International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN: 2349-9745

Implementation of MIML Framework using Annotated Image Dataset

Prof. Praveen Bhanodia[1], Prof. Pritesh Jain[2], Keyur Tank[3]

[1]*Asst. Professor & Head (CSE), PCST, Indore, India, pcst.praveen@gmail.com*
[2]*Asst. Professor (CSE), PCST, Indore, India, pritesh.arihant@gmail.com*
[3]*Computer Science & Engineering, PCST, Indore, India, keyur.kt@gmail.com*

Abstract— As MIL (Multi-Instance Learning) considers only input ambiguity and MLL (Multi-Label Learning) consider only output ambiguity, we require a framework which consider both ambiguities together and solve the complex problems. MIML (Multi-Instance Multi-Label) framework can solve this problem, but the implementation of MIML dataset is more complex as it considers multiple labels and its multiple instances both together. This research work focuses on implementation of MIML framework using 2014 annotated natural scene image dataset. An image annotation task is closely related to MIML learning problem. Multi class SVM (MSVMpack) used to handle classification of more than two classes without depending on different decomposition methods. Bag of Regions (BoR) is used as a bag generator which is well known framework to generate local features from images. SIFT Scale-Invariant Feature Transform (SIFT) good descriptor can handle intensity, rotation and scale with variations. During experiment for each image SIFT descriptors are extracted for each shot. As a result it also provide vector of predicted labels, accuracy rate during classification, hamming loss, one-error, coverage and R-loss after testing the model.

Keywords- MIML Classification Framework, Image Annotation, Multi class SVM, SIFT, BoR.

I. INTRODUCTION

A data mining classification is an important task of data mining or machine learning that is used to predict group membership for data instances. Classification separates data into learning (training) and classification (testing) sets. Training set is a dataset that is derived from original set and Testing set is a dataset that will be use to evaluate the performance of classifier or a model. Classification can be a Supervised or Unsupervised learning. In Supervised Classification the set of possible classes is known in advance. In Unsupervised Classification set of possible classes is not known. There are three important classification frameworks are currently available Multi-Label Learning (MLL) [1, 2], Multi-Instance Learning (MIL) [3, 4], and Multi-Instance Multi-Label Learning (MIML) [1] framework. In this paper we focus on a MIML framework implemented in an image annotation process. Figure 10 shows the comparison between MIL, MLL and MIML using image example. An annotation is one type of metadata that can be attached to any video, image, text, audio or other data in the form of explanation, comments, navigation or presentational markup. Annotations are the part of the original data. For example, YouTube video annotations are a new way to add interactive commentary to your videos.

Generally image contains multiple regions as a feature vector, so image annotation task is basically a MIML learning problem. New research in MIML classification framework deals

e-ISSN: 2349-9745

with such a problem and ge nerates annotation and learning methods more smoothly and accurately. It is convenient to implement such a framework on dataset which reduces learning efficiency and consider index ing, browsing and retrieval of annotated image/video from the database because nowadays YouTube, storage devices, networks, compression techniques, lots of images/videos have be en generated and transmit or shared on the internet.

The main problem with MIL and MLL is that both suffers from the input and output ambiguities, respectively. Mu lti-Instance Multi-Label (MIML) is a new framework in data mining classification, where multiple objects are represented by a bag of in stances (input) and the objects are allowed to have multiple labels (output) simultaneously. Th e main goal of this research work is to implement MIML framework that can consider both input and output ambiguities of annotated video dataset together. Furthermore, after imp lementation checks the learning accuracy of train ing and testing model, and then need to fi nd local and global features of annotated dataset u sing different parameters.

II. MIML FRAMEWORK AND IMAGE ANNOTAT ION

For MIL and MLL framewor ks many efficient algorithms proposed by v arious researchers. But with efficient learning alg orithms, efficient learning model are also req uire generating the testing and training set. It means highly efficient learning model effect s learning methods more accurately. MIL framework describe real-world object by a number of instances is associated with one class labe l only. Likewise, in MLL a real-world obje ct described by one instance is associated with a number of class labels. Image is useful resou rce which contains multiple region or sections, so MIML task can solve such a task more acc urately than MIL or MLL. MIML framework is easier in representation and many-to- many mapping of complicated objects in compare of SISL framework as shown in fig. 1 and 2.

