International Journal of Soft Computing 12 (4): 218-223, 2017
ISSN: 1816-9503

A Comparative Analysis Between the three Main Approaches that are being used to Solve the Problem of Multi Label Classification

Raed Alazaidah, Farzana Kabir Ahmad and Mohamad Farhan Mohamad Mohsen
School of Computing, Universiti Utara Malaysia (UUM), Sintok, Kedah, Malaysia

Abstract: In this study, a comparative analysis between the three main approaches that are being used to solve the problem of Multi Label Classification (MLC) have been conducted. The goal of doing this comparative analysis is to evaluate the performance of the three main approaches and decide which approach to use with respect to the characteristics of the multi label data set. Different from other comparative analysis that have been conducted in the field of MLC which focused on doing comparative analysis between different methods or algorithms, this study is focusing on the performance of the three main approaches that these methods and algorithms are belonging to. These three main approaches are: first, second and high order approaches. Results show that the high order approach is more suitable for solving the problem of MLC and has a better predictive performance than the other two approaches.

Key words: Multi label classification, first order, high order, second order, MLC

INTRODUCTION

MLC was initially motivated by two domains: text classification (Rubin *et al*., 2012; Vogrincic and Bosnic, 2011) and medical diagnosis (Abbas *et al*., 2013; Shao *et al*., 2013) but since then, it was widely applied to many different real life domains such as: automatic image and video annotation (Peters *et al*., 2010) classification of song according to the emotions it invoke (Trohidis and Kalliris, 2008). Gene functionality detection (Barutcuoglu *et al*., 2006; Skabar *et al*., 2006) and protein functionality detection (Chan and Freitas, 2006).

Three main approaches that are being used to solve the problem of MLC: first order, second order and high order approaches (Gibaja and Ventura, 2015). These approaches have been categorized according to the degree of correlations among labels or among labels space and feature space that has been considered in developing the algorithm. Recently, it has been believed that any multi label classifier must take the correlations among labels into consideration, in order to enhance the predictive performance of the multi label classifier (Devi *et al*., 2015; Doppa *et al*., 2014; Jing *et al*., 2015; Xu *et al*., 2016; Zhang and Zhou, 2014) as well as to deal efficiently and effectively with the huge number of labels combinations, especially when the number of labels in a multi label data set is high or moderate. Although, many state-of-art MLC methods state that exploiting correlations among labels can help in improving the predictive performance but there is no systematic study about how exploiting labels correlations solely can improve the predictive performance, regardless to the base classifiers that have been used (Chekina *et al*., 2011; Xie *et al*., 2013). No systematic study correlates between the effectiveness of exploiting correlations among labels and the characteristics of the multi label data sets. According to Doppa *et al*. (2014), there is a lack of awareness between the sparsity property of multi label data sets and the gain of exploiting correlations among labels. This study intends to determine the effectiveness of the three main approaches in solving the problem of MLC as well as to identify the gained benefits from capturing and exploiting the correlations among labels in term of predictive performance.

Literature review: Based on the degree of correlations among labels that has been considered in the MLC algorithm or method, MLC algorithms and methods could be categorized into three main approaches as follow.

First order approach: Algorithms that are considered as first order, handle the problem of MLC through an individual and separate learning of each label (Alazaidah and Ahmad, 2016). These algorithms ignore any correlations or dependencies among labels which may lead to relatively low predicative accuracy, especially when the number of labels is too high and Label Cardinality (LC) is also large. Despite this, relatively low predictive accuracy, these algorithms tend to be simple and efficient in term of speed. Examples of the first order

Corresponding Author: Raed Alazaidah, School of Computing, Universiti Utara Malaysia (UUM), Sintok, Kedah, Malaysia

algorithms could be found by Boutell *et al.* (2004) Clare and King (2001) Comite *et al.* (2003), Zhang and Zhou (2007).

Second order approach: In the second order approach, only the pairwise relationships are considered among labels. Two types of algorithms that consider the second order of correlations among labels, could be found in the literature. The first type produces bipartitions of relevant and irrelevant labels while the second type performs a pairwise comparison between labels to produce a final ranking (Elisseeff and Weston, 2001; Furnkranz *et al.*, 2008; Ghamrawi and McCallum, 2005; Schapire and Singer, 2000; Ueda and Saito, 2002).

