A Survey and Analysis of Multi-Label Learning Techniques for Data Streams

S.K.Komagal Yallini
Research Scholar (Bharathiar University)
School of Computing
Sri Rama Krishna College of Arts and Science
Nava India Coimbatore, Tamil Nadu, India

Dr. B.Mukunthan
Associate Professor,
Department of Computer Science
Sri Rama Krishna College of Arts and Science
Coimbatore, Tamil Nadu, India

Abstract:- Multi-Label Learning (MLL) solves the challenge of characterizing every sample via a particular feature which relates to the group of labels at once. That is, a sample has manifold views where every view is symbolized through a Class Label (CL). In the past decades, significant number of researches has been prepared towards this promising machine learning concept. Such researches on MLL have been motivated on a pre-determined group of CLs. In most of the appliances, the configuration is dynamic and novel views might appear in a Data Stream (DS). In this scenario, a MLL technique should able to identify and categorize the features with evolving fresh labels for maintaining a better predictive performance. For this purpose, several MLL techniques were introduced in the earlier decades. This article aims to present a survey on this field with consequence on conventional MLL techniques. Initially, various MLL techniques proposed by many researchers are studied. Then, a comparative analysis is carried out in terms of merits and demerits of those techniques to conclude the survey and recommend the future enhancements on MLL techniques.

Keywords:- Multi-label learning, Label correlations, Multiple instances, Machine learning, Multi-label problem transformation.

I. INTRODUCTION

Conventional supervised learning is the most well-known machine learning concepts where every item is denoted by a certain feature vector related to the specific CL. Though this learning is popular and successful, in several uses, specific feature might have many CLs. For exemplar, a scene photo is normally interpreted with many CLs [1]; a file might contain various themes [2] and a music segment might fit to various fields [3]. To handle these types of data, MLL has been emerged which is also the learning concept and concerned more interest in modern decades [4].

In MLL, every item is denoted by a particular feature when related to the group of CLs rather than the specific CL in the traditional supervised learning. The process is identifying the appropriate CL groups for unknown features. In the previous years, MLL has increasingly involved considerable interests from machine learning and broadly used in various issues such as automated interpretation for audio-video data [5] including image, bioinformatics, web mining, rule mining, information retrieval and tag recommendation, etc. Earlier studies of MLL mostly focused on the issue of ML manuscript labelling and considered the predetermined set of CLs. But, in several real-time applications, a dynamic scenario is taken into account in which fresh CLs might occur with well-known CLs in a predicted feature of a DS.

In a dynamic scenario, a learning technique has the ability to reprocess and adjust a pre-trained framework to the varying atmosphere [6]. In the ML configuration, the technique should have the ability to reform a pre-learned framework into the fresh features are found and novel classifiers are constructed for each fresh CLs. In the dynamic MLL configuration, there are no ground truths for CLs in the DS at each time excluding the actual training set. So, the primary challenges are identifying and modelling fresh CLs. Particularly, the most complex is identifying the features with any fresh CL. Because, there is no past information of the fresh CL and it often co-appears with few well-known CLs, it is extremely complex to split features with fresh CLs from those with well-known CLs only. Due to the inappropriate identification, the error will increase as increasing fresh CLs in a DS. Therefore, designing the effective frameworks for enhancing the identification performance in a DS is also a difficult process. To tackle this problem, various MLL techniques with promising solutions have been accounted to identify the correlations between the labeled and unlabeled features. This paper discusses different MLL techniques used for increasing the accuracy while using multiple CLs for data features. It also focuses on the merits and limitations of these techniques to suggest further improvement on MLL.

II. A REVIEW ON VARIOUS TECHNIQUES FOR MULTI LABLE LAEARNING

Tsoumakas et al. [7] proposed a scheme, namely RAndom k labELsets (RAkEL) where k was the subsets size. In this scheme, two different strategies were considered that leads to disjoint and overlapping labelsets. The main goal of this scheme was splitting a huge amount of CLs into the amount of small-sized labelsets in a random manner. Then, the ML classifier was constructed via the Label Powerset (LP) scheme for training each labelset. Moreover, decisions of all LP classifiers were accumulated and fused for classifying the CLs of an unknown features.

