Mahasiswa-9-Tesis Oxapisi

by Budi Warsito

Submission date: 31-May-2020 10:53AM (UTC+0700)
Submission ID: 1334969942
File name: MAHASISWA-9-THESIS-C40-2019-Oxa_dkk-PID6235061.pdf (601.5K)
Word count: 3795
Character count: 20419

Comparative Experimental Study of Multi Label Classification using Single Label Ground Truth with Application to Field Majoring Problem

Oxapisi Adikhresna[1], Retno Kusumaningrum[2], Budi Warsito [3]
[1]*Department of Master of Information System,* [2]*Department of Informatics,*
[3] *Department of Statistics*
Diponegoro University
Semarang, Indonesia
[1]oxapisiv@gmail.com, [2]retno@live.undip.ac.id, [3]budiwrst2@gmail.com

***Abstract*— Researches on multi label classification methods generally use training data that already have multi label output as ground truth, but there are real-world problems where it is required to produce multi label prediction output but the available training data only have single label as ground truth. This study compared the performance of various multi label classification methods i.e. Ranking Support Vector Machine (Rank-SVM), Backpropagation for Multi Learning (BP-MLL), Multi Label K-Nearest Neighbor (ML-KNN), and Multi Label Radial Basis Function (ML-RBF) that were trained using multi label training data as intended and which were trained using single label training data. The dataset used in this research is an example of real-world problem, namely the personality-aptitude psychological test results is used to predict suitable majors in vocational high school. The results showed that hamming loss between the two was not far adrift so that it can be concluded that in certain problems, multi label classification methods can train single label and still produce multi label predictions with fairly good accuracy.**

Keywords—multi label classification; training multi label using single label; field majoring problem

I. INTRODUCTION

In machine learning, multi label classification tasks are ubiquitous in real life problem. For example, in text categorization, a text article or news may have various topics simultaneously [1], [2]; in scene classification, an image may have many objects in it [3], [4]; in medical field, a set of symptoms can be associated with many diseases [5], [6]. In all of those cases, the multi label ground truth can be obtained easily. Nevertheless, there are unique cases where the ground truth is difficult to obtain, and sometimes only one label may be obtained, e.g. in field majoring problem i.e. finding someone's suitable work or study field. A person may be suitable for more than one field, hence the multi label classification method is needed. However in real world practice, majority of individuals only try one field in a lifetime and can only assess whether he or she is suitable for that field, hence only one label can be obtained.

Multi label classification is tantamount to multi class classification. Multi class classification categorizing instances into precisely one of more than or equal to two classes, gives the assumption that each instance is allocated to one and only one label, for instance an animal can be either a cat or a dog but not both at the same time. In the multi-label classification there is no limit on number of the classes the instance can be allocated to [7]. Several traditional multi class classification algorithms have been adapted to tackle multi label classification problem by implementing the characteristic of multi label learning, that is the labels relevant to the instance in question is supposed to ranked higher than labels that is not relevant to that same instance [8].

This paper presents comparative experimental study of four cutting-edge multi label classification methods i.e. (1) Ranking Support Vector Machine (Rank-SVM), (2) Backpropagation for Multi Learning (BP-MLL), (3) Multi Label K-Nearest Neighbor (ML-KNN), and (4) Multi Label Radial Basis Function (ML-RBF). Each of those algorithms were derived from another algorithms that were originally used to solve multi class classification, which naturally are expected to perform better at multi label classification using single label ground truth training data than other algorithms.

Rank-SVM [9] is extended from basic Support Vector Machine by adapting maximum margin strategy to minimize ranking loss. It transforms the ranking problem into pairwise classification problem.

BP-MLL [8] is adapted from the popular Backpropagation (Basic-BP) multi-layer feed-forward neural network algorithm by introducing improved error function that considers multi label learning characteristics into account. There are also some revisions from Basic-BP's learning algorithm to minimize the new error function.