High order approach: In the high order MLC algorithms, the learning step considers all the high order correlations and dependencies among label such as the influence of all labels on a specific label and vice versa or addressing connections among random subsets of labels (Alazaidah *et al.*, 2015). High order MLC algorithms capable of exploiting strong correlations but at the same time usually suffer from high complexity and less scalability. Examples of algorithms that exploit high order correlations among labels could be found by Godbole and Sarawagi (2004), Read *et al.* (2011), Tsoumakas *et al.* (2008) and Yan *et al.* (2010).

MATERIALS AND METHODS

Experimental work: To conduct this experiment, 5 different data sets from different domains and sizes were used, 3 of these data sets are of regular size (emotion, scene, yeast) and the other two data sets are large size data sets. Table 1 shows the different characteristics of the multi label data sets that were used in the experiment. In this experiment, 9 different methods and algorithms have been used to represent the three main approaches. These methods have been chosen due to their popularity in the field of MLC as well as their high predictive performance which has been proven, either by the authors of the methods themselves or by others. For the first order approach, two algorithms have been used: BR and ML-KNN. The first method is a Problem Transformation Method (PTM) while the second one is an Algorithm Adaptation Method (AAM). High order Approach has been represented using 5 different methods or algorithms: LP, PS, EPS, CC and RAKEL. Lastly, the second approach has been represented using two algorithms: BP-MLL and CLR.

It has been believed that the evaluation process of any multi label classifier should be based on using different evaluation metrics and not only one metric. Based on that it has been decided to use 5 different evaluation metrics in the experiment. These evaluation metrics are: accuracy, F1-measure, Hamming Loss (HL), one-error and coverage.

For the PTM, the Support Vector Machine (SVM) was used as the base classifier for all PTMs to eliminate the effect of the base classifier on the final result. Mulan was used in the implementation of all methods and algorithms that were used in this experiment. Mulan is a Weka based (Witten *et al.*, 2011) package of Java classes for MLC. The experiment was conducted using 10-fold cross validation.

The experiment has been conducted in two phases, the first phase considered only those multi label data sets with regular size. Regular size means data sets with a total number of labels <15 labels and total number of instances <5000 instances. Data sets that were used to represent the regular size in the experiment are: yeast, scene and emotions. The second phase of the experiment considered large size data sets where the number of labels is ≥15 class labels and the number of instances is ≥5000. Data sets that were used to represent the large size in the experiment are: EMC2007 and Ohsumed.

RESULTS AND DISCUSSION

Regular size data sets: Table 2-4 depict the results of the experiment on the regular size data sets. The best values for each metrics are those in bold. It is worth mentioning here, for accuracy and F1-measure, the higher the value of the metric, the better the performance of the classifier. For HL, coverage and one-error metrics, the lower the value of the metrics, the better the performance of the classifier. By analyzing the previous results on the three regular size data sets it can be concluded that: for the yeast data set, no approach could be described as the optimal or the best approach. It can be obvious noted that the first order

Table 1: Multi label data sets characteristics

Data set	Instances	Attributes	Label	DLS	LC	LD	Domain
Yeast	2417	103	14	198	4.327	0.302	Biology
Scene	2712	294	6	15	1.074	0.179	Media
Emotions	593	72	6	27	1.868	0.311	Media
TMC2007	28596	500	22	1344	2.160	0.098	Text
Ohsumed	13929	1002	23	1142	1.660	0.072	Text

approach has the best results for the HL, one-error and F1-measure metrics. The second best approach was the high order approach with the best results for the accuracy and F1-measure metrics. The second order approach was the third best approach since it has only the best value in only one metric which is the coverage. For the emotions data set, it is clearly noted that the best results belong to the high order approach, since the high order approach has the best values for, all the five evaluation metrics. For the scene data set the situation is nearly similar to the situation in the emotions data set. High order approach has the most of the best results with only one exception the coverage metric where the best value belongs to the first order approach. A general analytical comparison between the three approaches leads to the following result: the high order approach is the best approach, since it wins eleven times on the three data sets. The second best approach is the first order approach which wins four times on the three data sets. The least best approach is the second order approach with number of wins equals to one only.

