Evaluating Conditional and Unconditional Correlations Capturing Strategies in Multi Label Classification

Raed Alazaidah[1], Farzana Kabir Ahmad[1], Mohamad Farhan Mohamad Mohsen[1], Ahmad Kadri Junoh[2], and Mo'ath Allwaise[2]
[1]*School of Computing, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia.*
[2]*Institute of Engineering Mathematics, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia.*
raeddiab@yahoo.com

***Abstract*—In the last few years, multi label classification has attracted many scholars and researchers; due to the increasing number of modern domains that are applicable to this general type of classification. Recently, it has been believed by many researchers that the best way to handle the problem of multi label classification is by exploiting the correlations among labels. Two main strategies have been utilized to capture these correlations: conditional correlations and unconditional correlations capturing strategies. In this paper, an extensive evaluation of both strategies has been conducted, to determine the best strategy to handle multi label classification, with respect to the size of the data set and the optimized loss function. Results showed that the unconditional correlations capturing strategy overcomes the conditional correlations capturing strategy in all multi label data sets that have been used in this experiment.**

***Index Terms*—Classification; Conditional Correlation; Multi Label Classification; Unconditional Correlations.**

I. INTRODUCTION

From statistical point of view, and according to the strategy that has been used in capturing the correlations among labels, the algorithms that captured and exploited correlations among labels in Multi Label Classification (MLC) could be categorized into two main groups: conditional and unconditional correlations strategies. While the later captures dependencies and correlations among label, regardless and independent from features values, the former captures and exploits these correlations with respect to specific observations (instances) [1].

Most MLC algorithms that capture and exploit correlations among labels and claim that these correlations enhanced the predictive performance of the classifier, do not consider or correlate this enhancement to the type of correlations being used [2]. Few researches stated the specific type of correlations capturing strategy which was used in their proposed algorithms [3]. Many researches in the domain of MLC claim enhancement and improvement of the predictive performance, but unfortunately never carefully questioned the conditions that make exploiting correlations among labels beneficial.

This Paper aims to elaborate on the issue of labels correlations and dependencies in depth, which will result in a better understanding of exploiting correlations among labels in MLC. Firstly, a formal definition of both conditional and unconditional correlations capturing strategies will be presented. Then, a classification of Multi Label Learning (MLL) algorithms according to the strategy that has been used in capturing the correlations will be discussed.

To give a formal definition of unconditional correlations capturing strategy, consider Y= $(y_1, y_2,\ldots,y_n)$, to be a vector of labels in a multi label data set.
Then, $\mathbf{P(Y)} = \prod_{i=1}^{n} p(i)\,(yi)$, is called unconditionally correlations capturing strategy.

As it can be seen from the formal definition of unconditional correlations capturing strategy, the correlations among labels do not consider the instances (observations) of the data set. On the other hand, the conditional correlations capturing strategy considers the concrete observations (instances), and it is defined formally as follows:

For a given vector of labels in a multi label data set, Y= $(y_1, y_2,\ldots,y_n)$. Then, $\mathbf{P_x(Y)} = \prod_{i=1}^{n} px(i)\,(yi)$, is called conditionally correlations capturing strategy.

To make it simpler, conditional correlations capturing strategy considers the instances in the data set when capturing the correlations among labels, while unconditional correlations capturing strategy ignores the instances (observations) in the data set, when capturing the correlations among labels. In fact, both types could be beneficial in improving the predictive performance of a multi label classifier.

The paper is organized as follows: next section briefly describes the related work of both strategies, and several well-known algorithms that represent both types of correlation capturing strategy. Section 3 describes the methodology of the research, and Section 4 presents the conclusion and the future work.

II. RELATED WORK

This section will discuss the two main strategies that are being used in capturing the correlations among labels in MLC.

A. MLC Algorithms Based on Unconditional Correlations Strategy

Capturing correlations among labels by following the unconditional type of correlations has attracted many researchers. The reason for that is, the small search space for the unconditional correlations when compared to the conditional correlations [4]. Unconditional correlations strategy focuses only on the label space of the multi label data sets, and ignores the observations (instances), which makes it simpler and easier to implement.

