A Review on enhanced Multilabel Classification Techniques using Artificial Neural Networks

G.NIRMALA JOYCEE [1] , **P.PRASHANTHI** [2]

[1,2] Assistant Professor ,
Department of Computer Science,
St.Ann's College for Women (A),Mehdipatnam,
Osmania University Hyderabad, India

Abstract: This paper considers the multilabel classification problem, which is a generalization of traditional two-class or multi-class classification problem. In multilabel classification a set of labels (categories) is given and each training instance is associated with a subset of this label-set. The task is to output the appropriate subset of labels (generally of unknown size) for a given, unknown testing instance. Some improvements to the existing neural network multilabel classification algorithm, named BP-MLL, are proposed here. The modifications concern the form of the global error function used in BP-MLL. The modified classification system is tested in the domain of functional genomics, on the yeast genome data set. Experimental results show that proposed modifications visibly improve the performance of the neural network based multilabel classifier. The results are statistically significant.

Keywords: Classifications,Artificial Neural Networks,Machine Learning.

1. Introduction

Classification in machine learning is the problem of identifying the function f(x) that maps each attribute vector x_i to its associated target label y_i, $i = 1,2,….,n$, where n is the total number of training samples [1]. Traditional classification problems in machine learning involve associating each of the sample instances with a single target label. i.e. unique target association. This type of classification is called single label classification. On the contrary, several real world classification problems involve data samples which correspond to a subset of target labels [6]. This results in the emergence of a new category of machine learning classification called the multi-label classification. The multi-label classification problems are gaining much importance and attention in the recent years due to the rapidly increasing real world application areas. Some of the real world application domains that require multi-label classification are medical diagnosis, text categorization [4-8],genomics, bioinformatics, multimedia, emotion, music categorization, scene and video categorization, map labeling, marketing etc. Due to the omnipresence of multi-label problems in a wide range of real world scenarios, multi-label classification is an emerging field in machine learning classification [15].

Classification errors occur when the classes overlap in the selected feature space. Various classification methods have been developed to provide [9] different operating characteristics, including linear discriminate functions, artificial neural networks (ANN), k-nearest-neighbor (k-NN), radial basis functions (RBF) and support vector machines (SVM).

2. Background

The sparse literature on multi-label classification is primarily geared to text classification or bioinformatics. For text classification, Schapire and Singer [21] proposed BoosTexter, extending AdaBoost to handle multi-label text categorization. However, they note that controlling complexity due to overfitting [7] in their model is an open issue. McCallum [14] proposed a mixture model trained by EM, selecting the most probable set of labels from the power set of possible classes and using heuristics to overcome the associated computational complexity [12]. However, his generative model is based on learning text frequencies in documents, and is thus specific to text applications. Bosubabu Sambana' approach is most similar to ours in that he uses a set of binary SVM classifiers [14]. He finds that SVM classfiers achieve higher accuracy than others. However, he does not discuss multi-label training models or specific testing criteria. In bioinformatics, Bosubabu Sambana [11][15] extended the definition of analysis to include multi-label data, but they use a conclusionhierarchy as their baseline algorithm approaches [6]. As they stated, they chose a decision tree because of the sparseness of the data and because they needed to learn accurate rules, not a complete classification. However we desire to use Support Vector Machines for their high accuracy in classification [8].

A related approach to image classification consists of segmenting and classifying image regions (e.g., sky, grass). A seemingly natural approach to multi-label scene classification is to model such scenes using combinations of these labels. For example, if a mountain scene is defined as one containing rocks and sky and a field scene as one containing grass and sky, then an image with grass, rocks, and sky would be considered both a field scene and a mountain scene [6].

However, in some classification tasks, it is likely that some data belongs to multiple classes, causing the actual classes to overlap by definition. In text or music categorization, documents may belong to multiple genres, such as government and health, or rock and blues [4-6] [18][19][20]. Architecture may belong to multiple genres as well. In medical diagnosis, a disease may belong to multiple categories, and genes may have multiple functions, yielding multiple labels [15]. A problem domain receiving renewed attention is semantic scene classification categorizing images into semantic classes such as beaches, sunsets or parties. Semantic scene classification finds application in many areas, including content-based indexing and organization and content-sensitive image enhancement. Many current digital library systems allow a user to specify a query image and search for images "similar" to it, where similarity is often defined only by color or texture properties [7].