Based on the previous mentioned results, it can be clearly noted that the high order approach is the best approach and the most appropriate approach to be used with the regular size multi label data sets. The second best approach is the first order approach. Using the second order approach does not reflect any benefits with regular size data sets. Regarding to the computational complexity, first order approach tends to be simple, efficient and of a low computational complexity when considering regular size data sets. On the other hand, high order approach is of high computational complexity in general but since the data sets are of regular size thus, its computational complexity will be relatively accepted. Based on that, there is no urgent need to consider the computational complexity of the approach when dealing with regular size

Table 2: Yeast data set results

Algorithm	ACC	HL	F1-measure	One error	Coverage
First order					
BR	0.499	0.199	0.637	0.256	9.096
ML-KNN	0.520	0.193	0.654	0.234	6.301
High order					
LP	0.530	0.206	0.643	0.267	8.065
PS	0.533	0.205	0.647	0.986	11.830
EPS	0.537	0.207	0.654	0.265	7.841
CC	0.489	0.211	0.619	0.256	8.674
RAKEL	0.487	0.207	0.625	0.255	9.273
Second order					
BP-MLL	0.185	0.322	0.210	0.805	2.523
CLR	0.514	0.226	0.405	NG	NG

Table 3: Emotions data set results

Algorithm	ACC	HL	F1-measure	One error	Coverage
First order					
BR	0.532	0.194	0.642	0.292	2.401
ML-KNN	0.366	0.262	0.493	0.386	2.327
High order					
LP	0.584	0.198	0.687	0.310	2.235
PS	0.599	0.192	0.701	0.902	4.364
EPS	0.599	0.193	0.703	0.300	2.138
CC	0.554	0.207	0.655	0.347	2.318
RAKEL	0.592	0.186	0.706	0.260	1.989
Second order					
BP-MLL	0.276	0.433	0.389	0.668	3.159
CLR	0.557	0.214	0.668	NG	NG

Table 4: Scene data set results

Algorithm	ACC	HL	F1-measure	One error	Coverage
First order					
BR	0.594	0.106	0.627	0.346	0.990
ML-KNN	0.691	0.085	0.759	0.226	0.456
High order					
LP	0.735	0.090	0.755	0.246	0.733
PS	0.751	0.084	0.769	0.904	4.225
EPS	0.751	0.085	0.769	0.225	0.689
CC	0.696	0.103	0.714	0.278	0.894
RAKEL	0.671	0.097	0.602	0.237	0.637
Second order					
BP-MLL	0.212	0.579	0.663	0.466	7.447
CLR	0.695	0.101	0.736	NG	NG

Table 5: TMC2007 data set results

Algorithm	ACC	HL	F1-measure
First order			
BR	0.541	0.064	0.547
High order			
EPS	0.549	0.069	0.573
ECC	0.517	0.068	0.496
RAKEL	0.549	0.068	0.577
Second order			
CLR	0.506	0.078	0.577

Table 6: Ohsumed data set results

Algorithm	ACC	HL	F1-measure
First order			
BR	0.361	0.074	0.370
High order			
EPS	0.424	0.074	0.366
ECC	0.426	0.063	0.414
RAKEL	0.383	0.075	0.392
Second order			
CLR	0.383	0.075	0.407

data sets. Therefore, high order approach still the best choice to be selected with the regular size multi label data sets.

Large size data sets: Table 5 and 6 depict the result of the experiment on the large size multi label data sets. The best values for each metrics are those in bold. By analyzing the previous results on the two large size data sets, it can be concluded that: for the TMC2007 data set, the high order approach is the best choice, since it has the best value for both the accuracy and F1-measure metrics. The first order and the second order approaches share the second best choice. First order approach has the best HL metric and second order approach has the best F1-measure metric.

For the Ohsumed data set, it is obvious clear that the high order approach has the best values for all metrics. The second best choice is the second order approach, since, it has a better performance for accuracy and F1-measure metrics while the first order approach has a better HL value than the second order approach.

In general, it can be stated that: the high order approach is more appropriate to be used with large size data sets. The high order approach wins 5 times out of six for both data sets while the second order approach wins only 1 time which is the same number of wins for the first order approach.

