Sparse coding with earth mover’s distance

Xuejie Liu\textsuperscript{a}, Jim Jing-Yan Wang\textsuperscript{b}

\textsuperscript{a}College of Computer Science and Technology, Jilin University, Changchun 130012, China
\textsuperscript{b}Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

Abstract

Sparse coding (Sc) has been studied very well as a powerful data representation method. It attempts to represent the feature vector of a data sample by reconstructing it as the sparse linear combination of some basic elements, and a $l_2$ norm distance function is usually used as the loss function for the reconstruction error. In this paper, we investigate using Sc as the representation method within multi-instance learning framework, where a sample is given as a bag of instances, and further represented as a histogram of the quantized instances. We argue that for the data type of histogram, using $l_2$ norm distance is not suitable, and propose to use the earth mover’s distance (EMD) instead of $l_2$ norm distance as a measure of the reconstruction error. By minimizing the EMD between the histogram of a sample and the its reconstruction from some basic histograms, a novel sparse coding method is developed, which is refereed as Sc-EMD. We evaluate its performances as a histogram representation method in tow multi-instance learning problems — abnormal image detection in wireless capsule endoscopy videos, and protein binding site retrieval. The encouraging results demonstrate the advantages of the new method over the traditional method using $l_2$ norm distance.

Keywords: Multi-instance Learning, Histogram Representation, Sparse Coding, Earth Mover’s Distance

1. Introduction

Sparse coding (Sc) has been recently proposed and studied well as an effective data representation method \cite{1, 2, 3}. Given a set of basic elements and a data sample with its feature vector, Sc tries to represent the sample by reconstructing it as the linear combination of these basic elements. The linear combination coefficient vector could be used as the new representation of the sample. To this end, the basic elements and the coefficient vector (also called coding vector)
are learned by minimizing the reconstruction error. At the same time, we also hope that the coding vector could be as sparse as possible. To measure the reconstruction error, a $l_2$ norm based distance is usually applied to compare the original feature vector and its sparse linear combination as a loss function. At the same time, a $l_1$ norm regularization term is also imposed to the coding vector to seek its sparsity. The advantage of using the $l_2$ norm distance to design the loss function and using $l_1$ norm regularization for sparsity pursers lies on that it is easy to optimize and interpret. The feature-sign search method has been proposed to solve the Sc problem by Lee et al. in [4]. And some different Sc versions have also be proposed since then, by adding bias terms to the original Sc loss function based on $l_2$ norm distance [5, 6, 7].

In the multi-instance learning framework, each sample is given as a bag of many instances, instead of one single instance in the traditional machine learning problem [8, 9]. For example, in image classification and retrieval problems, an image could be split into many small image patches, and each patch is an instance. In this case, we usually first learn a set of instance prototypes by clustering the instances of the training samples, and then represent a sample by quantizing its instances into the instance prototypes, and obtain a quantization histogram. The normalized histogram will be used as the feature vector of the sample for further classification or retrieval task. When we try to apply the Sc to represent these histograms, directly using $l_2$ norm distance may not be suitable anymore, and other distance functions which is specially suitable for histogram type is desired. In fact, many distance functions have been studied for histogram comparison, such as Kullback-Leibler divergence [10], $\chi^2$ distance [11, 12], Earth mover’s distance (EMD) [13], etc. Among these distance functions, the EMD metric has been known to quantify the errors in histogram comparison better than other distance metrics.

In this paper, we propose the first sparse coding method with EMD metric as the loss function of the reconstruction error, for the representation of data type of histograms. Instead of using $l_1$ norm distance, we model the sparse coding problem by using the EMD to constructed the loss function. Moreover, we replace the $l_1$ norm regularization term by $l_\infty$ norm for the sparsity of the coding vector so that it could be optimized easily. The newly proposed method, Sc-EMD is especially suitable for the histogram representation within multi-instance learning framework.