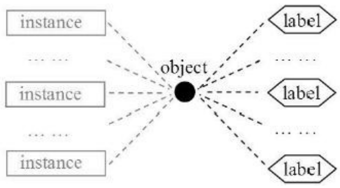


Fig 1: MIML Framew ork

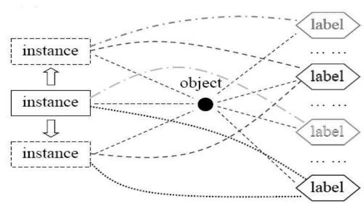


Fig 2: Many-to-man y mapping

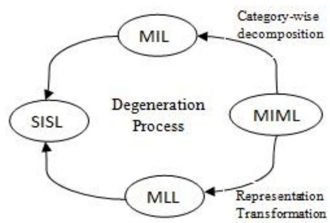


Fig 3: MIML Degeneration Process

There are mainly two important methods to solve MIML problems: (1) Solving MIML Problems by Degeneration. (2) Solving MIML Problems by Regulariz ation. SISL can be considering as a degenerate version of either MIL or MLL which is again a degenerate version of MIML shown in fig. 3. MIML can also deal with ambiguo us data effectively. Image usually contains multiple regions each can be represented by an instance and labels. For example fig. 4 is an image related with jungle of the Africa. It contai ns instances like an Elephant, Lion, Tropic, Afric a, Grassland etc, and labels like animal, forest, grassland etc. All the instances defined left side and all the labels are defined right side. Col oring arrows define that how different instances are depends on different labels in the im age. For example,

e-ISSN: 2349-9745

instance lion can be defined under all the labels animal, forest and grassland. Similarly Africa and Amazon both fall under all the categories like animal, forest and grassland. It solves both input and output ambiguities together without losing any information.

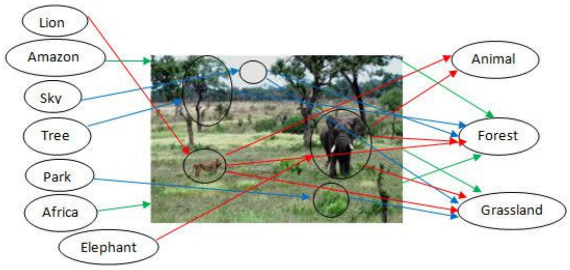


Fig 4: Example of MIML framework in image

An annotation is one type of metadata that can be attached to any video, image, text, audio or other data in the form of explanation, comments, navigation or presentational markup. In this research work image annotation [6] is used for classification. Image annotation captures uploaded image or real-time image which is either 2D or 3D. It also provide image as an overlay or portion of the image. Image annotations should be possible in any web image formats like jpg, png, gif, svg, pdf. There are different image annotation techniques like, (1) Image Annotation Based On Ontology. (2) Making use of Textual Information. (3) Automatic Image Annotation (4) Manual Image Annotation. Fig. 4 shows example of different regions and annotation of image like,

1. Image annotation with navigation and zooming range.
2. Annotation of entire image.
3. Overlays of user-generated graphics and text.
4. Overlays of other images or videos
5. User-defined, custom annotation markers

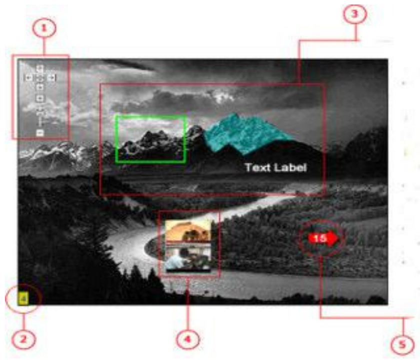


Fig 5: Image Annotation

III. LITERATURE SURVEY

Zhi-Hua Zhou et al. [1] define MIMLBOOST algorithm, it provide independent labels that decompose MIML task into a series of multi-instance learning tasks where all labels will be treat as a task. In the first step of MimlBoost, each MIML example is transformed into a set of number of multi-instance bags, where bag contains number of instances and labels. Zhi-Hua Zhou et al. [1] define, MIMLSVM algorithm which provide spatial distribution of the bags. Each bag provide relevant information for label discrimination which measure distance between each bag and each representative bags identified using clustering methods. M.-L. Zhang et al. [7] proposed MIML-NN algorithm which provide dependencies between

e-ISSN: 2349-9745

different categories during decomposition into multiple set of classification problems using well-known Back-Propagation learning method (BP-MLL). Zhang et al. [8], also provide M3MIML: A Maximum Margin Method for Multi-Instance Multi-Label Learning. This method defines connection between instances and labels. In this method learning task is formulated as a quadratic programming (QP) drawback and implemented in its twin type.