ML-KNN [10] is the first multi-label lazy learning algorithm that is adapted from traditional K-Nearest Neighbor algorithm. Lazy learning is a learning method where generalization of the training data will only begin when a query is inserted into the system [11]. It is originally motivated to solve problem like "people who viewed/purchased/listened to this item also like ...".

ML-RBF [12] is adapted from the popular Radial Basis Function neural network. ML-RBF consist of two layer, the first layer of ML-RBF is established by conducting clustering inspection on instances of every possible class, and then the second layer weights are learned by reducing a sum-of-squares

error function. By using this architecture, each output neuron is linked with all basis function of training vectors in all possible labels.

The mentioned methods have been applied in most of benchmark datasets in multi label learning such as YEAST dataset [9] that classify gene annotations [13], [14], [15], EMOTIONS dataset [16] that classify music into the feel emotions [17], [18], SCENE dataset [19] that contain images with multi scenes [18], REUTERS dataset [20] that classify news articles into topics [8], [21], and MEDIAMILL dataset [22] that contain video clips with multi scenes [17], [23]. In all of those datasets, the multi label ground truth is available. In our dataset on field majoring problem, only 1 label ground truth is available. This condition is still applicable in multi label classification methods, as in [8] stated that every sample needs to have at least one label and no sample may have all labels.

Our work sheds some light on wether if multi label classification methods trained with single label ground truth performs almost as good as which trained with multi label ground truth, and which methods gave best performance. This is achieved by training the mentioned methods using multi label training data as common architecture in multi label learning and using single label training data, using real-world problem dataset namely the personality-aptitude psychological test results used to predict suitable majors in vocational high school. Our contributions are as follows.

- We demonstrate that multi label classification methods trained using single label ground truth are able to predict multi label output with good performance.
- We demonstrate the methods performance slope on five different evaluation criterion of each methods when trained using single label and multi label ground truth.
- We demonstrate which methods performs superior when trained using single label ground truths, and which when trained using multi label ground truths, which is surprisingly result in two different methods.

The rest of this paper is organized as follows. Section II presents background knowledge and definition of multi label classification, such as illustrations of the methods used in this experiments. Section III presents all the details of the experimental configuration, such as dataset detail, parameter settings, and evaluation methodology used in this paper. Section IV presents experimental results and their discussion. Finally, section V concludes and gives direction for future research.

II. Multi Label Classification

Let X is a d-dimentional input instance space of categorical or numeric features, and $Y = \{y_1, y_2, \dots, y_q\}$ is label space where q is a finite possible class labels and $q > 1$. The goal of multi label classification methods is to output a multi label classifier $h : X \rightarrow 2^y$ from given training set $D = \{(x_i, Y_i) | 1 \leq i \leq m\}$. For each training instance, $x_i \in X$ and $Y_i \subseteq y$. For new unseen instance, the multi label classifier $h(.)$ predicts $h(x) \subseteq y$, giving a set of relevant labels for x. A perfect system will deliver output $Y_i > \overline{Y}_i$, where Y_i is set of label belongs to instance x_i and $\overline{Y}_i$ is the complementary set of Y_i [24].

Rank-SVM [9] defines a set of q linear classifiers $\{h_j(x) = \langle w_j, x \rangle + b_j = w_j . x + b_j | 1 \leq j \leq q\}$, each with weight vector w_j and bias b_j. They are adjusted to reduce the ranking loss utilizing quadratic programming and kernel to oversee non-linearity. The learning multi-label margin on the entire training set (1) [9] marks its capacity to correctly rank each label pair that belonging and not belonging to each training instance (x_i, Y_i) in the training set D. The margins for these pair of labels belonging and not belonging corresponds to the hyperplane $\langle w_j - w_k, x_i \rangle + b_j - b_k$.

$$\min_{(x_i, Y_i) \in S} \max_{(y_j, y_k) \in Y_i \times \overline{Y}_i} \frac{\langle w_j - w_k, x_i \rangle + b_j - b_k}{\|w_j - w_k\|} \quad (1)$$