This rest parts of this paper continues as follows: The formal objective function of Sc-EMD, and the linear programming-based optimization and algorithm, are presented in Section 2. Experiments with two actual multi-instance learning tasks are discussed in Section 3. Conclusions are given in Section 4.

2. Sparse Coding with Earth Mover’s Distance

In this section, we will introduce the novel sparse coding method using EMD as distance metric instead of the traditional squared $l_2$ norm distance for the representation of histogram data.
2.1. Objective function

Given a data sample, we represent it as a bag of multiple instances under the framework of multi-instance learning [8, 14]. Assuming that we have a training set of \( N \) data samples, to extract the feature vector from the \( n \)-th sample, we quantize the instances of the \( n \)-th sample into a set of instance prototypes, and use the quantization histogram as the feature vector. We assume the number of instance prototypes is \( D \), thus the feature vector of the \( n \)-th sample is a normalized \( D \)-dimensional histogram, denoted as \( x_n = [x_{n1}, \cdots, x_{nD}]^\top \in \mathbb{R}_+^D \), where \( x_{ni} \) is the \( i \)-th bin of the histogram. Note that \( x_n \) is normalized as \( \sum_{i=1}^D x_{ni} = 1 \) so that it is a distribution. The set of the histograms of all the training samples is denoted as \( X = \{x_1, \cdots, x_N\} \), where \( x_n \) is the histogram of the \( n \)-th sample.

In the framework of sparse coding, we try to represent each histogram in \( X \) as a sparse linear combination of a set of basic histograms. We denote the set of basic histograms as \( U = \{u_1, \cdots, u_M\} \), where \( u_m = [u_{m1}, \cdots, u_{mD}]^\top \in \mathbb{R}_+^D \) is the \( m \)-th basic histogram. Similar to \( x_n \), \( u_m \) is also normalized as \( \sum_{i=1}^D h_{mi} = 1 \). The basic histograms are further organized as a basic matrix \( U = [u_1, \cdots, u_M] \in \mathbb{R}_+^{D \times M} \). With the basic histograms, we try to reconstruct each \( x_n \) as the weighted linear combination of these basic histograms as

\[
x_n \approx \sum_m u_m v_{nm} = Uv_n
\]

(1)

where \( v_n = [v_{n1}, \cdots, v_{nM}]^\top \) is the reconstruction coefficient vector, which is also called coding vector of \( x_n \). And \( v_{nm} \) is the coefficient of the \( m \)-th basic histogram for the reconstruction of the histogram of the \( n \)-th sample. Similarly, the sparse coding vectors for all the samples in \( X \) could also be organized as a coding matrix as \( V = [v_1, \cdots, v_N] \in \mathbb{R}_+^{D \times N} \), with the \( n \)-th column as the coding vector of the \( n \)-th sample. Given the histogram \( x_n \) of \( n \)-th sample, the target of sparse coding problem is to learn a basic histogram matrix \( U \) and a sparse coding vector \( v_n \), so that the original \( x_n \) and its reconstruction \( Uv_n \) should be as close to each other as possible to reduce the reconstruction error. At the same time, we also expect the coding vector \( v_n \) should be as sparse as possible. To this end, we discuss the following two issues to build the objective function for the learning of \( U \) and the sparse coding matrix \( V \) for all the training samples’ histograms:

**Reconstruction Error** To measure the reconstruction error between \( x_n \) and \( Uv_n \), traditional sparse coding methods have used the \( l_2 \) norm distance as

\[
l_2(x_n, Uv_n) = \|x_n - Uv_n\|_2^2
\]

