ENSEMBLE-BASED INSTANCE RELEVANCE ESTIMATION IN MULTIPLE-INSTANCE LEARNING

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

10-07-2021 / 14-07-2021

CITATION

Waqas, Muhammad; Qureshi, Rizwan (2021): ENSEMBLE-BASED INSTANCE RELEVANCE ESTIMATION IN MULTIPLE-INSTANCE LEARNING. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.14945712.v1>

DOI

[10.36227/techrxiv.14945712.v1](https://doi.org/10.36227/techrxiv.14945712.v1)

ENSEMBLE-BASED INSTANCE RELEVANCE ESTIMATION IN MULTIPLE-INSTANCE LEARNING

Muhammd Waqas^{*1}, Muhammad Atif Tahir¹, Rizwan Qureshi²

¹National University of Computer and Emerging Sciences, Karachi Campus, Pakistan

²Department of Electronic Engineering, City University of Hong Kong, Hong Kong

Email: {^{*}waqas.sheikh, atif.tahir}@nu.edu.pk, engr.rizwanqureshi786@gmail.com

ABSTRACT

The objective of Multiple-instance learning (MIL) is to learn a mapping function from weakly labeled training data, the training data in MIL is arranged in the form of labeled bags, and every bag holds several instances. The label of the bag depends upon the characteristics of unlabeled instances. This formulation has been used in decision-making applications, such as medical image classification and molecular activity prediction. This data formulation leads to a complex hypothesis, and many existing MIL algorithms are not robust to complex hypothesis space. To deal with this limitation, this paper proposes a Fisher vector-based stacking ensemble design with an instance relevance estimation process, called relevance-based multiple-instance Fisher vector encoding (RMI-FV). The ensemble design builds on top of the instance relevance estimation mechanism. The instance relevancy calculation process employs a Gaussian mixture-based subspace clustering approach, which helps to identify instances with higher relevance to the bag label. The experiments show that the proposed RMI-FV achieves better performance than state-of-the-art MIL approaches.

Index Terms— Multiple instance learning, Instance relevance, Instance selection.

1. INTRODUCTION

The supervised learning techniques are used to develop predictive algorithms that require labeled data instances [1]. In general, these methods have proven to be effective for several machine learning problems, such as regression, classification, and segmentation. However, a high labeling cost makes it difficult to mark all the instances in diverse and challenging situations. In this case, multiple-instance learning (MIL) helps to train the model with restricted labels [2]. The data in MIL is composed of labeled bags

$D = \{(B_1, Y_{B1}), (B_2, Y_{B2}), (B_3, Y_{B3}) \dots (B_n, Y_{Bn})\}$, each $B_j = \{\mathbf{x}_{1,j}, \mathbf{x}_{2,j} \dots \mathbf{x}_{n,j}\}$ is the collection of instance, where $\mathbf{x}_{i,j}$ is the i th instance vector of the j th bag, and n represents the number of instances in the bag. Each instance $\mathbf{x}_{i,j}$ is associated with a label $y_{i,j} \in \{-1, +1\}$. A bag is labeled

as positive if has at least one positive instance, otherwise the bag is associated with a negative label [3]. Additionally, the total number of instances in all bags is denoted by ni . This paper focuses on binary classification problems in MIL.

The concept of MIL was initially introduced for molecular activity prediction [3]. Soon after this, the use of MIL is extended to numerous other applications, including face and action recognition, video classification, and image analysis [4, 5, 6, 7]. In MIL, the bag representation leads to complex hypothesis space, especially in the image classification problem. In image classification, the image patches are extracted and treated as a set of instances, and the instances from different classes may contain similar information. This problem is known as instance label ambiguity [8, 9]. In this case, the relevance of an instance to the bag label becomes essential in the classification process. The example of label ambiguity is illustrated in Figure 1.

Fig. 1. The label ambiguity problem, where the instance of different bags share similar information.

Early MIL approaches such as support vector-based MIL algorithms (mi-SVM and MI-SVM) [10], booting-based tech-

niques (MI-boosting) [11], Multi-instance graph(mi-graph) [12] are not robust to handle complex hypothesis space encountered in MIL [13].