T. Sumathi et al. provide survey on "Automatic Image Annotation and Retrieval using MIML", using different algorithms like MIMLBOOST, MIMLSVM, D-MIMLSVM, InsDif and SubCod algorithms. Cam-Tu Nguyen et al. proposed "Multi-Modal Image Annotation with Multi-Instance Multi-Label LDA (Latent Dirichlet Allocation)". Z. H. Zhou et al. proposed the MIMLBOOST and MIMLSVM algorithms which achieve good performance in an application to scene (image) classification using MIML framework. Ameesh Makadia et al. [9] introduce a new baseline technique for image annotation that treats annotation as a retrieval problem.

IV. IMAGE DATASET

Here MIML image data set consists of 2,014 natural scene images [10], where all the possible six class labels are road, desert, mountains, sea, sunset and trees. Original part contains each image smoothed by a Gaussian filter and a combination of different label set. Processed part contains bags and instances of the scene dataset implemented in MATLAB. Each image is sub-sampled into 9x15 dimensional feature vectors, where region of each defined as a blob 2x2 matrix with a four neighbors.

V. EXPERIMENTAL SETUP

In image annotation generally we can consider each image as a bag and the subparts of the images as instances of the same image. Each image is a collection of regions and then each region consider as an instance. In this experiment, each image considered as a bag and regions of image as an instance. During experiment 2000 bags were used.

A. SIFT descriptors

A good descriptor can handle intensity, rotation and scale with variations. During experiment for each image Scale-Invariant Feature Transform (SIFT) [11] used as a detection and descriptor. The SIFT detector extracts features, frames or key-points from the image. The SIFT descriptors provide regions, edges or appearance of image which are rough in texture. It describes and detects local features in frames or images. After extraction global/local features from each image, bags of instances will be create. Andrea Vedaldi [12] provide library to implement SIFT in MATLAB interface. This package provide different SIFT function to detect and descript features from the image. SIFT returns descriptors in 128 x K matrix format and returns frames in 4 x K matrix format.

B. Performance measures

Many different measures are available for evaluating the performance of classification model to retrieve information like, predicted labels, accuracy rate during classification, hamming loss, one-error, coverage, ranking loss and correlation coefficient after testing the model.

C. Multi-class SVM package (MSVMpack)

Support Vector Machine has been used in different real-world problems like text or hypertext categorization, classification of image or video, bioinformatics, hand-written character recognition etc. It is sensitive to noise and generally use binary classification (two classes).

To solve this problem for multiclass, combination of different binary classifiers is a better approach. Multi-class SVM package (MSVMpack) [13] is an open source software package used for multi class SVM. It handles classification of more than two classes without depending on different decomposition methods. This package provides a framework to implement all the different multi class SVM models and algorithms. It can implement M-SVMs defined by Weston and Watkins (1998), Crammer and Singer (2001), Lee et al. (2004), and the M-SVM2 of Guermeur and Monfrini (2011). It supports Linear, Polynomial and Gaussian RBF kernel functions. It provide features like k-fold cross validation, data normalization, MATLAB interface, API to use in other programs, handle multiple data formats, parallel implementation etc. This package used as a classifier and to construct MIML model using MATLAB define in phase-4 in fig 6.

