

Applications of Multi-Label Classification

Ayesha Mariyam, SK Althaf Hussain Basha, S Vishwanadha Raju

Abstract: *The absence of labels and the bad quality of data is a prevailing challenge in numerous data mining and machine learning problems. The performance of a model is limited by available data samples with few labels for training. These problems are ultra-critical in multi-label classification, which usually needs clean data. Multi-label classification is a challenging research problem that emerges in several applications such as multi-object recognition, text categorization, music categorization and image classification. This paper presents a literature review on multi-label classification, various evaluation metrics used for analyzing performance and research challenges.*

Keywords: *Multi-label classification, Deep Learning, Machine Learning, Feature Extraction.*

I. INTRODUCTION

Multi-label classification is a primary but challenging problem in machine learning. In contradiction to single-label classification, multi-label classification not only associate labels to corresponding instances but also makes use of the dependencies present between labels[1]. For example, in multi-label image categorization, an image consisting of sky and beach normally appear together but not airplane and cat occur often. Supervised learning is used to train classifiers, which takes labeled data as training and learns, its performance depends on the quality of data provided for training. There are two main challenges related to data quality, which includes absence of labeled data and noisy data.

[10] There exist two types of methods for multi-label classification; they are 1) Problem Transformation Methods (PTM) and 2) Algorithm Adaptation Methods (AAM). On one hand problem transformation methods work by fixing data to algorithms, i.e., coercing multi-label data to integrate other well-developed learning frameworks. On the other hand, algorithm adaptation methods work by fitting algorithms to data, i.e., adapting popular standard supervised learning algorithms to deal with multi-label data. Problem Transformation can be carried out in three different ways as: 1). Binary Relevance (BR) 2). Classifier Chains(CC) 3). Label Power set(LP).

The baseline approach called the binary relevance(BR) method, amounts to independently training one binary classifier for each label.

The label power set (LP) transformation creates one binary classifier for every label combination attested in the training set.

Revised Manuscript Received on March 18, 2020.

* Correspondence Author

Ayesha Mariyam*, CSE, Jawaharlal Nehru Technological University Hyderabad, Hyderabad, India. ayesha.mariyam84@gmail.com

Sk Althaf Hussain Basha, CSE, A1 Global Institute of Engineering and Technology, Markapur, India. althafbashacse@gmail.com

S Vishwanadha Raju, CSE, Jawaharlal Nehru Technological University Hyderabad, Jagtial, India. svraju.jntu@gmail.com

The Classifier Chain model (CC) involves $|L|$ binary classifiers as in BR. Classifiers are connected along a chain so that each classifier deals with the binary relevance problem related with label $l_j \in L$.

Some classification algorithms or models have been adapted to the multi-label task, without requiring problem transformations such as boosting, k-nearest neighbors, decision trees, neural networks etc.

[3] A multi-label classification task for unmanned aerial vehicle (UAV) images obtained from urban areas uses auto encoder(AE) which results into reduced dimensionality and then fed into multi layer perceptron (MLP) for getting the list of objects present in the given image tile. This method of processing results good accuracies and lesser processing time.

In natural language processing, generally bag-of-words paradigm has been used to depict documents, but it has experienced problems such as sparsity and ambiguity[4],[7] Micro blogging websites have a lot of textual data and classification of such data is crucial in various applications such as user profiling, content tagging, topic analysis and information filtering. To overcome the problems associated with BOW, with the growth of word embedding method and deep learning in recent years, variant CNN models were used for sentence classification and modeling. Convolutional neural networks were also been utilized in the areas such as image classification and object recognition, where CNNs outperforms the state-of-art algorithms [5].

[8] In multi-label classification, generative adversarial network (GAN) is used to model the label allocation.

II. LITERATURE REVIEW

In 2017, Guanglei Zhan et al [1] proposed a new multi-label classification approach based on CNNs by making use of invisible semantic enclosing different labels. The invisible semantic between multi-label is imported to increase the performance. The dataset used for experimentation is CIFAR-100. To some amount

the proposed model got better classification performance in contrast to CNN model for single label only.

Wang Zhan et al [2] proposed a multi-label learning model named LIFTACE which creates label-specific features by considering label correlations via clustering ensemble techniques. The experiments done on different datasets have proved the potential of the proposed model when compared with traditional algorithms.

Abdallah Zeggada et al [3] have proposed an effective multi label classification approach for UAV images above urban areas. This approach implements a two-stage signature representation for each tile of the query image, by employing a proper feature extraction method and feature learning



Step that improves the features discrimination capability using an auto encoder. After that a multilayer perceptron network is trained for the multi-label classification task. The proposed approach has shown better results specially in terms of processing time, which is beneficial for real-time urban monitoring applications.