One of the first algorithms that captured correlations among labels by following the unconditional strategy of correlations capturing is the Label Powerset (LP) algorithm. LP transforms the problem of MLC to a problem of multi class classification, by considering each new label set as a new class in a multi class classification, and then applies any single label classifier in the learning phase [5].

Another algorithm that captured the unconditional correlations among labels, is the RAndom K labELsets (RAKEL) algorithm [6]. In fact, RAKEL is an ensemble algorithm that comprises several LP classifiers. These classifiers are defined on different subsets of labels that are randomly chosen. RAKEL depends heavily on two main parameters: number of the base classifiers and the size of label subsets. The final prediction made by RAKEL is obtained through combining the prediction of the several ensemble members on the label subsets [7]. Although, RAKEL has a competitive performance, its theoretical principle is not clear; as it is not obvious clear what loss function it tries to optimize [2].

Classifier Chains (CC) and its two main extensions: Ensemble of Classifier Chains (ECC) [8] and Probabilistic Classifier Chains (PCC) [9] follow the unconditional correlations strategy. CC transforms the multi label data set into a single label data set, and then trains a binary classifier for each label. A chain of classifiers is then built, where binary attributes are added to each classifier for all of the predictions of the previous classifiers [8]. PCC proposed a solution for optimizing the best chains orders by predicting labels combinations in a stepwise manner through augmenting (k-1) binary features to the input space. In fact, this solution is infeasible and thus, leads to high computational complexity [2].

ECC depends for its final prediction on averaging several CC predictions. Hence, it has the same limitations and drawbacks of the CC algorithm, in addition to the ambiguity of the loss function that ECC intends to optimize.

The Pruned Set (PS) method -as its name indicates- prunes all the label sets that have a frequency less than a specific user predefined threshold [10]. This strategy may solve the problem of the high computational complexity of LP, and the problem of the imbalance class distribution, but at the same time it imposed a new problem, which is the huge information loss due to the pruned labels combinations. The Ensemble of Pruned Set (EPS) method constructs a number of pruned sets through sampling the training set, and build the final prediction using a voting schema and a user predefined threshold, in order to form new combinations of labels [10].

B. *MLC algorithms Based on Conditional Correlations Strategy*

The benefits of capturing conditional correlations between labels and features encouraged many researchers to adapt several single label classification algorithms to handle multi label data sets. It can be clearly seen that, conditional correlations capturing strategy is intensively used with Algorithm Adaptation Methods (AAMs), while unconditional correlations strategy tends to be more suitable to Problem Transformation Methods (PTMs).

The problem of MLC could be seen as a special type of structured output, and based on this point of view, it could be solved by using Structural Support Vector Machine (SSVM) as in [11]. RANK-SVM is another multi label ranking algorithm based on SVM. This algorithm utilizes a set of linear classifiers in order to minimize the ranking loss metric, with the help of kernel trick to handle nonlinear problems [12].

Also, several Multi Label Ranking (MLR) algorithms have utilized conditional correlations in the ranking process of relevant labels to a test instance. For example, Ranking by Pairwise Comparisons (RPC) that depends on performing exhaustive pairwise comparisons to achieve the ranking of all labels to a given test case [13]. RPC has a limitation of considering only a second order correlation, which makes it suitable only for small data sets. *Calibrated* Label Ranking (CLR) is another pairwise method that enhanced RPC by introducing a calibration label. This virtual label (L0) works as a split point between relevant labels, and irrelevant labels [13]. As in RPC the CLR method suffers from space complexity, and computational complexity too.

An algorithm that followed the conditional correlations strategy was presented in [14]. This algorithm was called (LEAD), short for Multi-Label Learning by Exploiting Label Dependency. LEAD is based on using a Bayesian network structure that aims to capture and encode the conditional correlations between labels and features set. The main goal of the LEAD algorithm is to enhance the predictive performance of MLL by exploiting conditional correlations. It manages to do that by using a Bayesian network, which is in its essence a Directed Acyclic Graph (DAG).

Back Propagation for MultiLabel Learning (BP-MLL) algorithm [15] is an adaptation of the traditional multi-layer, feed-forward neural network to handle multi label data. The net was trained with gradient descendent and error back propagation, with an error function that took into account the multi label data. Experimental results showed a competitive performance of the BP-MLL algorithm in genomics and text categorization domains.