Kong et al. [8] recommended a TRAsductive Multi-label (TRAM) classification for assigning many CLs to every feature through labeled and unlabeled data. In TRAM, the CLs of the unlabeled features were estimated efficiently using the labeled and unlabeled data. Initially, the TRAM classification was formulated as an optimization dilemma of identifying the CL notion configurations. After that, the closed-form result was derived and efficient technique was proposed for assigning the CLs to the unlabeled features.

Zhang et al. [9] proposed the MLL with Label specific FeaTures (LIFT). At first, specified features to every CL was extracted using clustering scheme on its desirable and undesirable features. After, training and testing were performed via querying the clustering outcomes. Moreover, a group of classifiers were encouraged with each deriving from the basic features of known CL.

Liu et al. [10] proposed the maximum CL distance Back-Propagation (BP) scheme for classifying many CLs. This scheme was devised via fine-tuning the overall mistake factor of the classical BP via taken into consider the penalty term recognized through increasing the space between the desirable and undesirable CLs. Also, the weights were controlled and the network's overview efficiency was improved.

Pham et al. [11] proposed an appropriate yet efficient implementation of the maximum likelihood method in terms of handling the determination of the system factors and repeatedly learning a feature-level classifier one by one on consecutive fashion for all CLs including the fresh CL. The major contributions were developing a system that accounts the occurrence of fresh CL features and proposing an accurate inference method.

Xu et al. [12] suggested a Three-way Incremental Learning Algorithm (TILA) for radar emitter detection which is flexible for increase in emitter characteristics, types and samples. This algorithm was dealt with fundamental cases of incremental learning such as sample increment, type increment and feature increment. In TILA, the data description variables were incrementally updated based on which discriminating features were chosen and the emitter types were detected.

Mu et al. [13] proposed an alternate method via unsupervised learning for solving the categorization under Streaming Emerging New Classes (SENC). In this method, a fully-random trees, namely SENCTrees were employed that can able to operate in the unsupervised and supervised learning independently. Also, the isolation-based anomaly detection scheme was used for constructing the classifier and detector. The anomalies of recognized CLs from features of fresh CLs were explicitly differentiated by using the SENCForest which is composed of SENCTrees. This model was modernized with no primary training set since there was no necessary of training fresh models.

Huang et al. [14] proposed an efficient Bayesian framework for categorizing many CLs using Local desirable and undesirable Pairwise Label Correlations (LPLC). During the training phase, the desirable and undesirable CL correlations of every ground truth CL were obtained for each training sample. During the test phase, K-Nearest Neighbor (KNN) and their equal desirable and undesirable PLC for every test sample was initially detected. After that, the prediction was achieved by increasing the posterior likelihood computed on the CL allocation and the LPLCs represented in the KNN.

Zhu et al. [15] suggested a novel MLL with GLObal and loCAL CL correlation (GLOBAL) that deals with the complete- and missed-label scenarios. The major functions of GLOBAL were: i). Used the low-rank structure of the CL matrix for obtaining a denser and abstract latent CL representation including the normal result to missed CL retrieval, ii). Achieved the global and local CL correlations and thus the CL classifier might use data from each CL, iii). Trained the CL correlations with no usual and complex manual description of the correlation matrix, iv). Integrated all these functions into a single joint learning challenge and used effective alternate reduction scheme.

Tan et al. [16] developed the Semi-supervised Multi-label categorization via Incomplete Label Information (SMILE). Initially, label correlation was estimated from incompletely labeled features and their missed CLs were restored. Afterwards, the labeled and unlabeled features were used for constructing the region graph. Subsequently, the recognized CLs and restored labeled features including the unlabeled features were obtained for training a graph-based semi-supervised linear classifier. The missed CLs of training features were restored according to the region graph. Also, the CLs of fully unlabeled fresh features were directly predicted.

