzeroski, "Predicting gene function using hierarchical multi-label decision tree ensembles", presented at BMC Bioinformatics, 2010.