Conditional Dependency Network- Logistic Regression (CDN-LR) and Conditional Dependency Network- Support Vector Machine (CDN-SVM) are two a cyclic directed graphical based model that were presented in [16]. CDN-LR and CDN-SVM effectively captured and exploited conditional correlations among labels and instances in order to improve the predictive performance of MLC.

Instance-Based Learning by Logistic Regression (IBLR) is a hybrid approach that is capable of capturing inter correlations among labels. The main idea of this approach is to consider class labels of the neighbouring examples as features of unseen test cases along with the reduction of IBL to LR [17].

To summarize, two main strategies are being used in MLC to capture correlations and dependencies. The first strategy considers only those correlations among labels and ignores any correlations between the labels and the observations. This strategy has been called as unconditional correlations strategy. The second strategy considers the correlations between labels and observations, and it has been called as conditional correlations strategy. Table 1 summarizes the main differences between the two strategies.

Table 1
Conditional and unconditional correlations capturing strategies

Conditional Correlations Strategy	Unconditional Correlations Strategy
Captures correlations among labels and observations (Features).	Captures correlations among labels only.
More oriented toward AAMs	More oriented toward PTMs
Usually, more oriented to specific domains	More general to any domain
More time consuming	Less time consuming
More Complexity	Less Complexity

III. The Research Methodology

A. Data Collection

In order to evaluate the performance of both strategies on the predictive performance of MLC algorithms, data has been collected from several published articles. The algorithms have been divided into two groups: algorithms that capture conditional correlations, and algorithms that capture unconditional correlations. For the first group the following algorithms have been chosen to be used in this paper: LP, RAKEL, CC, ECC, PS and EPS. For the second group, the selected algorithms are: CLR, BPMLL, LEAD, RANK-SVM, CDN-LR, CDN-SVM and IBLR. Six evaluation metrics have been used in the evaluation process of the selected algorithms. These metrics are: Hamming Loss (H.L), Accuracy (ACC), F1-Measure, Micro-F1, Macro-F1and Exact Match (EM). More details about these evaluation metrics could be found in [5]. Data sets were divided into two groups: regular size data sets that have 15 or fewer labels, and large size data sets that have more than 15 labels. Table 2 describes the main characteristics of the data sets.

Table 2
Data set main characteristics

Size	Data set	Instances	Attributes	Label	LC	Domain
Regular	Yeast	2417	103	14	4.327	Biology
	Scene	2712	294	6	1.074	Media
	Emotions	593	72	6	1.868	Media
Large	TMC2007	28596	500	22	2.16	Text
	Ohsumed	13929	1002	23	1.66	Text

B. Unconditional Correlations Algorithms Data Collection

Table 3 to Table 5 depict the collected data on the regular size data sets, using several algorithms that capture the unconditional correlations among labels. Bold values represent the best values, where "NG" represents the "Not Given" values.

Table 3
Yeast data set

Algorithm	H.L↓	ACC↑	F1 Measure↑	Micro-F1↑	Macro-F1↑	Exact Match↑
LP	0.206	0.530	0.643	0.643	0.418	**0.260**
RAKEL	0.207	0.487	0.625	0.624	0.333	0.128
CC	0.211	0.489	0.619	0.620	0.403	0.196
ECC	0.623	0.298	0.458	0.459	0.469	0.001
PS	**0.205**	0.533	0.647	0.645	0.396	0.258
EPS	0.207	0.537	0.654	0.650	0.515	0.253
Average	0.276	0.479	0.607	0.606	0.422	0.182

Table 4
Scene data set

Algorithm	H.L↓	ACC↑	F1 Measure↑	Micro-F1↑	Macro-F1↑	Exact Match↑
LP	0.090	0.735	0.755	0.745	0.754	0.696
RAKEL	0.097	0.671	0.706	0.724	0.734	0.602
CC	0.103	0.696	0.714	0.705	0.714	0.659
ECC	0.470	0.159	0.235	0.247	0.243	0.007
PS	0.084	0.751	0.769	0.760	0.766	0.717
EPS	0.085	0.751	0.769	0.759	0.765	0.715
Average	0.154	0.627	0.658	0.656	0.662	0.566