In 2018, Yun Chen et al [4] explored, proposed and improved different neural network architectures to take contextual data. Different neural networks are applied to text present in Chinese public text dataset and got good results. Initially, the architecture called inception is designed to extract features of n-gram and then a simple modification is done on top of RNN models to extract more information.

Xinsheng Wang et al [5], proposed an improved CNN via hierarchical dirichlet process (HDP) model to manage the multi-label classification problem in NLP. At first an HDP model is applied to remove the words which are semantically less important. Then word embedding methods are used to transform words to vectors. Finally, CNN is trained based on word vectors. Experimental results have shown that this method is better than traditional multi-label classification methods and Text CNN in terms of performance.

Rajni Jindal et al [6], proposed a new method for automated categorization of text documents. The proposed method is constructed on lexical and semantics concepts. It is a new idea in the area of categorization of research articles. It can be applied to both single-label and multi-label text documents. The proposed method is experimented and tested with a dataset of IEEE research articles. It has given a good performance with an accuracy of 75%.

In 2019, MD. Aslam Parwez et al [7], presented various neural network models for multi-label classification of micro blogging data. These models are based on convolutional neural network (CNN) architectures, which make use of pre-trained word embeddings from generic and domain-specific textual data sources. To predict class labels, the word embeddings are handled separately and in different combinations through different channels of CNN. A comparative analysis of the proposed CNN models with state-of-art machine learning models and one of the existing CNN architectures is done. The proposed models are assessed over a real Twitter dataset, and the results have shown their efficiency to classify micro blogging texts with improved accuracy.

Che-Ping Tsai et al [8], proposed a creative structure to train a multi-label classifier. A Generative Adversarial Network (GAN) is used to imitate the label dissemination for multi-label classification. This structure is built upon a conditional GAN (cGAN). The classifier acts as a conditional generator, which has some instance as input and it outputs a set of labels like a multi-label classifier. This framework also has a discriminator which is trained to model label dependency, that is it takes an object and a set of labels as input, and outputs a score. The discriminator learns to discriminate the real and generated label sets. The generator then learns to fool the discriminator by generating label sets with what seems to be the correct dependencies, given an input instance by the discriminator. The classifier and discriminator are learned iteratively as in a typical GAN. Large scale experiments and comparisons on two large-scale image classification datasets have shown that with this

discriminator, F1 scores are improved across different classifier models.

Marciele M. Bittencourt et al [9], proposed the ML-MDL Text, a new text categorization method which has the ability to classify multi-label documents without transforming the problem into a binary or multiclass task. A model can be trained incrementally, which allows its application in dynamic and large-scale text classification problems. Experiments are done by imitating the online learning process in real world applications, examining situations in which incremental training is done to correct wrong predictions of the classifier. ML-MDL Text requires lower computational cost than PA, SGD, and MLP mainly in tasks with large number of classes.

The table –I below shows the summary of literature review carried out.

III. EVALUATION METRICS

Multi label problems deals with more than one labels consequently classification of a given sample may be partly correct or partly incorrect. Mainly there are two types of evaluation measures for multi label classification problems: Example Based Measures and Label Based Measures.

A. Example Based Measures:

Let (X, Y) be a multi-label example, $Y \subseteq L$. Let M be a multi-label classifier. Let $Z=M(X)$ is a set of labels predicted by M for (X, Y) . Let N be the total number of instances.

▪ Accuracy:

Accuracy for each instance is defined as the ratio of the predicted correct labels to the total number (predicted and actual) of labels for that instance. Overall accuracy is the averages across all instances. Accuracy

$$(M, N) = \frac{1}{|N|} \sum_{i=1}^{|N|} \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|}$$

$$M_{\text{micro}} = \frac{M(\sum_{\lambda=1}^{|L|} tp_{\lambda} + \sum_{\lambda=1}^{|L|} fp_{\lambda} + \sum_{\lambda=1}^{|L|} tn_{\lambda} + \sum_{\lambda=1}^{|L|} fn_{\lambda})}{\sum_{\lambda=1}^{|L|} (tp_{\lambda} + fp_{\lambda} + tn_{\lambda} + fn_{\lambda})}$$

▪ Precision:

Precision is the percentage of predicted labels that were correct.

$$\text{Precision}(M, N) = \frac{1}{|N|} \sum_{i=1}^{|N|} \frac{|Y_i \cap Z_i|}{|Z_i|}$$

▪ Recall:

Recall is the percentage of correct labels that were predicted.

$$\text{Recall}(M, N) = \frac{1}{|N|} \sum_{i=1}^{|N|} \frac{|Y_i \cap Z_i|}{|Y_i|}$$

▪ Hamming loss:

Hamming Loss reports how many times on average, the relevance of an example to a class label is incorrectly predicted. Therefore, hamming loss takes into account the prediction error (an incorrect label is predicted) and the missing error (a relevant label not predicted), normalized over total number of classes and total number of examples.

$$\frac{1}{|N|} \sum_{i=1}^{|N|} \frac{|Y_i \Delta Z_i|}{|L|}$$

Hamming Loss (M, N) = $\frac{1}{|N|} \sum_{i=1}^{|N|} \frac{|Y_i \Delta Z_i|}{|L|}$
Where, Δ stands for the symmetric difference of two sets.