This so-called query by example process has often proved to be inadequate. Knowing the category of a scene helps narrow the search space dramatically, reducing the search space, and simultaneously increasing the hit rate and reducing the false alarm rate. Knowledge about the scene category can find also application in context-sensitive image enhancement. While an algorithm might enhance the quality of some classes of pictures [6], it can degrade others. Rather than applying a generic algorithm to all images, we could customize it to the scene type [8]. In the scene classification domain, many images may belong to multiple semantic classes. The Figure -2 shows an image that had been classified by a human as a beach scene. However, it is clearly both a beach scene and an urban scene. It is not a fuzzy member of each (due to ambiguity), but is fully a member of each class (due to multiplicity) issimilar [14].

Automatic data classifiers, where a tested object is assigned to one of pre-defined classes, are broadly used worldwide and they are very useful in many applications. However, some data do not fit into this classification scheme [10]. For instance, when listening to a piece of music from an audio database, one can feel various emotions, and when such data are classified with respect to these emotional states, multi-label classification is much more useful [11].

In this case, each piece of music can be labeled with various emotions associated to this music. Therefore, the authors decided to investigate multi-label classification of data, how one can produce a classifier, and how the classification quality can be tested in multi-label case [7].
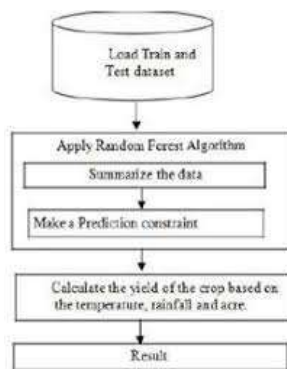
Figure.1: Working Mechanism

The traditional single label classification problems maps each of the input samples to a unique target label from the pool of available target labels. The single label classification problems can be categorized into binary and multi-class classification [8]. When the number of available target labels is two, it is called binary classification. Binary classification is the most fundamental classification problem in which the input sample belongs to either of the two target class labels [5]. Examples of binary classification problems include biometric security, medical diagnosis, etc.

When the number of available target labels is greater than two, the classification problem is called multi-class classification. Biometric identification, character recognition and other similar classification problems are examples of multi-class classification. Binary classification is a special case of multi-class classification in which the number of target labels is two [11]. There are several real world applications in which the target labels are not mutually exclusive and require the need for multi-label classification.

Multi-label classification involves associating each of the input samples with a set of target labels. Therefore, multi-label classification forms the superset of binary and multi-class classification problems [13]. When compared to single label classification, multi-label classification is more difficult and more complex due to the increased generality of the multi-label problems [16]. Several machine learning techniques is available in the literature for multi-label classification problems. The existing multi-label classifiers available in the literature are based on Support Vector Machines (SVM), Decision Trees (DT), and Extreme Learning Machines (ELM) etc.

The machine learning techniques available can be broadly categorized into two categories: Batch learning and online learning. Batch learning techniques involve collection of all the data samples in prior and estimating the system parameters by processing all the data concurrently [14]. Batch learning techniques require all the training data beforehand and cannot learn from streaming data [15].
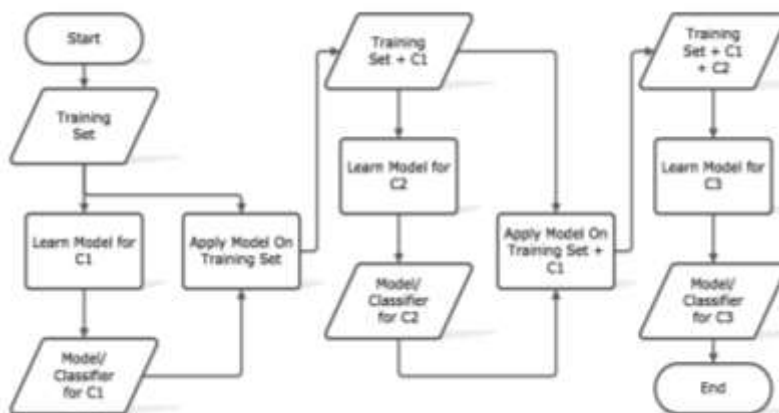