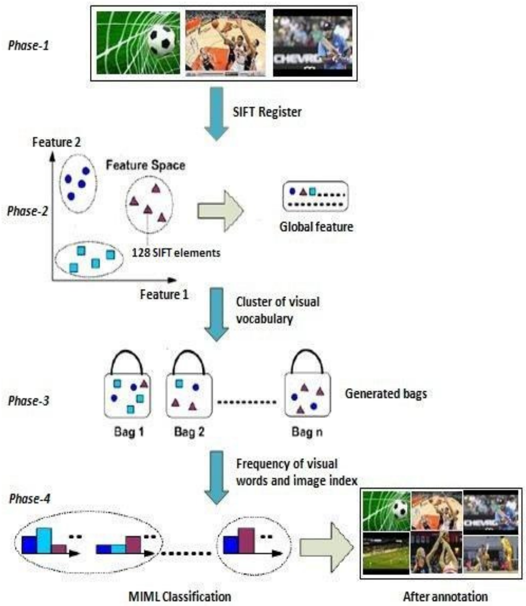


Fig 6: Generation of bags and Classification

D. Generation of bags and Classification

Generation and classification of bags is shown in fig. 6. In this research work, MIL framework is used to generate positive and negative bags and instance from given MIML dataset. The bag is labeled as a positive bag (interested image) or positive instance (interested region) if it contains at least one positive instance. If not then it is labeled as a negative bag (non-interested image) or negative instance (non-interested region).

Bag of Region (BoR) [14] is another recently proposed approach which uses multi-level pooled representation. It creates overlapped order less region parts. In this approach the regions are not concatenated into an image descriptor but taken as a collection of regions known as bags. The main advantage of BoR is that it covers a larger variance in scale, translation, rotation of frames, viewpoint, and make clear by enlarging the training pool. The disadvantage of the BoR method is that it increases the number of classification operations and the size of the training pool. This approach requires a number of classification operations increased with an order equal to the total number of regions from BoR. In this research BoR method used to represent frame or image and then the classification score is computed for each class and each region.

e-ISSN: 2349-9745

VI. CONCLUSION

This research work proves successfully implementation of MIML framework using annotated image dataset. MIML framework can efficiently solve input and output ambiguities together. Multi-class SVM is efficient classifier and implements all the different multi class SVM models and algorithms. Bag of region generate bag of instances and SIFT used as a descriptor for dimensionality reduction process and generate accurate feature vectors from the images. MATLAB is efficient tool to solve MIML problems and handle MIL and MLL framework together. LIBSVM support classification process more effectively and Radial Basis Function (RBF) or Gaussian kernel provide similarity between feature vectors more accurately. Finally MIML framework/classification can solve complex problems related image retrieval more accurately and achieve highest classification accuracy of 90%.

Future research challenges in image annotation and MIML framework improve implementation and development of new framework for other datasets like 2D or 3D and spatial datasets. Explore other efficient learning algorithms for MIML framework. Efficient methods for generation of bags and combination of mix dataset.

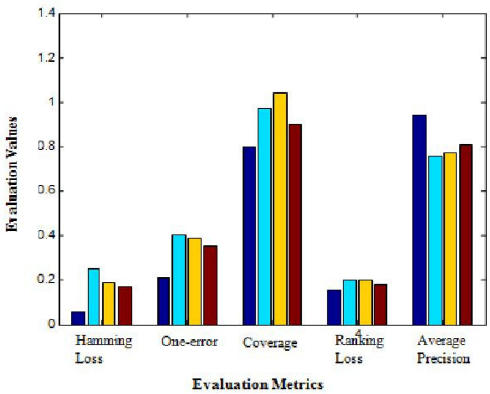


Fig 7: Comparison of evaluation using labels

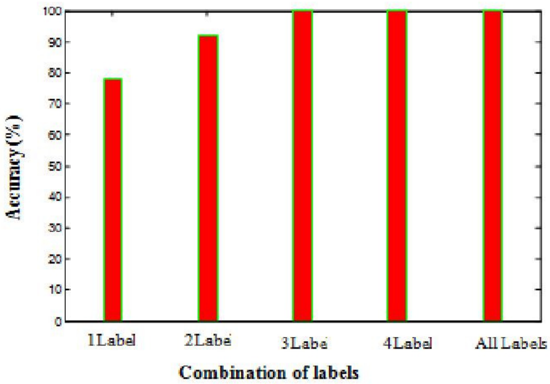
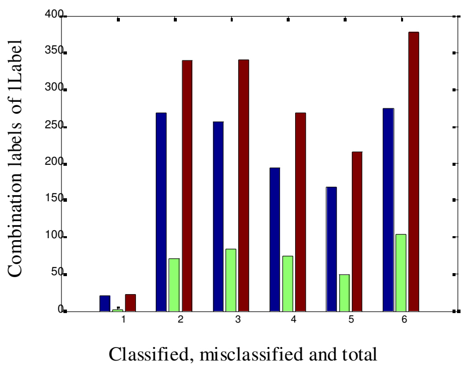
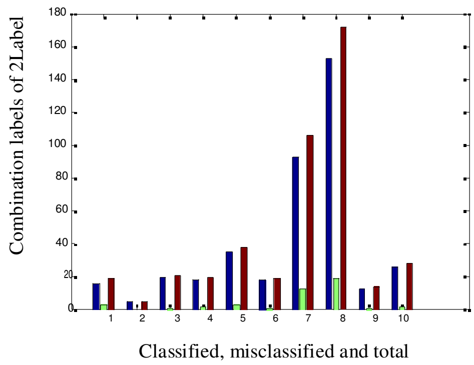