Regarding to computational complexity, high order approach is known to have a high computational complexity with large data sets but at the other hand, first order approach has to deal with large number of labels in the large size data sets, in addition to the huge loss of valuable information, due to neglecting the correlations among labels. For the second order approach, the number of pairwise comparisons grows in an exponential way and the benefits of discovering these pairwise correlations are very limited as it can be seen from the results in Table 5 and 6. Therefore, the high order approach is again the best choice to be used when considering large size data sets.

From both experiments on regular size and large size data sets, it can be concluded that in general the high order approach is the best choice for solving the problem of MLC, especially with large size data sets. Also, the second order approach is the second best choice especially with large data sets. Based on the last two conclusion, it can be assure that exploiting correlations among labels in the multi label data sets is a very important issue in enhancing the predictive performance of any MLC algorithms.

CONCLUSION

In this study, an experiment has been conducted to determine what is the best approach that should be used to handle the problem of MLC properly. The experiment considered both types of multi label data sets: regular size and large size data sets. The experiment showed that the high order approach is more appropriate and leads to better predictive performance than the other two approaches, especially with large data sets. Another finding of the experiment is that second order approach is more appropriate to be used with multi label data sets than the first order approach. Our main conclusion of this experiment is that exploiting correlations among labels is of great benefits in enhancing the predictive performance of a multi label classifier.

SUGGESTIONS

For future work, more investigations should be carried out to consider other characteristics of the multi label data sets that determine when exploiting correlations among labels could be of great benefits. These characteristics include the average number of significant relationships for each label in the label set as well as Label Cardinality (LC) and Label Diversity (LD).

REFERENCES

Abbas, Q., M.E. Celebi, C. Serrano, I.F. Garcia and G. Ma, 2013. Pattern classification of dermoscopy images: A perceptually uniform model. Pattern Recognit., 46: 86-97.

Alazaidah, R. and F.K. Ahmad, 2016. Trending challenges in multi label classification. Intl. J. Adv. Comput. Sci. Appl., 1: 127-131.

Alazaidah, R., F. Thabtah and A.Q. Radaideh, 2015. A multi-label classification approach based on correlations among labels. Intl. J. Adv. Comput. Sci. Appl., 6: 52-59.

Barutcuoglu, Z., R.E. Schapire and O.G. Troyanskaya, 2006. Hierarchical multi-label prediction of gene function. Bioinf., 22: 830-836.

Boutell, M.R., J. Luo, X. Shen and C.M. Brown, 2004. Learning multi-label scene classification. Pattern Recognit., 37: 1757-1771.

Chan, A. and A.A. Freitas, 2006. A new ant colony algorithm for multi-label classification with applications in bioinfomatics. Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, July 08-12, 2006, ACM, New York, USA., ISBN:1-59593-186-4, pp: 27-34.

Chekina, L., L. Rokach and B. Shapira, 2011. Meta-learning for selecting a multi-label classification algorithm. Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops, December 11-11, 2011, IEEE, Vancouver, British Columbia, Canada, ISBN:978-1-4673-0005-6, pp: 220-227.

Clare, A. and R.D. King, 2001. Knowledge discovery in multi-label phenotype data. Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery, September 3-5, 2001, Springer, Berlin, Germany, pp: 42-53.

Comite, D.F., R. Gilleron and M. Tommasi, 2003. Learning multi-label alternating decision trees from texts and data. Proceedings of the 3rd International Workshop on Machine Learning and Data Mining in Pattern Recognition, July 5-7, 2003, Springer, Berlin, Germany, pp: 35-49.

Devi, P.S., R. Baskaran and S. Abirami, 2015. Multi-label learning with class-based features using extended centroid-based classification technique (CCBF). Procedia Comput. Sci., 54: 405-411.

Doppa, J.R., J. Yu, C. Ma, A. Fern and P. Tadepalli, 2014. HC-Search for multi-label prediction: An empirical study. Proceedings of the 28th International Conference on AAAI Artificial Intelligence, June 0-21, 2014, Oregon State University, Corvallis, Oregon, USA., pp: 1795-1801.

Elisseeff, A. and J. Weston, 2001. A Kernel Method for Multi-Labelled Classification. In: Advances in Neural Information Processing Systems, Kearns, M.S., S.A. Solla and D.A. Cohn (Eds.). Bradford Bks, USA., pp: 681-687.

Furnkranz, J., E. Hüllermeier, L.E. Mencia and K. Brinker, 2008. Multilabel classification via calibrated label ranking. Mach. Learn., 73: 133-153.