Table 5
Emotions data set

Algorithm	H.L↓	ACC↑	F1 Measure↑	Micro-F1↑	Macro-F1↑	Exact Match↑
LP	0.198	0.584	0.687	0.688	0.675	0.351
RAKEL	0.186	0.592	0.706	0.701	0.681	0.341
CC	0.207	0.554	0.655	0.663	0.633	0.310
ECC	0.640	0.268	0.409	0.422	0.416	0.002
PS	0.192	0.599	0.701	0.704	0.692	0.367
EPS	0.193	0.599	0.703	0.705	0.691	0.366
Average	0.269	0.532	0.643	0.647	0.631	0.289

Table 6 and Table 7 depict the collected data on the large size data sets, using three algorithms that capture the unconditional correlations among labels.

Table 6
TMC2007 data set

Algorithm	H.L↓	ACC↑	F1 Measure↑	Micro-F1↑	Macro-F1↑	Exact Match↑
ECC	0.068	0.517	0.496	NG	NG	0.767
EPS	0.069	0.549	0.573	NG	NG	0.740
RAKEL	0.068	0.549	0.577	NG	NG	0.744
Average	0.068	0.538	0.548	NG	NG	0.750

Table 7
Ohsumed data set

Algorithm	H.L↓	ACC↑	F1↑ Measure	Micro F1↑	Macro F1↑	EM↑ ↑Match
ECC	0.063	0.426	0.414	NG	NG	0.784
EPS	0.074	0.424	0.366	NG	NG	0.797
RAKEL	0.075	0.383	0.392	NG	NG	0.830
Average	0.070	0.411	0.390	NG	NG	0.803

C. Conditional Correlations Algorithms Data Collection

Table 8 to Table 10 describe the data that has been collected on the three regular size data sets, that have been tested against several common conditional correlations-based algorithms.

Table 8
Yeast data set

Algorithm	H.L↓	ACC↑	F1 Measure↑	Micro-F1↑	Macro-F1↑	Exact Match↑
CLR	0.226	0.514	0.405	NG	NG	0.970
BPMLL	0.322	0.185	0.210	0.202	0.459	0.185
LEAD	0.202	NG	NG	NG	NG	NG
Rank-SVM	NG	NG	NG	0.587	0.387	0.161
CDN-LR	NG	NG	NG	**0.640**	0.438	0.174
CDN-SVM	NG	NG	NG	0.638	0.357	0.164
Average	0.25	0.3495	0.3075	0.51675	0.41025	0.3308

Table 9
Scene data set

Algorithm	H.L↓	ACC↑	F1 Measure↑	Micro-F1↑	Macro-F1↑	Exact Match↑
CLR	0.101	0.695	0.405	NG	NG	0.391
BPMLL	0.579	0.212	0.663	0.233	0.219	0.212
LEAD	0.098	NG	NG	NG	NG	NG
Average	0.259	0.453	0.534	0.233	0.219	0.301

Table 10
Emotions data set

Algorithm	H.L↓	ACC↑	F1 Measure↑	Micro-F1↑	Macro-F1↑	Exact Match↑
CLR	0.214	0.557	0.668	NG	NG	0.740
BPMLL	0.433	0.276	0.389	0.381	0.426	0.276
LEAD	0.197	NG	NG	NG	NG	NG
Rank-SVM	NG	NG	NG	0.651	0.566	0.225
CDN-LR	NG	NG	NG	0.629	0.615	0.225
CDN-SVM	NG	NG	NG	0.654	0.641	0.241
Average	0.281	0.416	0.528	0.578	0.562	0.341

Table 11 and Table 12 depict the collected data on the large size data sets, using two algorithms that captured conditional correlations among labels

Table 11
TMC2007 data set

Algorithm	H.L↓	ACC↑	F1 Measure↑	Micro-F1↑	Macro-F1↑	Exact Match↑
CLR	0.078	0.506	0.577	NG	NG	0.853
IBLR	0.077	0.479	0.434	NG	NG	0.797
Average	0.077	0.492	0.505	NG	NG	0.825