In BP-MLL [8], the network is trained with gradient descent and backpropagation global error improved from simple sum-of-squares that occupy the characteristics of multi label learning (2) [8], where c_j^i is the output of one label belonging to instance i and c_k^i is the output of one label not belonging to instance i. BP-MLL architecture is consist of three layers: an input layer, a hidden layer, and an output layer. The input layer consists of d units corresponding each one to instance's feature vector. The output layer has q units corresponding each one to possible label/class.

$$E = \sum_{i=1}^{m} \frac{1}{|Y_i||\overline{Y}_i|} \sum_{(j,k) \in Y_i \times \overline{Y}_i} \exp\left(-(c_j^i - c_k^i)\right) \quad (2)$$

ML-KNN [10] consist of two stage procedures. First, it determines $N = \{(x_i, Y_i) | 1 \leq i \leq k\}$ as the set of k nearest neighbours and obtains a membership counting vector $c = (c_i, \dots, c_q) | c_j = \sum_{(x_i, Y_i) \in N} [\![y_j \in Y_i]\!]$ that stores the number of examples in the neighbourhood in each label, then it identifies labels of the unseen instance using MAP principle based on the prior and posterior probabilities (3) [10].

$$y_j = \begin{cases} 1 & if\ P(c_j | y_j = 1) P(y_j = 1) \geq P(c_j | y_j = 0) P(y_j = 0) \\ 0 & otherwise \end{cases} \quad (3)$$

ML-RBF [12] comprises of two layers of neurons. The first layer is consists of hidden neurons which is a basis function. Every each one of them represent a features vector and every each one of the output neuron represent a possible class. It is trained in two phase. To begin with, basis functions are learnt by applying k-means clustering on instances of every possible class, resulting in replacing the features vectors of first-layer basis functions with centroids of the clustered groups. Next, the second layer's weights are updated using new minimized sum-of-squares error function (4) [12].

$$E = \frac{1}{2} \sum_{i=1}^{m} \sum_{l=1}^{L} (yl(x_i) - t_l^i)^2 \quad (4)$$

III. Experimental Setup

A. Dataset

The dataset used in this study is the Personality-Aptitude Psychological Test of Semarang, Indonesia Public Vocational High School Dataset available at `www.psikologiporos.com/dataset` on field majoring problem. This dataset consist of 2387 data of vocational high school student's personality aptitude psychological test result and his/her current major. The personality aptitude psychological test result consists of 30 feature vectors which is used as inputs in this experiments. There are 19 possible major in this dataset that is used as outputs label/class in this experiments.

All students in this dataset had assess that him/herself is suitable in his/her current major and received above average grade in their semesterly report card. This current major is used as the single label ground truth trained in this experiments. This dataset also contains prediction of suitable majors from an expert i.e. pyschologist comprehensive analysis. Expert's prediction always containing the real major as well. This expert's prediction is used as the multi label ground truth trained in this experiments. Each methods is trained with single label and multi label ground truth, but all methods is evaluated using multi label ground truth.

B. Parameters Settings

Rank-SVM is trained using 8 degree polynomial kernels which cede the best performance in [9]. As recomended in [8], BP-MLL is trained with hidden neurons equal to 20% of the input neurons, 100 epochs, and 0.05 learning rate. As recomended in [10], ML-KNN is trained with 10 nearest neighbors and smoothing factor equal to 1 which yields the Laplace smoothing. As recomended in [12], ML-RBF is trained with parameters of scaling factor equal to 1 and number of centroids of the i-th class is set to be 0.01.

C. Evaluation Methodology

The dataset is split into a training and testing set using 5-Fold Cross Validation. The following popular multi label evaluation criterion used in [8], [12], [25] are utilized:

(1) *Hamming loss* (5) [8] rates the occurences of misclassifed labels, i.e. label that is relevant is not predicted or label that is not relevant is predicted. The method's performance is flawless when hamming loss is equal to 0. The method's performance is considered to be more superior when the value of hamming loss is lower.

$$hloss(f) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{Q}|f(x_i)\Delta Y_i| \quad (5)$$

Δ stands for the number of misclassified labels and Q stands for the number of possible label/class. p denotes number of training data.