Wu et al. [17] designed a new Cost-sensitive MLL with Positive and Negative Label (CPNL) pairwise correlations for resolving the MLL challenge. Initially, the cost-sensitive loss matrix was computed and integrated with the loss function for resolving the class-imbalance issue. After that, 2 sparse symmetric similarity matrices were computed related to the PNL correlations, respectively. Also, 2 regularizers were added for explicitly exploring the PNL pairwise correlations in the multiple assumptions of the CL distance. Additionally, the kernel addition of the linear framework was proposed for exploring the complex nonlinear input-output correlations.

Zhang et al. [18] suggested a new technique for joint learning of CL-specific features and CL correlations. The objective of this technique was designing an optimization method for learning the weight distribution strategy and the correlations among CLs were considered by constructing the additional features simultaneously. The CL-specific features were learned by the sparsity regularized optimization in ML setting. The MLL challenge was modeled as an optimization method where the feature's weights and CL correlations-based features were

represented as 2 groups of fresh features. These fresh features were updated by using an iterative optimization method. Moreover, CL correlations were denoted via extra features created in the optimization and a KNN was applied for obtaining the CL correlations-based features of test sample.

Weng et al. [19] recommended a LF-LPLC for MLL that merges the CL-specific features and LPLC simultaneously. Initially, the actual feature space was converted to the low-dimensional CL-specific feature space. After that, the local correlation between every couple of CLs was exploited using KNN. Then, the CL-specific features of every CL were extended through merging the relevant information from another CL-specific features. At last, a binary classifier was constructed to test the unlabeled features for every CL according to its CL-specific features.

Wang et al. [20] designed a novel scheme to learn the ML classifiers using the confidential data. In this scheme, the similarity constraints were used for obtaining the correlation between accessible data and confidential data. Also, the sorting criteria were used for obtaining the dependences among many CLs. By merging both the criteria into the classifier's training, the valid data and the dependences among many CLs were achieved for building an improved classifier or group of classifiers during training. During testing, only accessible data was considered.

Ma & Chow [21] proposed a new CL retrieval scheme in the semi-supervised configuration. This scheme has the ability to execute the CL matrix prediction in the labeled and unlabeled space at the same time. Additionally, the semantic correlation were exploited for increasing the sturdiness to semantic breaks and variable CL correlations. During the fitness factor formulation, l_1-norm and nonnegative limits were used for obtaining the secret interactive graphs in semantic-level and revealing the interpretation. Also, an iterative method was applied for ensuring all variables were reliable.

Zhu et al. [22] recommended the MLL with Emerging New Labels (MuENL) for detecting and classifying the features with ENLs. In MuENL, three functionalities were performed such as classifying the features on presently recognized CLs, identifying the occurrence of a fresh CL and building a novel classifier for every fresh CL which operates cooperatively with the classifier for recognized CLs. Also, this method was extended to MuENLHD for handling sparse high-dimensional DSs by dimensionality reduction via streaming kernel Principal Component Analysis (PCA).

Table 1 gives the merits and demerits of the studied MLL techniques.

Ref. No.	Techniques	Merits	Demerits	Dataset Used	performance
[7]	RA*k*EL	Computationally efficiency.	It cannot be used in practice due to complex training issue.	Scene, yeast, TMC2007, medical, Enron, mediamill, reuters and bibtex	Indicative micro F1 and macro F1 measures
[8]	TRAM	Efficient performance using both labeled and unlabeled data.	It cannot generalize to fresh samples.	Annotation, yeast, yahoo, RCV1-v2 dataset and scene	Micro F1, Hamming Loss (HL), Ranking Loss (RL) and average precision
[9]	LIFT	High efficiency.	The CL correlations were not considered during creation of CL-specific features.	CAL500, language log, Enron, image, scene, yeast, Slashdot, corel5k, RCV1 dataset, bibtex, corel16k, eurlex, tmc2007 and mediamill	HL, one-error, coverage, RL, mean precision and macro-averaging Area Under Curve (AUC)
[10]	Maximum CL distance BP algorithm	More effective.	Overall computational cost was high while increasing the number of training features.	Yeast, human and plant	One-error, RL, average precision, HL, F1 and AUC
[11]	Maximum likelihood method	It can find many fresh CLs rather than one fresh CL only.	Computational cost was increased linearly while increasing the amount of bag CLs.	MSCV2, letter carroll, and letter frost datasets, MNIST handwritten dataset	HL and AUC