Table 12
Ohsumed data set

Algorithm	H.L↓	ACC↑	F1 Measure↑	Micro-F1↑	Macro-F1↑	Exact Match↑
CLR	0.088	0.374	0.407	NG	NG	0.914
IBLR	0.097	0.230	0.127	NG	NG	0.937
Average	0.092	0.302	0.267	NG	NG	0.925

D. Data Analysis

After collecting the data, the collected data is aggregated and summarized. The data has been grouped according to the data set, and the correlations capturing strategy that has been used in the algorithm. Then, the results of each metric are averaged over all the algorithms that have been used. Table 13 depicts the final summarization of the collected data, where "C" represents the conditional correlations strategy, and "U" represents the unconditional correlations strategy.

Table 13 clearly shows that the unconditional correlation capturing strategy overcomes the conditional correlations capturing strategy on most evaluation metrics using all data sets. The only metric that is preferred by conditional correlations capturing strategy is the Exact Match metric, since conditional correlations strategy wins 4 times and unconditional correlations strategy wins only 1 time. In fact, this is rational, since most conditional correlations capturing strategies are designed to optimize the Exact Match metric, while unconditional correlations capturing strategies are usually designed to optimize the Hamming Loss metric.

Table 13
Data final summarization

Size	Data Set	Metric	H.L↓	Acc↑	F1↑	Micro↑	Macro↑	EM↑
Regular	Yeast	U	0.276	0.479	0.607	0.606	0.422	0.182
		C	0.25	0.349	0.307	0.516	0.410	0.330
		Winner	C	U	U	U	U	C
	Scene	U	0.154	0.627	0.658	0.656	0.662	0.566
		C	0.259	0.435	0.534	0.233	0.219	0.301
		Winner	U	U	U	U	U	U
	Emotions	U	0.269	0.532	0.643	0.647	0.631	0.289
		C	0.281	0.416	0.528	0.578	0.562	0.341
		Winner	U	U	U	U	U	C
Large	TMC2007	U	0.068	0.538	0.548	NG	NG	0.750
		C	0.077	0.492	0.505	NG	NG	0.825
		Winner	U	U	U	NG	NG	C
	Ohsumed	U	0.070	0.411	0.390	NG	NG	0.803
		C	0.092	0.302	0.267	NG	NG	0.925
		Winner	U	U	U	NG	NG	C

Table 14 summarizes the total number of wins for both conditional and unconditional correlations capturing strategies. Regarding Hamming Loss metric, unconditional correlations strategy wins 4 times, while conditional correlations strategy wins only 1 time. For Accuracy and F1-measure, it is clearly noted that, the unconditional correlations strategy totally overcomes the conditional correlations strategy with a number of wins equals to 5 for the unconditional correlations strategy and zero for the conditional correlations strategy. Same situation for the Micro-F1 and Macro-F1 metrics. For the Exact Match metric, it is clearly seen that, the conditional strategy overcomes the unconditional strategy.

Table 14
Number of wins comparison

Metric	H.L↓	Acc↑	F1↑	Micro↑	Macro↑	EM↑
U	4	5	5	3	3	1
C	1	0	0	0	0	4

Also, it is obviously clear that, the size of the data sets does not affect the previous results. On all cases and whatever the size of the data set is, the unconditional correlations strategy always shows a better predictive performance than the conditional correlations strategy on all evaluation metrics with only one exception, that is the Exact Match metric. Table 15 shows a comparison between the number of wins for both strategies with respect to the size of the data sets.

Table 15
Data set-size based comparison

Size	Regular	Large
U	15	6
C	3	2

Also, it is very important to determine the best strategy to use with respect to the data set which is being used. Table 16 depicts a brief comparison between the unconditional correlations capturing strategy and the conditional correlations capturing strategy, with respect to the data set being used. It can be clearly noted that, unconditional correlations strategy is preferred to be used with all data sets regardless to the size and the characteristics of the data set.