(2) *One-error* (6) [8] rates the occurences of the predicted label that is first in rank is in fact not a relevant label. The method's performance is flawless when one-error is equal to 0. The method's performance is considered to be more superior when the value of one-error is lower.

$$one\text{–}error(f) = \frac{1}{p}\sum_{i=1}^{p}\left[\left[\arg\max_{y \in Y} f(x_i, y)\right] \notin Y_i\right] \quad (6)$$

(3) *Coverage* (7) [8] rates how far (how many iteration) we need to check the list of predicted labels to find all the actual relevant labels. The method's performance is considered to be more superior when the value of coverage is lower.

$$coverage(f) = \frac{1}{p}\sum_{i=1}^{p}\max_{y \in Y_i} rank_f(x_i, y) - 1 \quad (7)$$

(4) *Ranking loss* (8), (9) [8] rates the occurences of label pairs where their order are reversed, i.e. label that is not relevant is ranked higher than label that is relevant . The method's performance is flawless when ranking loss is equal to 0. The method's performance is considered to be more superior when the value of ranking loss is lower.

$$rloss(f) = \frac{1}{p}\sum_{i=1}^{p}\frac{|D_i|}{|Y_i||\bar{Y}_i|} \quad (8)$$

$$|D_i| = \{(y_1, y_2)|f(x_i, y_1) \leq f(x_i, y_2),\ (y_1, y_2) \in Y_i \times \bar{Y}_i\} \quad (9)$$

$|Y_i|$ denotes number of relevant label in instance i and $|\bar{Y}_i|$ denotes number of label that are not relevant.

(5) *Average precission* (10), (11) [8] rates the occurences of relevant labels ranked above other relevant labels. The method's performance is flawless when average precission is equal to 1. The method's performance is considered to be more superior when the value of average precission is higher.

$$avgprec(f) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{|Y_i|}\sum_{y \in Y_i}\frac{|L_i|}{rank_f(x_i, y)} \quad (10)$$

$$L_i = \{y'|rank_f(x_i, y') \leq rank_f(x_i, y),\ y' \in Y_i \quad (11)$$

IV. Result and Discussion

Table I presents result of algorithms trained using multi label ground truth on different evaluation criterions, where the best result of each criterion is presented in bold. ML-RBF achieves the best result in all evaluation criterion. This result is no surprise as ML-RBF achieved superior performance in most of benchmark database such as Yeast [9], Natural Scene [10], and Yahoo [26] dataset in [12].

Table II presents result of algorithms trained using single label ground truth on different evaluation criterions, where the best result of each criterion is presented in bold. BP-MLL achieves the best result in almost all evaluation criterion, in exception of hamming loss where ML-RBF dominates the rest of the algorithms.

TABLE I. RESULT USING MULTI LABEL GROUND TRUTH

Evaluation Criterion	Algorithm			
	Rank-SVM	*BP-MLL*	*ML-KNN*	*ML-RBF*
Hamming loss	0,2743	0,1712	0,1963	**0,1281**
One-error	0,4186	0,0721	0,1857	**0,0591**
Coverage	9,9169	7,0391	8,6537	**6,9977**
Ranking loss	0,2296	0,0840	0,1551	**0,0771**
Average precission	0,6333	0,8537	0,7513	**0,8683**

TABLE II. RESULT USING SINGLE LABEL GROUND TRUTH

Evaluation Criterion	Algorithm			
	Rank-SVM	*BP-MLL*	*ML-KNN*	*ML-RBF*
Hamming loss	0,2875	0,3160	0,2628	**0,2626**
One-error	0,7331	**0,4860**	0,6276	0,5281
Coverage	13,6250	**11,8366**	14,1395	12,1196
Ranking Loss	0,3949	**0,2795**	0,4076	0,3124
Average precission	0,4121	**0,5522**	0,4327	0,5134