The instance selection approaches are used to tackle complex hypothesis space, such as constructive clustering ensemble (CCE) [14]; this algorithm uses clustering to obtain a binary vector, where each bit in the binary vector represents a bag link to the cluster. The instance selection approach based on random subspaces (RSIS) is proposed in [15], where a representative instance from the positive bag is selected based on the clustering process. However, the performance of RSIS and CCE is comparatively low. Additionally, there is a class imbalance issue encountered in the instance selection process of RSIS.

The multi-point-based embedding method in [16] describes the instance selection process as a target concept detection. The algorithm selects instances based on the relevance measure and maps selected instances to the embedding. the algorithm asserts essential concepts in all bags. This assertion is not always correct, particularly in image classification, where a limited number of instances may contain relevant information, and this information is not available in other images.

To manage large hypothesis-space, MIL problems can be transformed to bag-to-vector encoding mechanism, such as, multi-instance Fisher vector (miFV) and multi-instance locally aggregated descriptors (miVLAD) [17], however, there is no mechanism of score calculation to eliminate the irrelevant instances within the bags, since, incorporating irrelevant instances in encoding generation process degrades the overall ability of the classification algorithms.

This paper proposes a probabilistic approach, for instance relevance estimation, based on subspace Gaussian mixture models (GMM) to overcome the above-discussed challenges. Next, we couple this relevance score in the instance selection process to design an ensemble-based Fisher vector (FV) encoding scheme, collectively called relevance-based multi-instance fisher vector(RMI-FV). The proposed work is motivated by miFV [17] and RSIS [15]. The main contributions of this paper are:

- The subspace-based GMM approach for probabilistic instance relevance calculation is developed.
- The Relevance guided instance selection process is presented.
- The combination of stacking ensemble design and instance selection process with Fisher vector encoding process is introduced.

The proposed RMI-FV is compared with the state-of-the-art MIL approaches including, multi-instance neural networks (mi-Net, MI-Net) [18] and attention based MIL [19]. The experiments show that the performance of the proposed

approach is better than the deep neural network-based approaches, with fewer parameters.

The paper is arranged as follows. The literature review is discussed in Section 2, the details of proposed methodology is given in Section 3, The experimental results and discussion are presented in Section 4 , and Section 5 concludes the paper along with future prospects.

2. LITERATURE REVIEW

The MIL was first used to solve predicting drug activity prediction problem [3], where molecules are viewed as a collection of conformations. Later, MIL-based approaches are also used in object tracking, web mining, and speech localization [8].

Diversity-density-based MIL techniques [20] and boosting approach [11] consider higher witness rates (proportion of positive instances in the bag). However, these algorithms do not perform well if this assumption is not valid. The sparse-MIL approach [21] is designed to handle low witness rates, but, higher witness rate, on the other hand, induces performance degradation. The γ -rule [22] estimates a constant witness rate for all the bags during the training process; however, the witness rate varies among the bags of the same classes [15].

In addition, deep neural networks(NN) are also used to solve MIL problems. The initial NN-based approach for MIL is proposed in [23]. Afterward, the architecture was modified in [18]. A ranking-based loss estimation and network adjustment technique for MIL is proposed in [24]. The attention-based architecture is presented in [19], where the weighted average is computed for some input instances. Moreover, the gate mechanism is also used as an alternate in [19]. In these studies, the posterior probabilities of instance are aggregated using pooling operations to estimate bag probabilities. However, the process of pooling may result in the loss of important information.

3. PROPOSED APPROACH

In this section, the proposed RMI-FV is discussed, the essential part of the proposed algorithm is relevance estimation process, which is discussed in Section 3.1. Given these relevance scores, the instance selection process is initiated, and encodings are generated from selected instances, the process of instance selection and encoding generation process is detailed in Section 3.2 and Section 3.3 respectively. Finally, the ensemble design is detailed in Section 3.4. The proposed RMI-FV algorithm is illustrated in Figure 3.

3.1. Instance Relevance Estimation

First a random subspace $RSS \subset N_f$ is generated, where N_f represents the number of features, and all the training in-

Fig. 2. Parameter analysis, (a) Illustrates the impact of ensemble size, (b) shows the effect of subspace size, (c) depicts influence of number of subspaces, and analysis of temperature is given in (d).