Table 16
Comparison based on the data sets being used

Data Set	Yeast	Scene	Emotions	TMC2007	Ohsumed
U	4	6	5	3	3
C	2	0	1	1	1

E. Findings

This section summarizes those main findings of the data analysis process. The following are the main findings:

1. In general, the unconditional correlations capturing strategy shows a better predictive performance over the conditional correlations capturing strategy.
2. When optimizing the Exact Match metric, it is better to use the conditional correlations capturing strategy otherwise, the unconditional correlations capturing strategy is preferred.
3. The size of the data sets does not affect the truth of preferring unconditional correlations capturing strategy over conditional correlations capturing strategy. In fact, as the size of the data sets get larger, it is more preferred to use unconditional correlations capturing strategy but, this could be an assumption that needs concrete proof.

IV. CONCLUSION AND FUTURE WORK

In MLC, two main strategies are being used to capture the correlations among labels: conditional and unconditional correlations capturing strategies. This paper investigates which of these two strategies is more preferred, with respect to different evaluation metrics and the size of the data sets. In general, unconditional correlations strategy shows a better predictive performance than conditional correlations strategy on most evaluation metrics, regardless of the size of the data set. As a future work, much more researches should be conducted to determine the characteristics of the data sets that make capturing and exploiting the correlations among labels of great benefit.

REFERENCES

[1] S. Bengio, J. Weston, and D. Grangier, "Label embedding trees for large multi-class tasks," *in Proc. Advances in Neural Information Processing Systems*, Vancouver, BC, Canada , 2010, pp. 163-171.

[2] K. Dembczynski, W. Waegeman, W. Cheng, and E. Hullermeier, "On label dependence and loss minimization in multi-label classification," *Machine Learning*, vol. 88, no. 1, pp. 5-45, 2012.

[3] R. Alazaidah and F. Ahmad, "Trending Challenges in Multi Label Classification," *International Journal of Advanced Computer Science and Applications*, vol. *7*, no. 10, pp. 127-131, 2016.

[4] R. Alazaidah, F. Thabtah, and Q. Al-Radaideh, "A Multi-Label Classification Approach Based on Correlations Among Labels," *International Journal of Advanced Computer Science and Applications*, vol. *6*, no. 2, pp. 52-59, 2015.

[5] E. Gibaja, and S. Ventura, "A Tutorial on Multilabel Learning," *ACM Computing Surveys (CSUR)*, vol. 47, no. 3, Apr., 2015.

[6] G. Tsoumakas, and I. Vlahavas, "Random k-labelsets: An Ensemble Method for Multilabel Classification," in *Proc. 18th European Conference on Machine Learning*, Warsaw, Poland, 2007, pp. 406–417.

[7] L. Rokach, A. Schclar, and E. Itach, "Ensemble methods for multi-label classification," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7507-7523, 2014.

[8] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Machine Learning*, vol. 85, no. 3, pp. 333-359, 2011.

[9] W. Cheng, E. Hüllermeier, and K. J. Dembczynski, "Bayes Optimal Multilabel Classification via Probabilistic Classifier Chains," in *Proc. 27th International Conference on Machine Learning (ICML-10)*, Haifa, 2010, pp. 279–286.

[10] J. Read, "A pruned problem transformation method for multi-label classification," in Proc. *2008 New Zealand Computer Science Research Student Conference (NZCSRS 2008)*, New Zealand, 2008, vol. 143150.

[11] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun, "Large Margin Methods for Structured and Interdependent Output Variables," *Journal of Machine Learning Research*, vol. 6, pp. 1453-1484, 2005.

[12] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification,", in *Proc. Advances in Neural Information Processing Systems*, Vancouver, BC, Canada , 2001, pp. 681-687.

[13] J. Fürnkranz, E. Hüllermeier, E. Loza Mencía, and K. Brinker, "Multilabel classification via calibrated label ranking," *Machine Learning*, vol. 73, no. 2, pp. 133-153, 2008.

[14] M. L. Zhang, and K. Zhang, "Multi-label learning by exploiting label dependency," in *Proc. the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, Washington, DC, 2010, pp. 999–1008.

[15] M. L. Zhang, and Z. H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.

[16] Y. Guo, and S. Gu, "Multi-label Classification using Conditional Dependency Networks," in *Proc. The Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI'11),* Barcelona, Spain, 2011, pp. 1300-1305.

[17] W. Cheng, and E. Hüllermeier, "Combining instance-based learning and logistic regression for multilabel classification,". *Machine Learning*, vol. 76, no. 2, pp. 211-225, 2009.