Figure.2: Block diagram of internal Mechanism

Online Learning Methods

There is limited number of techniques available in the literature on multi-label classification for data streams. A simpler approach is to use batch learning classifiers that trains on new batches of data streams by replacing the classifiers of previous batches. This type of learning is called batch incremental learning. The first work on multi-label classifier for data streams is based on ensemble of classifiers which are trained on successive data chunks [11]. The paper by proposes multi-label stream classification by extending the heoffding tree by using batch multi-label classifier in each node.

Spyromitros-Xioufis proposes binary relevance and kNN based multi-label classifier for data streams. Microsoft developed an Active Learning framework for multi-label classification as the result of the increase in demand for the need of multi-label classification in real world multimedia datasets. A Passive-Aggressive method is proposed by Crammer et. al for multi-label classification [12]. The Passive Aggressive and the Bayesian Online Multi-label Classification techniques are application specific and are implemented only for text categorization datasets [22].

Extreme Learning Machines

ELM is a single-hidden layer feed forward neural network based learning technique. The special feature of ELM is that the initial weights and the hidden layer bias can be selected at random [8]. This results in high speed training and small number of tunable parameters thus enabling ELM to have fast learning speed and generalization of performance [11]. The universal approximation capability and generalization ability are the key distinguishing factors of ELM. Several variants of ELM have been developed [13].

Multi-label Classification Methods

Multi-label classification has already been performed in numerous applications in text mining and scene classification domains, where documents or images can be labeled with several labels describing their contents [1-3]. Such a classification requires considering additional issues, including the selection of the training model, as well as set-up of testing and evaluation of results on their classification [14][18].

We can group the existing methods for multi-label classification into two main categories: a) problem transformation methods, and b) algorithm adaptation methods. We call problem transformation methods, those methods that transform the multi-label classification problem either into one or more single-label classification or regression problems, for both of which there exists a huge bibliography of learning algorithm [6]s. We call algorithm adaptation methods, those methods that extend specific learning algorithms in order to handle multi-label data directly [16].

The second improvement of is SVM-specific and concerns the margin ofSVMs in multi-label classification problems [6]. They improve the margin by a) removing very similarnegative training instances which are within a threshold distance from the learnt hyper plane, and b)removing negative training instances of a complete class if it is very similar to the positive class,based on a confusion matrix that is estimated using any fast and moderately accurate classifier on a held out validation set. Note here that the second approach for margin improvement is actually SVMindependent. Therefore, it could also be used as an extension to PT4.
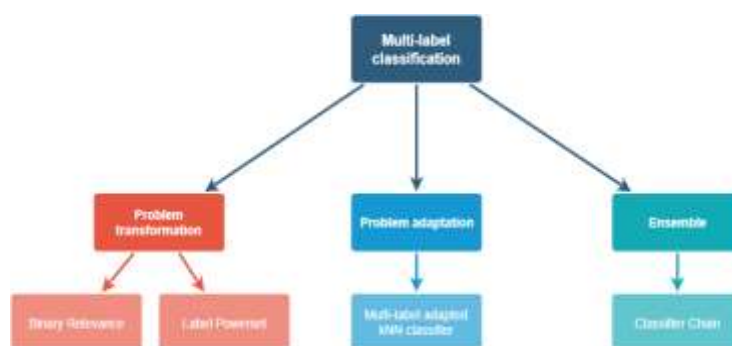

Figure.3: Multi-Label Text Classification

A probabilistic generative model according to which, each label generates different words. Based on this model a multi-label document is produced by a mixture of the word distributions of its labels [6][11]. The parameters of the model are learned by maximum a posteriori estimation from labeled training documents, using Expectation Maximization to calculate which labels were both the mixture weights and the word distributions for each label. Given a new document the label set that is most likely is selected with Bayes rule. This approach for the classification of a new document actually follows the paradigm of PT3, where each different set of labels is considered independently as a new class [17].