(2)
as a distance measure between \( x_n \) and \( Uv_n \), which has been a popular metric for histogram data. To define the EMD between two histograms \( x_n \) and \( Uv_n \), we treat each bin of \( x_n \) as a supply, while each bin of \( Uv_n \) as demand. We also denote \( d_{ij} \) as the ground distance from the \( i \)-th supply to \( j \)-th demand. The EMD between \( x_n \) and \( Uv_n \) is defined as the minimum amount of work needed to fill all the demands with all the supplies,

\[
EMD(x_n, Uv_n) = \min_{F^n} \sum_{i,j} f^n_{ij} d_{ij}
\]

\[
\text{s.t. } f^n_{ij} \geq 0, \sum_j f^n_{ij} \leq x_{ni}, \sum_i f^n_{ij} \leq \sum_m u_{mj} v_{nm}.
\]

(3)

where the variable \( f^n_{ij} \) denotes the amount transported from the \( i \)-th supply to the \( j \)-th demand for the \( n \)-th sample, and \( F^n = [f^n_{ij}]_{D \times D} \) is the matrix of transported amount. The constrain \( f^n_{ij} \geq 0 \) prevents the negative transportation, \( \sum_j f^n_{ij} \leq x_{ni} \) means that the mess moved out from the \( i \)-th supply should not be larger than \( x_{ni} \), while \( \sum_i f^n_{ij} \leq \sum_m u_{mj} v_{nm} \) means that the mess moved into the \( j \)-th demand should not be larger than \( \sum_m u_{mj} v_{nm} \).

**Sparsity Regularization** To encourage the sparsity of each coding vector \( v_n \), traditional sparse coding approaches have been applying the \( l_1 \) norm based sparse penalty to \( v_n \) as

\[
l_1(v_n) = ||v_n||_1 = \sum_{m=1}^{M} |v_{nm}|
\]

Using the \( l_1 \) norm sparsity penalty could impose most of the the elements of \( v_n \) to zeros, and only a few of them will be kept for the reconstruction of \( x_n \). However, the \( l_1 \) norm based term is not smooth and difficult to optimize when it is combined with the EMD objective function. To avoid this problem, we use the \( l_\infty \) norm to regularize the coding vector as

\[
l_\infty(v_n) = \max_{m=1, \cdots, M} |v_{nm}|
\]

(5)

It’s important to notice that the \( l_\infty \) norm does not promote sparsity directly but it could produce the coding vector with all the elements go to zero together and significantly reduces the computational cost.

By applying both the EMD based reconstruction error term in (3) and the \( l_\infty \) regularization term in (5) to each training sample in \( \mathcal{X} \), and summing them up, we have the following objective function for the EMD based sparse coding problem,
\[
\min_{U,V} \sum_n [EMD(x_n, Uv_n) + \gamma l_\infty(v_n)] \\
\text{s.t. } u_{mj} \geq 0, \sum_j u_{mj} = 1.
\]

(6)

where \(\gamma\) is a trade-off parameter, and the constrains \(u_{mj} \geq 0\) and \(\sum_j u_{mj} = 1\) are introduced to the basic histogram to guarantee that the learned basic histograms are normalized distributions. Please note that the \(EMD(x_n, Uv_n)\) itself is also obtained by solving a minimizing problem with regarding to \(F^1, \ldots, F^M\).

We substitute (3) and (5) to (6), so that the optimization problem in (6) is extended into the parameter-enlarged optimization with additional parameter of transported amount matrices as,

\[
\min_{U,V,F^1,\ldots,F^M} \left( \sum_n \sum_{i,j} f_{ij}^n d_{ij} + \gamma \sum_n \max_{m=1,\ldots,M} |v_{nm}| \right) \\
\text{s.t. } u_{mj} \geq 0, \sum_j u_{mj} = 1, \\
f_{ij}^n \geq 0, \sum_j f_{ij}^n \leq x_{ni}, \sum_i f_{ij}^n \leq \sum_m u_{mj} v_{nm}.
\]

(7)