While hamming loss is an instance based metrics, other evaluation criterion used are ranking based metrics. On that account, BP-MLL is prefered when output's rankings are mattered, while ML-RBF will achieve better result when output's rankings are overlooked. In this problem/dataset i.e. finding suitable major, rankings are mattered, seeing that in real-world there is a such things as "the most suitable major" and "less suitable major". Therefore, it is concluded that BP-MLL achieved best performance.

Fig.1 illustrates performance differences of algorithms trained using single label ground truth and multi label ground truth on different evaluation criterions. Each curves correspond to different algorithm, and the slope on each curves indicates how big is the performance difference between single label ground truth and multi label ground truth training.

Fig.1 shows that trainings using multi label ground truth always have better result as expected. However, in hamming loss, the average slope is at 9% which is fairly small. It is concluded that if output's rankings are overlooked, multi label learning algorithms that trained using single label ground truth can perform almost as good as those trained using multi label ground truth. In one-error, coverage, ranking loss, and average precission, the average slope is at 41%, 25%, 21%, and 30% respectively which is fairly big. It is concluded that if output's rankings are considered, other specific algorithm is needed.

V. CONCLUSION AND FUTURE WORK

Four cutting-edge multi label classification methods i.e. Rank-SVM, BP-MLL, ML-KNN, ML-RBF were trained using multi label ground truth as intended and using single label ground truth. When trained using multi label ground truth, ML-RBF shows superior performance than other methods. When trained using single label ground truth, BP-MLL shows superior performance than other methods. When compared, trainings using multi label ground truth always have better result as expected. However, hamming loss between the two is fairly small so it is concluded that multi label learning algorithms that trained using single label ground truth can perform almost as good as those trained using multi label ground truth. Unfortunately, performances slope in ranking based metrics is fairly big. Improvement from existing multi label classification methods to overcome this challenge is beneficial in future study.

REFERENCES

[1] J. Liu, W.C. Chang, Y. Wu, and Y. Yang, "Deep learning for extereme multi-label text classification," Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, 2017, pp. 115-124

[2] G. Chen, D. Ye, Z. Xing, J. Chen, and E. Cambria, "Ensemble application of convolutional and recurrent neural networks for multi-label text categorization," Proceedings of 2017 International Joint Conference on Neural Networks, Anchorage, USA, 2017, pp. 2377-2383

[3] Y. Liu, K. Wen, Q. Gao, X. Gao, and F. Nie, "SVM based multi-label learning with missing labels for image annotation," Pattern Recognition, vol. 78, pp. 307-317, 2018

[4] F. Zhu, H. Li, W. Ouyang, N. Yu, X. Wang, "Learning spatial regularization with image-level supervisions for multi-label image classification, Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017, pp. 5513-5522

[5] L. Dai, J. Zhang, C. Li, C. Zhou, and S. Li, "Multi-label feature selection with application to TCM state identification," Concurrency and Computation: Practice and Experience, 2018, e4634

[6] Q. Suo, F. Ma, Y. Yuan, M. Huai, W. Zhong, A. Zhang, J. Gao, "Personalized disease prediction using a CNN-based similarity learning method." Proceedings of 2017 IEEE International Conference on Bioinformatics and Biomedicine, Kansas City, USA, 2017, pp. 811-816

[7] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, no. 4, pp. 427-437, 2009

[8] M.L. Zhang and Z.H Zhou, "Multilabel neural networks with applications to function genomics and text categorization," IEEE Transactions on Knowledge and Data Engineering, vol. 18, no.10, pp. 1338-1351, 2006

[9] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, Vancouver, Canada, 2001, pp. 681-687

[10] M.L. Zhang and Z.H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," Pattern Recognition, vol. 40, pp. 2038-2048, 2007