Elisseeff and Weston present a ranking algorithm for multi-label classification. Their algorithm follows the philosophy of SVMs: it is a linear model that tries to minimize a cost function while maintaining a large margin [11]. The cost function they use is ranking loss, which is defined as the average fraction of pairs of labels that are ordered incorrectly. However, as stated earlier, the disadvantage of a ranking algorithm is that it does not output a set of labels [19].

Application to Classification

Bosubabu Sambana proposed a new work Multilevel classification problem generalizes traditional two-class or multi-class classification [15] since each instance in the training/testing set is associated with several (usually more than one) classes (labels). The problem is not easy to solve also because the size of the label-set associate with particular unseen instance is generally unknown [18]. Various approaches to tackle this problem were presented in the literature, but – up to our knowledge and there has been only one attempt to apply a neural network for solving this task [2][6]. In this paper a few modifications of the global error function proposed in are presented and experimentally evaluated. Generally, all of them improve performance of the multilevel neural classifier [12].

The improvements in case of the two most elaborate functions, and are noticeable and statistically significant [7]. Overall, including the threshold values into the error function and considering differences between the rank values and the thresholds proved to be a promising direction for improvement of the multilevel classifier performance [9]. Currently, we are focused on performing more tests (especially with other sizes of hidden layer) and on other data sets in order to further verify the efficacy of proposed modifications [10].

3. Proposed Approach Classification Analysis

This paper exploits the inherent high speed nature of the ELM and OS-ELM to develop an online sequential multi-label classifier for real-time streaming data applications. The key novelty of the proposed approach is that, there are no online techniques available thus far in literature to perform real-time multi-label classification [6]. In single label classification problems such as binary and multi-class classification, each input sample corresponds to a single target label.

A task that also belongs to the general family of supervised learning and is very relevant to multi-label Classification is that of ranking. In ranking the task is to order a set of labels L, so that the topmost labels are more related with the new instance [16]. There exist a number of multi-label classifications methods that learn a ranking function from multi-label data. However, a ranking of labels requires Post-processing in order to give a set of labels, which is the proper output of a multi-label classifier [8].

In certain classification problems the labels belong to a hierarchical structure. A web page may be labeled using one or more of those classes, which can belong to different levels of the hierarchy[19]. The top level of the MIPS (Munich Information Centre for Protein Sequences) hierarchy consists of classes such as: Metabolism, Energy, Transcription and Protein Synthesis. Each of these classes is then subdivided into more specific classes, and these are in turn subdivided, and then again subdivided, so the hierarchy is up to 4 levels deep[20]. When the labels in a data set belong to a hierarchical structure then we call the task hierarchical classification. If each example is labeled with more than one node of the hierarchical structure, then the task is called hierarchical multi-label classification. In this paper we focus on flat (non-hierarchical) multi-label classification methods [6].

Therefore the classifier is required to identify the single target label corresponding to the input sample. On the contrary, in multi-label classification[21][8], each of the input samples belongs to a subset of target labels. Therefore, the multi-

label classifier is required to identify both the number of labels and the identity of the labels in order to perform multi-label classification. This results in the increased complexity of the multi-label classification problems. Another key challenge in implementing a generic multi-label classifier is that, not all datasets are equally multi-labeled [15].The degree of multi-liableness varies for every dataset [18]. The increased complexity and the varying degree of multi-liableness are the two major challenges in developing a multi-label classifier [12].
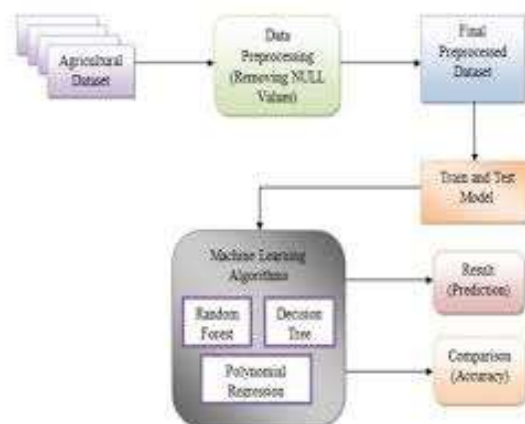

Figure.4: Multi-level preprocessing operations

Pre-processing and post-processing of data are of prime importance in extending the ELM based technique for multilevel classification.