B. Label Based Measures:

Calculate a binary evaluation measure separately for each label. Two averaging operations are used across all labels: micro and macro average. Binary evaluation measure is calculated using parameters of confusion matrix (true positive, true negative, false positive and false negative).

$$M_{\text{macro}} = \frac{1}{|L|} \sum_{\lambda=1}^{|L|} M(tp_{\lambda}, fp_{\lambda}, tn_{\lambda}, fn_{\lambda})$$

IV. APPLICATIONS

The following are the list of applications of multi-label classification in various industries as shown below in fig.1.

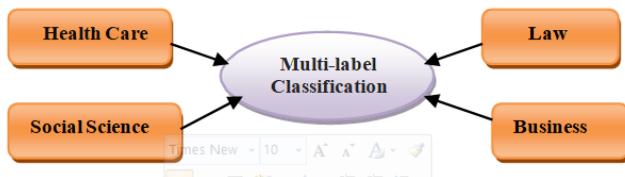


Fig.1. Multi-label Classification Applications

▪ **HealthCare:**

The medical images associated with the health care system is unlabeled. Hence it is essential to label the Medical diagnoses to specific class values, requires multi-label classification.

▪ **SocialScience:**

Sentiment analysis is done for observing the human behavior in societal issues, which mainly concentrates on mining the content present in tweets, short message service (SMS), clinical notes etc.

▪ **Law:**

It is necessary for legal practitioners to find pertinent laws and regulations before regulating the fines, penalties, charges and other sentences for the accused or suspects. Nonetheless, compound regulations and rules are clearly an obstruction for normal public. As a result, law articles labeling is needful and workable in legal intelligence learning.

▪ **Business:**

The industry experts are utilizing the social media as a stage for marketing. Opinion mining from social media networks is used in a large-scale to improve the business.

Table- I: Summary of Literature Review

Authors	Datasets	Methodology	Results	Future Scope
Guanglei Zhang et al [1]	CIFAR-100	This paper introduces a new CNN and linear sparsity auto-encoder.	It attains good performance in comparison with CNN model using single label only. The accuracy of the classification decreases with the increase of the super classes because of less features input to the coarse labels.	The higher performance can be achieved if the independent CNNs is used for the coarse labels or with the increase in the super classes, the complexity of coarse labels increases.
Wang Zhan et al [2]	Multiple datasets such as CAL500, language log etc are used for comparison	LIFTACE named multi-label learning method is proposed, which create label-specific features through label correlations.	Experimental studies prove the efficacy of the proposed method in comparison with the conventional algorithms.	Other strategies can be explored to learn from multi-label data by making use of label correlations in creating label-specific features.
Abdallah Zeggada et al [3]	two real data sets of UAV images picked up from two different locations	A new multi label classification method is proposed for UAV image.	One can refer from the results that the proposed method is favorable for UAV multi label applications.	various strategies for tile representation and the classification steps can be explored.
Yun Chen et al [4]	Zhihu corpora	Various neural network architectures are explored and presented to extract information and important components from text.	The proposed model initiation performs better than its forebear TextCNN in all angles. Extensive analysis and experiments on speed, memory-consumption, and accuracy conducted to know the power and quality of deep learning models for NLP tasks.	The model proposed can be used as a supporting network and be utilized for other NLP problems to prove its generalization.
Xinsheng Wang et al [5]	Ohsumed and Slashdot	An enhanced CNN via hierarchical dirichlet process (HDP) architecture is proposed to deal with the problem	TextCNN+HDP is better than TextCNN in converge time and performance.	More layers can be added to the TextCNN model.
Rajni Jindal et al [6]	Computer science domain 150 articles taken from IEEE Xplore digital library	A approach for multi label categorization of text documents based on semantic and lexical concepts is proposed.	IEEE Xplore digital library results are compared with the proposed method which has given accuracy of 75%.	The proposed method can be applied on other digital libraries and other domains of research articles.