Fig 8: Comparison of accuracy of labels

Finally, figure 5 shows workflow of feature extraction and classification. Figure 7 shows comparison of different metric evaluation. Figure 8 shows comparison of accuracy between combinations of different labels. Figure 9 shows comparison between classified, misclassified and total labels.





e-ISSN: 2349-9745

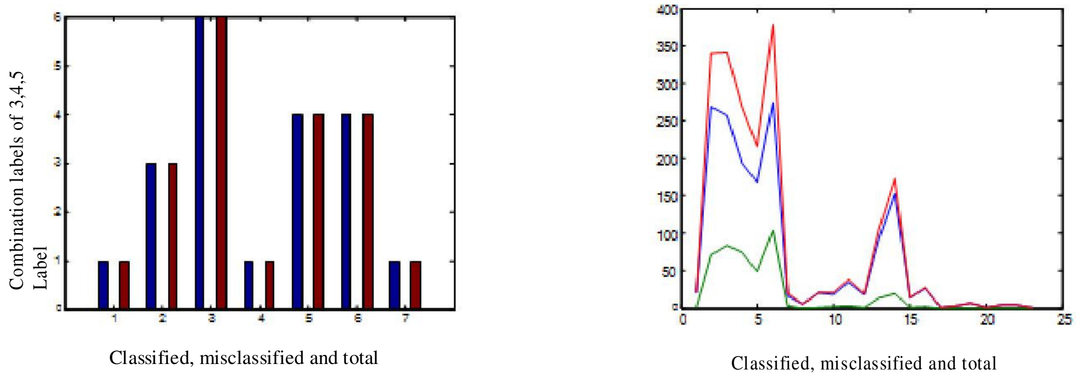


Fig 9: Comparisons of Classified, Misclassified and Total All Labels

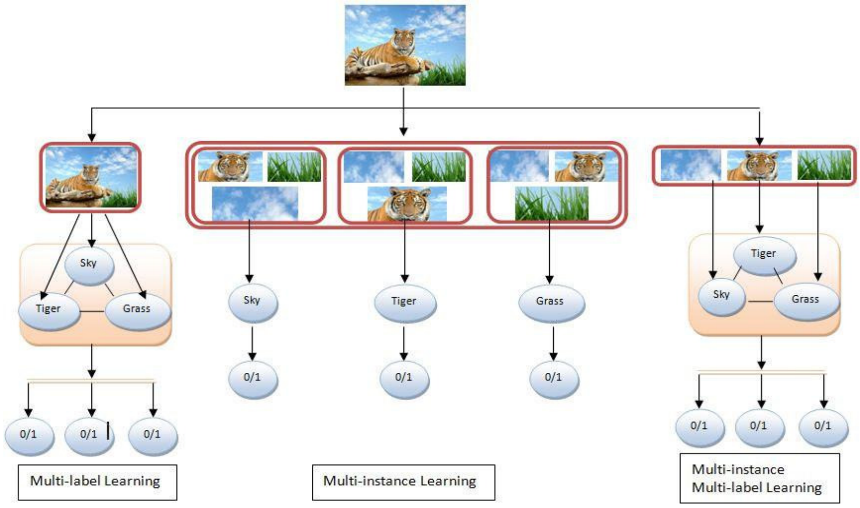


Fig 10: Example or Comparison of MLL, MIL, MIML framework using image classification.