Table 1. The Details of the Datasets.

Dataset	# of Bags	# of positive Bags	# of Negative Bags	# of Instances
Elephant	200	100	100	1391
Fox	200	100	100	1320

stances are mapped to a subspace RSS to generate new set of training instances called D_{RSS} . After that, GMM clustering is applied on D_{RSS} to divide data into K clusters. The clus-

tering process is performed for full and diagonal covariance matrix, and probability scores for instances in corresponding clusters for both matrices are averaged. GMM clustering performs soft assignment of instances to the clusters. Each cluster is used as a generative construct of mean and standard deviation. For the cluster K_i , represented by Gaussian distribution $N(\mu_k, \sigma_k)$, the probability of instance $x_{i,j}$ belonging to K_i is given by:

$$pr(K_i | x_{i,j}) = \frac{pr(x_{i,j} | K_i) * pr(K_i)}{pr(x_{i,j})}, \quad (1)$$

$$p(x_{i,j} | K_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_{i,j} - \mu_k)^2}{2\sigma_k^2}\right), \quad (2)$$

$$p(x_{i,j}) = \sum_k p(x_{i,j} | K_i) * p(K_i), \quad (3)$$

where, $pr(x_{i,j} | K_i)$ is likelihood, $pr(K_i)$ is prior information. Next, each cluster K_i receives the positive score C_i by:

$$C_i = \frac{\sum_{i=1}^{ni} \vartheta(x_{i,j}, K_i)}{\sum_{i=1}^{ni} p(K_i | x_{i,j})}, \quad (4)$$

where

$$\vartheta(x_{i,j}, K_i) = \begin{cases} p(K_i | x_{i,j}) & \text{if } x_{i,j} \in +bag \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Fig. 3. The framework of proposed RMI-FV algorithm, (a) illustrates the instance selection and encoding generation process, (b) shows the ensemble design.

After calculation of positive score for each cluster, the a cumulative posterior probability score $PS_{i,j}$ for $x_{i,j}$ is computed as:

$$PS_{i,j} = \sum_{i=1}^{|K|} C_i * \text{pr}(K_i | \mathbf{x}_{i,j}) \quad (6)$$

The process of score calculation is repeated for numerous times $\mathcal{N}$, where $\mathcal{N}$ is a hyper-parameter. The mean probability score MPS_i for each instance is obtained by:

$$MPS_{i,j} = \frac{\sum_{i=1}^{\mathcal{N}} PS_{i,j}}{\mathcal{N}}, \quad (7)$$

finally, the relevance of instances x_i is computed using the softmax transformation function[15] using equation (8), where T is temperature parameter.

$$Rev_{i,j} = \frac{e^{\frac{MPS_i}{T}}}{\sum_{k=1}^{|B_j|} e^{\frac{MPS_k \in B_j}{T}}} \quad (8)$$

3.2. Instance Selection

The instances from each bag are randomly sampled, the selection process is guided by previously estimated relevance score. The relevance scores work as a sieve procedure in instance selection process. It helps to highlight the likely positive instances in positive bags and difficult instances in negative bags.

3.3. Encoding Process

The FV encoding was first proposed for large scale image retrieval and categorization[25, 26]. The applications of FV encoding in MIL is proposed in [13, 17], where the bag generation process assumed B_i to be modeled by Gaussian Mixture Model (GMM), which works as a probability density function and denoted as $U(;;\theta)$, where $\theta = [\theta_1, \theta_2, \theta_3, \dots, \theta_n]$, denotes the estimated weights, diagonal covariances and mean vectors of a GMM model with K components. The details of assignment process of $\mathbf{x}_{i,j}$ to GMM component is given in Equations 1, 2, 3 respectively.