[11] I. Hendirckx and A.V.D Bosch, "Hybrid algorithms with instance-based classification," Proceeding of the 16th European Conference on Machine Learning, Porto, Portugal, 2005, pp. 158-169

[12] M.L. Zhang, "ML-RBF: RBF neural networks for multi-label learning," Neural Processing Letters, vol. 29, no.2, pp.61-74, 200

[13] S. Agrawal, J. Agrawal, S. Kaur, and S. Sharma, "A comparative study of fuzzy PSO and fuzzy SVD-based RBF neural network for multi-label classification," Neural Computing and Applications, vol. 29, no.1, pp. 245-256, 2018

[14] W.J. Chen, Y.H. Shao, C.N. Li, and N.Y. Deng, "MLTSVM: A novel twin support vector machine to multi-label learning," Pattern Recognition, vol. 52, pp. 61-74, 2016

[15] B. Liu, D. Ma, C. Ma, and S. Wu, "Gene functional annotation via matrix completion," Proceedings of 2017 IEEE International Conference on Communications, Paris, France, 2017, pp. 1-6

[16] K. Trochidis, G. Tsoumakas, G. Kaliris, and I. Vlahavas, "Multi-label classification of music into emotions," EURASIP Journal on Audio Speech and Music Processing, vol.1, pp. 325-330, 2008