Initialization: The fundamental parameters of the ELM network such as the number of hidden layer neurons and the activation function are initialized [15]. The number of hidden layer neurons is selected for each dataset so as to avoid the over fitting problem. The input weights and the bias value of the network are randomly initialized [8].

Pre-processing: In single label classification, each of the input samples corresponds to only one target class. Therefore, the dimension of the target output label is always fixed at 1. On the contrary, in multi-label classification, each input is associated with an M-tuple output label with each element of the set as 0 or 1 representing the belongingness of the input corresponding to the target labels. Therefore the dimension of the target output label is 'M'. The label set denoting the belongingness for each of the labels is converted from unipolar representation to bipolar representation [13].

ELM Training: The processed input is then supplied sequentially to the online sequential variant of the ELM technique.

Multi-label Identification: The multi-label identification step is the key step in extending the ELM based technique for multi-label classification. As foreshadowed, multi-label classifiers are required to predict both the number of target labels and the identity of the target labels corresponding to 5 each of the input samples [6]. Since the number of labels corresponding to each input is completely unknown and dynamic, a thresholding based technique is used.

The threshold value is selected during the training phase such that it maximizes the separation between the family of labels the input belongs to and the family of labels the input does not belong to. Setting up of the threshold value is of prime importance as it directly [15] affects the performance of the classifier. The raw output values Y obtained from the previous step is then compared to the threshold value [8]. The number of raw output values that are greater than the threshold determines the number of target labels corresponding the input sample and the index of the corresponding values determines the identity of the target labels. The overview of the proposed algorithm is summarized [6][21].

4. Conclusion

The proposed multi level classification technique is a real-time online multi-label classifier for streaming data applications. The performance of the proposed method is experimented on five datasets of different domains with a wide range of LC and LD. From the analysis of evident that the modified method is consistent and outperforms the existing state-of-the-art techniques in terms of speed and remains one of the top methods in terms of performance.

This paper introduces a step-by-step and end-to-end methodology of how to design and train a scalable multilabel classifier. This methodology also shows how to expand the capability of a single labeled deep learning model into a multilabel classifier by leveraging on transfer learning with low cost of labeling and achieving impressive accuracy. Using our methodology, a single model can achieve multiple classifications on an image without an increased cost. This method is quite useful for exploiting a single image for various classifications.

References

[1] " Multi-Label Classification: An Overview ", Grigorios Tsoumakas, Ioannis Katakis

[2] R. Grodzicki, J. Mańdziuk, and L. Wang , " Improved Multilabel Classification with Neural Networks "

[3] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014, pp. 675–678.

[4] Boutell, M., Shen, X., Luo, J., Brown, C.: Multi-label Semantic Scene Classification. Technical Report, Dept. of Computer Science, U. Rochester (2003)

[5] L. Zhang, Y. Jin, X. Yang, X. Li, X. Duan, Y. Sun, and H. Liu, "Convolutional neural network-based multi-label classification of pcb defects," The Journal of Engineering, vol. 2018, no. 16, pp. 1612–1616, 2018.

[6] Bosubabu Sambana," Block chain approach to cyber security vulnerabilities attacks and potential Countermeasures", International Journal of Security and Its Applications, ISSN: 2207-9629 IJSIA, Vol. 14, No. 1 (2020), pp.1-14, http://dx.doi.org/10.33832/ijsia.2020.14.1.01

[7] Bosubabu Sambana, "Internet of Things: Applications and Future Trends", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 3, March 2017, ISSN (Online): 2320-9801,

[8] Bosubabu Sambana, "Identifying Datasets by Pattern Recognition Techniques using Neural Networks" Asian Journal of Computer Science and Technology ISSN: 2249-0701 Vol. 6 No. 1, 2017, pp.34-39.