The FV representation $\mathbf{f}_\theta^{B_i} = [\phi_\omega^1, \phi_\mu^1, \phi_\sigma^1, \dots, \phi_\omega^K, \phi_\mu^K, \phi_\sigma^K]$ for B_i is obtained by:

$$\mathbf{f}_\theta^{B_i} = \sum L_\theta \nabla_\theta \log U_\theta(\mathbf{x}_{i,j}) \forall \mathbf{x}_{i,j} \in B_i. \quad (9)$$

The L_θ is computed as the square root of the inverse of the Fisher information matrix, and the entries $\phi_\omega^{k_i}$, $\phi_\mu^{k_i}$, and $\phi_\sigma^{k_i}$ is determined by:

$$\phi_\omega^{k_i} = \frac{1}{\sqrt{\omega_k}} \sum_{j=1}^{\|B_i\|} (\gamma_j(k) - \omega_k), \quad (10)$$

$$\phi_\mu^{k_i} = \frac{1}{\sqrt{\omega_k}} \sum_{j=1}^{\|B_i\|} \gamma_j(k) \left(\frac{\mathbf{x}_{ij} - \boldsymbol{\mu}_k}{\boldsymbol{\sigma}_k} \right), \quad (11)$$

$$\phi_\sigma^{k_i} = \frac{1}{\sqrt{\omega_k}} \sum_{j=1}^{\|B_i\|} \gamma_j(k) \frac{1}{\sqrt{2}} \left[\frac{(\mathbf{x}_{ij} - \boldsymbol{\mu}_k)^2}{\boldsymbol{\sigma}_k^2} - 1 \right], \quad (12)$$

where $\{\mu_k, \omega_k, \sigma_k\}$ are the estimated mean vectors, weights, and diagonal covariance parameters of a GMM model with K components, respectively. The $\gamma_j(k)$ is the soft assignment of $\mathbf{x}_{ij}$ to component k , the parameters of GMM are estimated on the training bags using maximum likelihood estimation. The detailed discussion of FV encoding for MIL is given in [13, 17].

The size of FV encoding is determined by the number of parameters in U , based on this characteristic, the Fisher encoding scheme transforms the input bag to fix $(2KD + 1)$ length vector, where D is the size of the instance.

3.4. Ensemble Design

We used stacking ensemble to train a set of heterogeneous classifiers. This stacking process is repeated for several iterations, and each base classifier is trained on different FV encodings, The final classifier is trained on the stacked predictions of base classifiers. For base classifiers, we used, KNN with $K = \{1, 3\}$, Random forest and SVM with polynomial and Gaussian kernels. The details of ensemble design is depicted in Figure 3(b).

4. EXPERIMENTAL RESULTS

In this paper, we used Fox and Elephant datasets of COREL images [10]. There are 200 bags in each dataset, which are related to animal images. The images in datasets are considered as bags, and segments of images are treated as instances. Each instance has 230 dimensions, which represent color, texture, and shape details. The performance of the proposed RMI-FV algorithm is evaluated using accuracy. The details of datasets are given in Table 1.

Table 2. Performance Comparison.

Algorithm	Classification Accuracy	
	Elephant	Fox
mi-Graph	0.869 ± 0.078	0.616 ± 0.079
CCE	0.793 ± 0.021	0.599 ± 0.027
miFV	0.852 ± 0.036	0.621 ± 0.050
RSIS	0.846 ± 0.080	0.611 ± 0.018
mi-Net	0.858 ± 0.037	0.613 ± 0.035
MI-Net	0.862 ± 0.034	0.622 ± 0.038
Attention MIL	0.868 ± 0.022	0.615 ± 0.043
Proposed RMI-FV	0.873 ± 0.0715	0.623 ± 0.093

As presented in Table 2, the proposed algorithms outperformed previous algorithms on both datasets. On the Elephant

Table 3. Accuracy values obtained by parameter tuning. The component is set to 1 for tuning PCA, and in components tuning, the PCA is fixed to 0.95.

Tuning PCA Energy		
PCA Energy	Elephant	Tiger
1	0.827	0.623
0.95	0.873	0.581
0.90	0.873	0.550
Tuning of GMM Components		
Number of Components	Elephant	Tiger
1	0.873	0.581
2	0.873	0.570
3	0.872	0.56

dataset, the previous highest reported accuracy was 86.9% by mi-Graph [12], the proposed RMI-FV achieved 87.3% accuracy. Similarly, on the Fox dataset, the proposed RMI-FV obtained 62.3% accuracy as compared to 62.2% accuracy reported by MI-net [18]. The GMM clustering assumes each cluster as a Gaussian component, and log-likelihood estimation accounts for the effect of instance dissimilarity inside each bag. The discriminatory abilities allow the encoding generation process to produce improved vector representation for the bags.