[12]	TILA	Better efficiency and insensitivity to data input sequence.	Total time complexity was high.	Airborne radar emitter dataset and ground radar emitter dataset.	Average true positive rate, runtime
[13]	SENCForest unsupervised and supervised learning	Performs effectively in long-streams with an adequate memory settings.	It cannot distinguish a number of fresh classes.	Synthetic, KDDCup 99, forest cover, MHAR and MNIST datasets	EN_Accuracy and F-measure
[14]	Effective Bayesian model using LPLC	Better performance.	Computational complexity was high.	Flags, cal500, emotions, yeast, corel5k, RCV1, corel16k, delicious, bookmark and imdb	HL, accuracy, exact-match, F1, macro F1 and micro F1
[15]	GLOBAL	Better accuracy.	It cannot handle the case that CL correlations were asymmetric.	Yahoo, Enron, corel5k and image datasets	RL, average AUC, coverage and average precision
[16]	SMILE	Reduced runtime and crucial to leverage unlabeled data with CL correlation.	It cannot exploit high-order CL correlations.	Cal500, bibtex and delicious datasets	RL, coverage, average precision, accuracy and adapted AUC
[17]	New Cost-sensitive MLL model with CPNL pairwise correlations	Better performance in terms of average precision and RL.	It does not discover CL correlations locally for incorporating with CPNL.	Emotions, image, scene, yeast, Enron, arts, education, recreation, science and business datasets	HL, subset accuracy, F1-example, RL and average precision
[18]	Sparsity regularized optimization method and KNN-like method	Better feasibility and efficient.	Improved method was needed for identifying the CL correlations-based features of test sample.	Emotions, genbase, medical, TCM1, TCM2, yeast, arts, computers, corel5k, education, science, social and society datasets	HL, coverage, one-error, RL and average precision
[19]	LF-LPLC	Improved learning performance.	Computational complexity was high.	Image, CAL500, emotions, language log, Enron, scene, yeast and Slashdot datasets	HL, RL, one-error, coverage and average precision
[20]	New scheme using the combined similarity constraints and ranking constraints	Reduced RL.	Accuracy was not effective.	Pascal VOC 2007, LabelMe, corel5k and CK+ datasets	Example-based accuracy, F1-score, accuracy and RL
[21]	New label recovery scheme under a semi-supervised configuration	Increased robustness and average accuracy.	High computational complexity.	Corel5k, ESP game, CAL500, yeast and REV1-v2 datasets	Average precision, AUC
[22]	MuENL and MuENLHD	High efficient in the dynamic learning setting.	It requires additional time for nonlinear mapping updation in the DS.	Birds, CAL500, emotions, Enron, yeast and Weibo datasets	Average precision and F1-measure

Table 1:- Comparison of Various MLL Techniques

III. CONCLUSION

In this article, a survey on recent MLL techniques is presented including with their merits and demerits to suggest the future scope. Based on this comparative analysis, it is concluded that the novel MuENL and MuENLHD can easily handle the sparse high-dimensional DSs via dimensionality minimization by streaming kernel PCA. Also, it can solve the practical issue of MLL using fresh labels efficiently. However, the choice of a proper configuration is an essential to improve the performance and it requires additional time for nonlinear mapping updation in the DS. As a result, it would require further research to solve these issues by utilizing advanced techniques that could increase the MLL performance efficiently.

REFERENCES

[1]. Younes, Z., & Denœux, T. (2010). Evidential multi-label classification approach to learning from data with imprecise labels. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (pp. 119-128). Springer, Berlin, Heidelberg.

[2]. Li, C., Wang, B., Pavlu, V., & Aslam, J. (2016). Conditional bernoulli mixtures for multi-label classification. In *International conference on machine learning* (pp. 2482-2491).