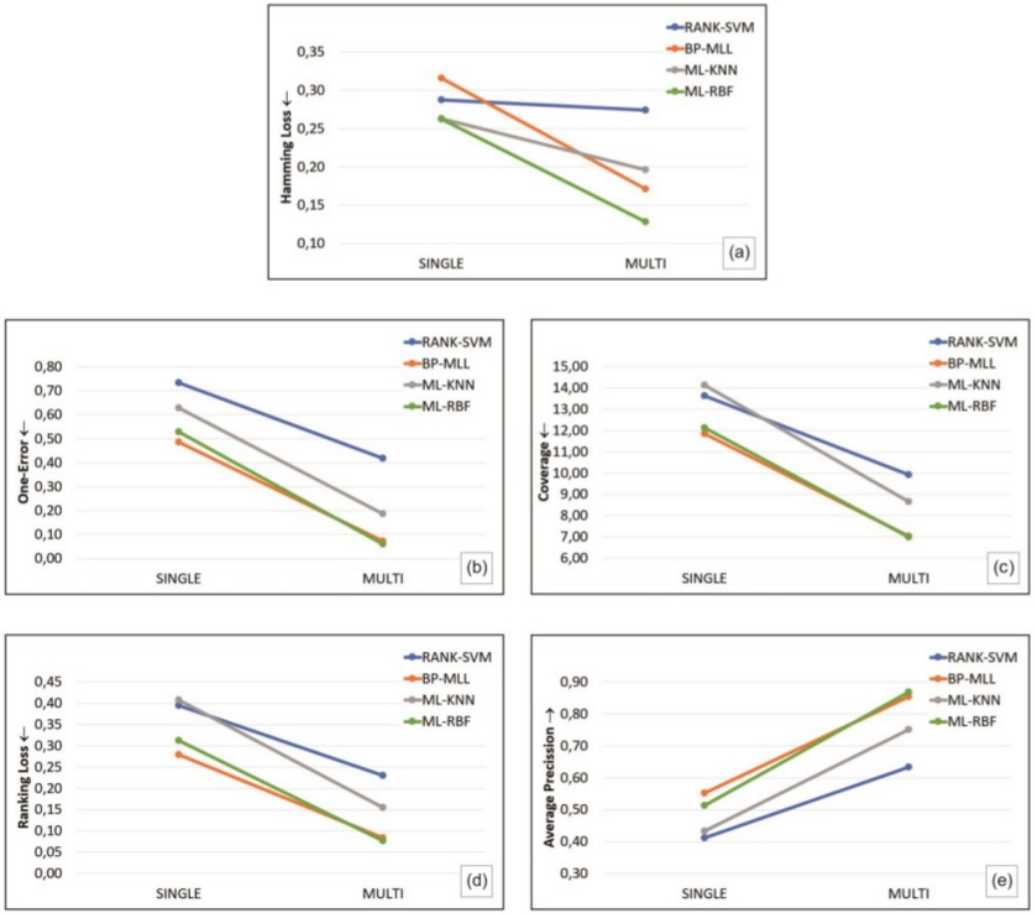

Fig. 1. The performance of multi label classification methods trained using multi label and single label ground truth. (a) *hamming loss;* (b) *one-error;* (c) *coverage;* (d) *ranking loss;* (e) *average precission*;

[17] B. Guo, C. Huo, J. Shan, and D. Yi, "Low rank multi-label classification with missing labels," Proceedings of 24th International Conference on Pattern Recognition, Beijing, China, 2018, pp. 417-422

[18] J. Xu, "An efficient multi-label support vector machine with a zero label," Expert System with Applications, vol. 39, no.5, pp. 4796-4804, 2012

[19] M. Bouttell, J. Luo, X. Shen, and C.M. Brown, "Learning multi-label scene classification," Pattern Recognition, vol.37, pp. 1757-1771, 2004

[20] D.D. Lewis, Y. Yang, T.G. Rose, and F. Li, "RCV1: A new benchmark collection for text categorization research," Journal of Machine Learning Research, vol.5, pp. 361-397, 2005

[21] S.T. Selvi, P. Karthikeyan, A. Vincent, V. Abinaya, G. Neeraja, and R. Deepika, "Text categorization using rocchio algorithm and random forest algorithm," Proceedings of Eighth International Conference on Advanced Computing, Chennai, India, 2017, pp. 7-12

[22] C.G.M. Snoek, M. Worring, J.C.V. Gemert, J.M. Geusebroek, and A.W.M Smeulders, "The challenge problem for automated detection of 101 semantic concepts in multimedia," Proceedings of ACM Multimedia, Santa Barbara, USA, 2006, pp. 421-430

[23] A. Dimou, G. Tsoumakas, V. Mezaris, I. Kompatsiaris, and I. Vlahavas, "An empirical study of multi-label learning methods for video annotation," Proceedings of 7th International Workshop on Content-Based Multimedia Indexing, Chania, Greece, 2009, pp. 19-24

[24] M.L. Zhang and Z.H. Zhou, "A review on multi-label learning algorithms," IEEE Transactions on Knowledge and Data Engineering, vol.26, no. 8, pp. 1819-1837, 2014

[25] C.A. Tawiah and V.S. Sheng, "Empirical comparison of multi-label classification algorithms," Proceedings of Twenty-Seventh AAAI Conference on Artificial Intelligence, Washington, USA, 2013, pp. 1645-1646

[26] N. Ueda and K. Saito, "Parametric mixture models for multi-labeled text," Proceedings of the 15th International Conference on Neural Information Processing Systems, Vancouver, Canada, 2002, pp. 721-728

Mahasiswa-12-Tesis Oxapisi

ORIGINALITY REPORT

7%	7%	7%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Eva Gibaja, Sebastián Ventura. "A Tutorial on Multilabel Learning", ACM Computing Surveys, 2015 Publication	2%
2	palm.seu.edu.cn Internet Source	1%
3	www.mirlabs.org Internet Source	1%
4	oa.upm.es Internet Source	1%
5	Sun, Xia, Jingting Xu, Changmeng Jiang, Jun Feng, Su-Shing Chen, and Feijuan He. "Extreme Learning Machine for Multi-Label Classification", Entropy, 2016. Publication	1%
6	Changsheng Li, Chong Liu, Lixin Duan, Peng Gao, Kai Zheng. "Reconstruction Regularized Deep Metric Learning for Multi-Label Image Classification", IEEE Transactions on Neural Networks and Learning Systems, 2019	1%

Publication

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		