Ghamrawi, N. and A. McCallum, 2005. Collective multi-label classification. Proceedings of the 14th ACM International Conference on Information and Knowledge management, Oct. 31-Nov. 5, Bremen, Germany, pp: 195-200.

Gibaja, E. and S. Ventura, 2015. A tutorial on multilabel learning. ACM. Comput. Surv., 47: 52-52.

Godbole, S. and S. Sarawagi, 2004. Discriminative Methods for Multi-Labeled Classification. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Dai, H., R. Srikant and C. Zhang (Eds.). Springer, Berlin, Germany, ISBN:978-3-540-22064-0, pp: 22-30.

Jing, L., L. Yang, J. Yu and M.K. Ng, 2015. Semi-supervised low-rank mapping learning for multi-label classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 07-12, 2015, IEEE, Boston, Massachusetts, USA., ISBN:978-1-4673-6964-0, pp: 1483-1491.

Peters, S., L. Denoyer and P. Gallinari, 2010. Iterative annotation of multi-relational social networks. Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining, August 9-11, 2010, IEEE, Odense, Denmark, ISBN:978-1-4244-7787-6, pp: 96-103.

Read, J., B. Pfahringer, G. Holmes and E. Frank, 2011. Classifier chains for multi-label classification. Mach. Learn., 85: 333-359.

Rubin, T.N., A. Chambers, P. Smyth and M. Steyvers, 2012. Statistical topic models for multi-label document classification. Mach. Learn., 88: 157-208.

Schapire, R. and Y. Singer, 2000. BoosTexter: A boosting-based system for text categorization. Mach. Learn., 39: 135-168.

Shao, H., G. Li, G. Liu and Y. Wang, 2013. Symptom selection for multi-label data of inquiry diagnosis in traditional Chinese medicine. Sci. China Inf. Sci., 56: 1-13.

Skabar, A., D. Wollersheim and T. Whitfort, 2006. Multi-label classification of gene function using MLPs. Proceedings of the International Joint Conference on Neural Networks IJCNN'06, July 16-21, 2006, IEEE, Vancouver, British Columbia, Canada, ISBN:0-7803-9490-9, pp: 2234-2240.

Trohidis, K., G. Tsoumakas, G. Kalliris and I.P. Vlahavas, 2008. Multi-label classification of music into emotions. Proceedings of the 9th International Conference on ISMIR Music Information Retrieval Vol. 8, September 14-18, 2008, Drexel University, Philadelphia, Pennsylvania, ISBN:978-0-615-24849-3, pp: 325-330.

Ueda, N. and K. Saito, 2002. Parametric Mixture Models for Multi-Labeled Text. In: Advances in Neural Information Processing Systems, Dietterich, T.G., S. Becker and Z. Ghahramani (Eds.). MIT Press, USA., ISBN: 9780262042062, pp: 721-728.

Vogrincic, S. and Z. Bosnic, 2011. Ontology-based multi-label classification of economic articles. Comput. Sci. Inf. Syst., 8: 101-119.

Witten, I.H., E. Frank and M.A. Hall, 2011. Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edn., Elsevier, New York, ISBN-13: 9780080890364, Pages: 664.

Xie, S., X. Kong, J. Gao, W. Fan and S.Y. Philip, 2013. Multilabel consensus classification. Proceedings of the 2013 IEEE 13th International Conference on Data Mining (ICDM), December 7-10, 2013, IEEE, Dallas, Texas, USA., ISBN:978-0-7695-5108-1, pp: 1241-1246.

Xu, S., X. Yang, H. Yu, D.J. Yu and J. Yang *et al*., 2016. Multi-label learning with label-specific feature reduction. Knowl. Based Syst., 104: 52-61.

Yan, Y., G. Fung, J.G. Dy and R. Rosales, 2010. Medical coding classification by leveraging inter-code relationships. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 25-28, 2010, ACM, New York, USA., ISBN:978-1-4503-0055-1, pp: 193-202.

Zhang, M.L. and Z.H. Zhou, 2007. ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognit., 40: 2038-2048.

Zhang, M.L. and Z.H. Zhou, 2014. A review on multi-label learning algorithms. IEEE. Trans. Knowl. Data Eng., 26: 1819-1837.