4.1. Discussion

In this subsection, the impact of parameters related to the proposed RMI-FV is analyzed. First, the parameters sensitivity analysis for relevance calculation are discussed, and later, the parameter related to FV encodings are studied. Additionally, the impact of these parameters are analyzed on Elephant dataset.

Figure 2(a) shows that increasing the number of ensemble iterations after ten has no major effect on the results.

The features in the generation of subspace reflect the subspace dimensions. Figure 2(b) reveals that lower values of RSS contribute to better performance.

In Figure 2(c), the impact of generated subspaces $\mathcal{N}$ is illustrated, which indicates that, a small number of subspaces are required.

Figure 2(d) shows that the temperature parameter (T) is a crucial parameter in the softmax function. With lower temperature, similar instances are picked; on the other hand, instances are selected randomly for a higher value of temperature.

In addition, Principal component analysis (PCA) is applied to the instances to reduce the number of the feature before encoding generation. The PCA energy and number of components are two essential parameters in FV encoding scheme for MIL, and these parameters are previously studied in [13, 17]. We used the same settings in our experiments. At

first, the number of components is kept fix to find PCA; afterward, The PCA parameter is fixed to tune the GMM component parameter. The impact of PCA and Number of components for FV encoding generation process is given in Table 3.

5. CONCLUSION

The MIL has achieved great success in domains where abstract structures like molecules, genes, or photographs are involved. However, existing algorithms have shortcomings: many of these are unable to accommodate a vast and complex hypothesis space and unable to find instance dissimilarity inside a bag. To overcome these limitations, this paper proposed a GMM clustering-based approach, for instance relevance estimation, and ensemble design to combine instance selection and FV encoding. The experiments are conducted on Elephant and Tiger dataset. The proposed approach outperformed several benchmark techniques such as miFV, CCE, mi-net, MI-net, and attention-based MIL. A limitation of the proposed approach is a large number of parameters. However, many of these parameters have a lower impact on performance. In the future, we intend to explore the impact of different clustering techniques in the relevance process and instance relationship inside the bag to generate discriminative vector representations.

Acknowledgements

This research work was funded by Higher Education Commission (HEC) Pakistan and Ministry of Planning Development and Reforms under the National Center in Big Data and Cloud Computing.

References

- [1] Richard Szeliski, *Computer vision: algorithms and applications*, Springer Science & Business Media, 2010.
- [2] Veronika Cheplygina, Marleen de Bruijne, and Josien PW Pluim, “Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis,” *Medical image analysis*, vol. 54, pp. 280–296, 2019.
- [3] Thomas G Dietterich, Richard H Lathrop, and Tomás Lozano-Pérez, “Solving the multiple instance problem with axis-parallel rectangles,” *Artificial intelligence*, vol. 89, no. 1-2, pp. 31–71, 1997.
- [4] Xinze Guan, Raviv Raich, and Weng-Keen Wong, “Efficient multi-instance learning for activity recognition from time series data using an auto-regressive hidden