[3]. Karamanolakis, G., Iosif, E., Zlatintsi, A., Pikrakis, A., & Potamianos, A. (2016). Audio-based distributional semantic models for music auto-tagging and similarity measurement. *arXiv preprint arXiv:1612.08391*.

[4]. Zhang, M. L., & Zhou, Z. H. (2014). A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, 26(8), 1819-1837.

[5]. Gibaja, E., & Ventura, S. (2015). A tutorial on multilabel learning. *ACM Computing Surveys (CSUR)*, 47(3), 1-38.

[6]. Zhou, Z. H. (2016). Learnware: on the future of machine learning. *Frontiers of Computer Science*, 10(4), 589-590.

[7]. Tsoumakas, G., Katakis, I., & Vlahavas, I. (2011). Random k-labelsets for multilabel classification. *IEEE Transactions on Knowledge and Data Engineering*, 23(7), 1079-1089.

[8]. Kong, X., Ng, M. K., & Zhou, Z. H. (2013). Transductive multilabel learning via label set propagation. *IEEE Transactions on Knowledge and Data Engineering*, 25(3), 704-719.

[9]. Zhang, M. L., & Wu, L. (2015). Lift: Multi-label learning with label-specific features. *IEEE transactions on pattern analysis and machine intelligence*, 37(1), 107-120.

[10]. Liu, X., Bao, H., Zhao, D., & Cao, P. (2015). A label distance maximum-based classifier for multi-label learning. *Bio-medical materials and engineering*, 26(s1), S1969-S1976.

[11]. Pham, A. T., Raich, R., Fern, X. Z., & Pérez Arriaga, J. (2015). Multi-instance multi-label learning in the presence of novel class instances. In *Proceedings of the 32nd International Conference on Machine Learning*, 2427-2435.

[12]. Xu, X., Wang, W., & Wang, J. (2016). A three-way incremental-learning algorithm for radar emitter identification. *Frontiers of Computer Science*, 10(4), 673-688.

[13]. Mu, X., Ting, K. M., & Zhou, Z. H. (2017). Classification under streaming emerging new classes: A solution using completely-random trees. *IEEE Transactions on Knowledge and Data Engineering*, 29(8), 1605-1618.

[14]. Huang, J., Li, G., Wang, S., Xue, Z., & Huang, Q. (2017). Multi-label classification by exploiting local positive and negative pairwise label correlation. *Neurocomputing*, 257, 164-174.

[15]. Zhu, Y., Kwok, J. T., & Zhou, Z. H. (2017). Multi-label learning with global and local label correlation. *IEEE Transactions on Knowledge and Data Engineering*, 30(6), 1081-1094.

[16]. Tan, Q., Yu, Y., Yu, G., & Wang, J. (2017). Semi-supervised multi-label classification using incomplete label information. *Neurocomputing*, 260, 192-202.

[17]. Wu, G., Tian, Y., & Liu, D. (2018). Cost-sensitive multi-label learning with positive and negative label pairwise correlations. *Neural Networks*, 108, 411-423.

[18]. Zhang, J., Li, C., Cao, D., Lin, Y., Su, S., Dai, L., & Li, S. (2018). Multi-label learning with label-specific features by resolving label correlations. *Knowledge-Based Systems*, 159, 148-157.

[19]. Weng, W., Lin, Y., Wu, S., Li, Y., & Kang, Y. (2018). Multi-label learning based on label-specific features and local pairwise label correlation. *Neurocomputing*, 273, 385-394.

[20]. Wang, S., Chen, S., Chen, T., & Shi, X. (2018). Learning with privileged information for multi-label classification. *Pattern Recognition*, 81, 60-70.

[21]. Ma, J., & Chow, T. W. (2018). Robust non-negative sparse graph for semi-supervised multi-label learning with missing labels. *Information Sciences*, 422, 336-351.

[22]. Zhu, Y., Ting, K. M., & Zhou, Z. H. (2018). Multi-label learning with emerging new labels. *IEEE Transactions on Knowledge and Data Engineering*, 30(10), 1901-1914.