Image label	Compared Methods on Scene			
	Proposed Model*	MIMLBOOST	MIMLSVM	Diverse Density**
Road	**0.90**±0.02	0.85±0.02	0.81±0.01	0.77±0.02
Desert	**0.87±0.03**	**0.87±0.01**	0.87±0.03	0.77±0.04
Mountains	**0.82**±0.03	0.80±0.02	0.82±0.02	0.72±0.03
Sea	**0.76**±0.03	0.73±0.03	0.73±0.03	0.59±0.04
Sunset	**0.87**±0.02	0.86±0.03	0.85±0.02	0.84±0.04
Trees	**0.85**±0.03	0.80±0.02	0.80±0.02	0.78±0.03
Overall Accuracy	**0.838**±0.028	0.818±0.022	0.813±0.021	0.745±0.033

*Accuracy with cross validation (10 fold CV) **MIL algorithm

Table 1: Comparison of accuracy - MIMLBOOST, MIMLSVM, Diverse Density

e-ISSN: 2349-9745

Evaluation Metric	Compared Methods on Scene			
	Proposed Model	MIMLBOOST*	MIMLSVM**	MIMLRBF
Hamming loss ↓	**0.06±0.01**	0.19±0.01	0.18±0.01	0.17±0.02
One-error ↓	**0.21±0.02**	0.40±0.01	0.33±0.02	0.35±0.02
Coverage ↓	**0.80±0.03**	0.97±0.02	1.04±0.02	0.90±0.03
Ranking loss ↓	**0.15±0.02**	0.20±0.03	0.19±0.01	0.18±0.02
Average precision ↑	**0.94±0.02**	0.76±0.02	0.78±0.02	0.81±0.03

* 25 boosting rounds, ** Gaussian kernel 0.2

Table 2: Evaluation on the scene data set (Arrow define lower-best and upper-best)

REFERENCES

[1] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. “Mining Multi-label Data”. O. Maimon, L. Rokach (Ed.), Springer, 2nd edition, 2010.

[2] Mohammad S Sorower, “A Literature Survey on Algorithms for Multi-label learning”, Computer Technologies and Information Sciences (CTIS), Oregon State University, OR 97330.Dec 2010.

[3] Jun Yang, “Review of Multi-Instance Learning and Its applications”, Int’l Conf. on Multimedia and Expo (ICME), Carnegie Mellon University (CMU), 2008.

[4] Oded Maron, Tomás Lozano-Pérez, “A framework for multiple-instance learning”, Proc. of the 1997 Conf. on Advances in Neural Information Processing Systems 10, p.570-576, 1998.

[5] Zhi-Hua Zhou, Min-Ling Zhang, Sheng-Jun Huang, Yu-Feng Li, Multi-Instance Multi-Label Learning. Nanjing University, Nanjing 210046, China. 2011.

[6] Reena Pagare and Anita Shinde. A Study on Image Annotation Techniques. In IJCA, 0975 – 8887, 2012.

[7] M.-L. Zhang and Z.-H. Zhou, “Multi-label neural networks with applications to functional genomics and text categorization,” IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 10, pp. 1338–1351, 2006.

[8] Zhang and Zhou, “M3MIML: A Maximum Margin Method for Multi-Instance Multi-Label Learning”, ICDM’08, 8th IEEE International conference on Data Mining, 2008, pp. 688-697.

[9] Ameesh Makadia, Vladimir Pavlovic, and Sanjiv Kumar. “A New Baseline for Image Annotation”, International Journal of Computer Vision 90(1): 88-105 (2010).

[10] Zhou, Z.-H., and Zhang, M.-L. 2007. “Multi-instance multi-label learning with application to scene classification”, Proc. Advances in Neural Information Processing Systems 19, 2007, pp 1609-1616.

[11] Lowe, David G. (1999). "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision 2. pp. 1150–1157.

[12] Andrea Vedaldi, “An open implementation of the SIFT detector and descriptor”, UCLA CSD Technical Report 070012, 2007.

[13] Jason Weston, Chris Watkins: Support vector machines for multi-class pattern recognition. ESANN 1999: 219-224

[14] Rémi Vieux, Jenny Benois-Pineau, and Jean-Philippe Domenger. 2012. Content based image retrieval using bag-of-regions. In Proceedings of the 18th international conference on Advances in Multimedia Modeling (MMM'12)

[15] Min-Ling Zhang and Zhi-Jian Wang, “MimlRbf: RBF Neural Networks for Multi-Instance Multi-Label Learning”, Neurocomputing, vol 72 (16-18): pp 3951-3956 (2009).