- markov model,” in *International Conference on Machine Learning*. PMLR, 2016, pp. 2330–2339.
- [5] Ivan Gonzalez-Diaz, Jenny Benois-Pineau, Jean-Philippe Domenger, Daniel Cattaert, and Aymar de Rugy, “Perceptually-guided deep neural networks for ego-action prediction: Object grasping,” *Pattern Recognition*, vol. 88, pp. 223–235, 2019.
  - [6] Yixin Chen, Jinbo Bi, and James Ze Wang, “Miles: Multiple-instance learning via embedded instance selection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 1931–1947, 2006.
  - [7] Yixin Chen and James Z Wang, “Image categorization by learning and reasoning with regions,” *The Journal of Machine Learning Research*, vol. 5, pp. 913–939, 2004.
  - [8] Marc-André Carbonneau, Veronika Cheplygina, Eric Granger, and Ghyslain Gagnon, “Multiple instance learning: A survey of problem characteristics and applications,” *Pattern Recognition*, vol. 77, pp. 329–353, 2018.
  - [9] Farrukh Hasan Syed, Muhammad Atif Tahir, Muhammad Rafi, and Mir Danish Shahab, “Feature selection for semi-supervised multi-target regression using genetic algorithm,” *Applied Intelligence*, pp. 1–24, 2021.
  - [10] Stuart Andrews, Ioannis Tsochantaridis, and Thomas Hofmann, “Support vector machines for multiple-instance learning,” in *NIPS*. Citeseer, 2002, vol. 2, pp. 561–568.
  - [11] Cha Zhang, John Platt, and Paul Viola, “Multiple instance boosting for object detection,” *Advances in neural information processing systems*, vol. 18, pp. 1417–1424, 2005.
  - [12] Zhi-Hua Zhou, Yu-Yin Sun, and Yu-Feng Li, “Multi-instance learning by treating instances as non-iid samples,” in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 1249–1256.
  - [13] Xiu-Shen Wei, Jianxin Wu, and Zhi-Hua Zhou, “Scalable multi-instance learning,” in *2014 IEEE International Conference on Data Mining*. IEEE, 2014, pp. 1037–1042.
  - [14] Zhi-Hua Zhou and Min-Ling Zhang, “Solving multi-instance problems with classifier ensemble based on constructive clustering,” *Knowledge and information systems*, vol. 11, no. 2, pp. 155–170, 2007.
  - [15] Marc-André Carbonneau, Eric Granger, Alexandre J Raymond, and Ghyslain Gagnon, “Robust multiple-instance learning ensembles using random subspace instance selection,” *Pattern recognition*, vol. 58, pp. 83–99, 2016.
  - [16] Liming Yuan, Guangping Xu, Lu Zhao, Xianbin Wen, and Haixia Xu, “Multiple-instance learning via multiple-point concept based instance selection,” *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 9, pp. 2113–2126, 2020.
  - [17] Xiu-Shen Wei, Jianxin Wu, and Zhi-Hua Zhou, “Scalable algorithms for multi-instance learning,” *IEEE transactions on neural networks and learning systems*, vol. 28, no. 4, pp. 975–987, 2016.
  - [18] Xinggang Wang, Yongluan Yan, Peng Tang, Xiang Bai, and Wenyu Liu, “Revisiting multiple instance neural networks,” *Pattern Recognition*, vol. 74, pp. 15–24, 2018.
  - [19] Maximilian Ilse, Jakub Tomczak, and Max Welling, “Attention-based deep multiple instance learning,” in *International conference on machine learning*. PMLR, 2018, pp. 2127–2136.
  - [20] Qi Zhang and Sally A Goldman, “Em-dd: An improved multiple-instance learning technique,” in *Advances in neural information processing systems*, 2001, pp. 1073–1080.
  - [21] Razvan C Bunescu and Raymond J Mooney, “Multiple instance learning for sparse positive bags,” in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 105–112.
  - [22] Yan Li, David MJ Tax, Robert PW Duin, and Marco Loog, “Multiple-instance learning as a classifier combining problem,” *Pattern recognition*, vol. 46, no. 3, pp. 865–874, 2013.
  - [23] Jan Ramon and Luc De Raedt, “Multi instance neural networks,” in *Proceedings of the ICML-2000 workshop on attribute-value and relational learning*, 2000, pp. 53–60.
  - [24] Amina Asif et al., “An embarrassingly simple approach to neural multiple instance classification,” *Pattern Recognition Letters*, vol. 128, pp. 474–479, 2019.
  - [25] Florent Perronnin, Jorge Sánchez, and Thomas Mensink, “Improving the fisher kernel for large-scale image classification,” in *European conference on computer vision*. Springer, 2010, pp. 143–156.
  - [26] Florent Perronnin, Yan Liu, Jorge Sánchez, and Hervé Poirier, “Large-scale image retrieval with compressed fisher vectors,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 3384–3391.