Applications of Multi-Label Classification

MD. Aslam Parwez et al [7]	Twitter	Different Neural network models for multi-label classification of micro blogging data are proposed.	The efficiency of CNN model outperformed from conventional algorithms in accuracy.	Online social media tweet streams or text streams can be labeled and aggregated with the help of CNN models which is worthwhile for the growth of online social media surveillance systems including prediction and disease surveillance.
Che-Ping Tsai et al [8]	NUS-WIDE with 81 labels and the MS-COCO benchmark with 80 labels	A futuristic framework on generative adversarial networks (GANs) is proposed for label dependency	F1 scores are enhanced over different classifiers with the discriminator, when Extensive experiments and comparisons are done on two large-scale image classification datasets	Apply the same approach for other domain datasets.
Marciele M. Bittencourt et al [9]	Various datasets used such as Reuters-ORGs, Reuters-places, Bibtext etc	A proposed ML-MDLText is a new method which classifies multi label documents without transforming the problem into a binary or multi class task and trained incrementally.	The results show that ML-MDLText is one of the best methods in online learning context.	Improve the prediction capability of ML-MDLText by investigating label dependency and by regulating the size of the label set.

V. CONCLUSION

The multi-label classification task is the most critical problem. In this paper we explored and discussed various recent advancements in the field of multi-label text classification which presents future scope as well and the evaluation metrics. The applications and research areas of text classification are also mentioned. It is also reasonable to implement the future scope presented in this paper.

REFERENCES

- Guanglei Zhang, Lei Chen, Yongsheng Ding, "A Multi-Label Classification Model Using Convolutional Neural Networks," 29th Chinese Control And Decision Conference (CCDC), *IEEE*, 2017, pp.2151-2156.
- Wang Zhan, Min-Ling Zhang, "Multi-Label Learning with Label-Specific Features via Clustering Ensemble," International Conference on Data Science and Advanced Analytics, *IEEE*, 2017, pp.129-136.
- Abdallah Zeggada, Farid Melgani, "Multi-labeling UAV Images with Auto encoder Networks," Geoscience and Remote Sensing Letters, *IEEE*, 2017, pp.1-5.
- Yun Chen, Bo Xiao, Zhiqing Lin, Cheng Dai, Zuochao Li and Liping Yan, "Multi-label Text Classification with Deep Neural Networks," Proceedings of IC-NIDC, 2018, pp.409-413.
- Xinsheng Wang, Lijun Sun, Zhihua Wei, "An Improved Convolutional Neural Network Algorithm for Multi-label Classification," ICALIP, *IEEE*, 2018, pp.113-117.
- Rajni Jindal, Shweta, "A Novel Method for Efficient Multi-Label Text Categorization of research articles," IEEE International Conference on Computing, Power and Communication Technologies, 2018, pp.333-336.
- MD. Aslam Parwez, Muhammad Abulaish, Jaheeruddin, "Multi-Label Classification of Micro blogging Texts Using Convolution Neural Network," *IEEE ACCESS*, 2019, pp.68678-68691.
- Che-Ping Tsai, Hung-Yi Lee, "Adversarial Learning of Label Dependency: A Novel Framework for Multi-Class Classification," *IEEE*, 2019, pp.3847-3851.
- Marciele M. Bittencourt, Renato M. Silva, Tiago A. Almeida, "ML-MDL Text: A Multi-label Text Categorization Technique with Incremental Learning," 8th Brazilian Conference on Intelligent Systems (BRACIS), *IEEE*, 2019, pp.580-585.
- A. P. L. F. de Carvalho and A. Freitas, "A Tutorial on Multi-label Classification Techniques," in Foundations of Computational Intelligence Volume 5, Springer Berlin Heidelberg, 2009, pp.177-195.

AUTHORS PROFILE



Ayesha Mariyam received the M.Tech degree from Jawaharlal Nehru Technological University Hyderabad in 2011. She is currently pursuing PhD from Jawaharlal Nehru Technological University, Hyderabad, India. Her research interest includes machine learning, deep learning, data mining.



SK Althaf Hussain Bashare received PhD (Computer Science and Engineering) from Acharya Nagarjuna University in the year 2013. He obtained his M.Tech (Computer Science and Engineering) from JNTU, Hyderabad in the year 2008, and obtained his M.Sc (CSC) from Acharya Nagarjuna University in the year 1999. His Research areas of Interest include Machine Learning, Information Retrieval, Data Mining and Web Mining. He is active member in FSF AP Chapter, IEI Hyderabad, IEEE Hyderabad and CSI.



S Vishwanadha Raju received M.Tech from JNTUH and Ph.D. from Acharya Nagarjuna University, Guntur. He is certified CMI in Education Leadership and Program Management from Dudley college London which is sponsored by MHRD-AICTE and UKIERI. His research interests include Programming methodology; Algorithms, Information Retrieval, Biometrics, Databases, Data Mining, Research Methodology and Quality in Higher Education are globally recognized.