[9] Bosubabu Sambana, "Signal Processing and Display of Linear Frequency Modulated Continuous Wave Radar on a Chip" Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, Special Issue, NCETC-17, April 2017.

[10] Bosubabu Sambana , "An Artificial Intelligence Approach to Social Networks Agent Task Scheduling Analysis in Map Reduce for Sentiment Opinion Analysis" , 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (IEEE-iSSSC 2020)

[11] Bosubabu Sambana et al. "Implementation of Securely Sharing Critical Information in Cloud Environment Using Mediated Certificate less Encryption System", International Journal & Magazine of Engineering, Technology, Management and Research, ISSN No: 2348-4845, Volume No: 2 (2015), Issue No: 11(November).

[12] Bosubabu Sambana, "An artificial intelligence approach to fake currency detection system", International Journal of Signal Processing, Image Processing and Pattern Recognition, ISSN: 2207-970X, IJSIP, Vol. 12, No. 3 (2019), pp.1-16 . http//dx.doi.org/10.33832/ijsip.2019.12.3.01.

[13] Shoulder Surfing Experiments: A Systematic Literature Review. Available from: https://www.researchgate.net/publication/343978396_Shoulder_Surfing_Experiments_A_Systematic_Literature_Review.

[14] Bosubabu Sambana, "Integrative Spectrum Sensing in Cognitive Radio using Wireless Network", Springer Nature Conference - International conference on Remote Sensing for Disaster Management (ICRSDM-2017) during 11-13 October 2017 at Andhra University-Visakhapatnam. First Online: 24 June 2018, Publisher Name: Springer, Cham. Print ISBN: 978-3-319-77275-2, Online ISBN: 978-3-319-77276-9

[15] Bosubabu Sambana, " Air Pollution Monitoring and Prediction System using the Internet of Things " , International Journal of Research and Development in Engineering Sciences , ISSN:2582-4201 , Volume: 2, Issue: 3 , Year: 2020

[16] Bosubabu Sambana, " Smart Agricultural activities Monitoring and Control system using an Internet of Things" , Journal of Xidian University, ISSN No:1001-2400 , Volume-14 Issue-7, July 2020,Page No: 1645 – 1654

[17] Bosubabu Sambana, " Artificial Intelligent Structure User-Friendly Multi-flexible Bed Cum Wheelchair Using Internet of Things ", Springesr Nature - Lecture Notes in Networks and Systems book series (LNNS, volume 105), Springer Nature Singapore Pte Ltd. 2020

[18] Bosubabu Sambana, " Vehicle Tracking and Accident Detection System using Internet of Things ", International Conference on Smart Technologies in Data Science and Communication (Smart DSC-2018) , Advanced Science and Technology Letters Vol.147 (SMART DSC-2017), pp.463-467.

[19] Bosubabu Sambana, "Implementation of Securely Sharing Critical Information in Cloud Environment Using Mediated Certificate less Encryption System", International Journal & Magazine of Engineering, Technology, Management and Research, ISSN No: 2348-4845, Volume No: 2 (2015), Issue No: 11(November).

[20] Bosubabu Sambana, "Identifying Datasets by Pattern Recognition Techniques using Neural Networks" , Asian Journal of Computer Science and Technology ISSN: 2249-0701 Vol. 6 No. 1, 2017, pp.34-39.

[21] Bosubabu Sambana, " Privacy-preserving Authentication Protocol using Trusted Third Party Secure Data Sharing in Cloud Computing ", International Journal & Magazine of Engineering, Technology, Management and Research , ISSN No: 2348-4845, Volume No: 3 (2016), Issue No: 3 (March)

[22] Bosubabu Sambana, "World Wide Web: Web Page Segmentation", International Journal & Magazine of Engineering, Technology, Management and Research, ISSN No: 2348-4845, Volume No: 3 (2016), Issue No: 3 (March).

[23] Rajasekar Venkatesan1*, Shiqian Wu , Mahardhika Pratama, " A Novel Online Real-time Classifier for Multi-label Data Streams".