By introducing a slack variable \(\xi_n\) as the upper boundary of \(|v_{nm}|\), so that \(|v_{nm}| \leq \xi_n, m = 1, \ldots, M\), we release the \(l_\infty\) term by minimizing the slack variable \(\xi_n\), instead of finding the maximum element of \(v_n\) first and then minimize it directly. In this way, the optimization problem of (7) is turned into the following one,

\[
\min_{U,V,F^1,\ldots,F^M,\xi} \left( \sum_n \sum_{i,j} f_{ij}^n d_{ij} + \gamma \sum_n \xi_n \right) \\
\text{s.t. } u_{mj} \geq 0, \sum_j u_{mj} = 1, \\
f_{ij}^n \geq 0, \sum_j f_{ij}^n \leq x_{ni}, \sum_i f_{ij}^n \leq \sum_m u_{mj} v_{nm}, \\
- \xi_n \leq v_{nm} \leq \xi_n, \xi_n \geq 0.
\]

(8)

where \(\xi = [\xi_1, \ldots, \xi_M]^T\) is the slack variable vector where \(\xi_n\) is the slack variable for the \(l_\infty\) regularization of \(n\)-th sample’s sparse coding vector.

2.2. Optimization

Directly optimizing the object of (8) is difficult, thus we adopt the alternate optimization strategy for learning of \(U\) and \(V\) in an iterative algorithm. In each iteration, one of \(U\) and \(V\) will be optimized while the other is fixed, and then their role will be switched. The iteration will be repeated until convergence or a maximum iteration number is reached.
2.2.1. Optimizing $V$ while fixing $U$

By fixing $U$, we could optimize the coding vectors together with the other additional variables. Similar to traditional sparse coding methods, we update each sparse coding vector individually. When the $n$-th sample’s coding vector $v_n$ is being optimized, the other ones $v_{n'} (n' \neq n)$ with their corresponding additional variables ($F_{n'}$ and $\xi_{n'}$) are fixed. Thus, the optimization (8) will be turned to

$$\min_{v_n, F_n, \xi_n} \left( \sum_{i,j} f_{ij}^n d_{ij} + \gamma \xi_n \right)$$

$$s.t. \sum_j f_{ij}^n \geq 0, \sum_j f_{ij}^n \leq x_{ni}, \sum_i f_{ij}^n \leq \sum_m u_{mj} v_{nm},$$

$$-\xi_n \leq v_{nm} \leq \xi_n, \xi_n \geq 0.$$  \(9\)

which could be solved as a linear programming (LP) [15, 16, 17]. Please notice that LP solves a problem for a given vector of unknown variables. Here we substitute the vector of variables in $F_n$ (the original variables of the EMD problem) with a longer vector, which contains $F_n$ variables and the entries of $v_n$ and $\xi_n$ together. We did not reformulate the EMD objective, but we changed its constraints to take care of the additional variables related to the NMF. This way the new LP is different from the original EMD, and the result contains both $F_n$, $v_n$ and $\xi_n$ values.

2.2.2. Optimizing $U$ while fixing $V$

By fixing $V$ and its corresponding variables, the optimization problem in (8) could be turned to

$$\min_{U, F_1, \ldots, F_M} \sum_m \sum_{i,j} f_{ij}^n d_{ij}$$

$$s.t. \sum_j u_{mj} = 1, u_{mj} \geq 0,$$

$$f_{ij}^n \geq 0, \sum_j f_{ij}^n \leq x_{ni}, \sum_i f_{ij}^n \leq \sum_m u_{mj} v_{nm}.$$  \(10\)

which could also be solved as a LP problem.

An important limitation of both the optimization problems in (9) and (10) is the large number of additional variables for the LP problem. For each sample $x_n$, a $D \times D$ transported amount matrix $F_n$ is solved in both (9) and (10), thus there are totally $N \times D \times D$ transportation amount variables in the LP problem for the $N$ training samples. When the dimension of the histogram $D$, or the training sample $N$ is large, there would be a large number of variables, which could cause serious computation problem. To overcome this shortage, we reduce the number of variables in $F_n$ by allowing the earth from the $i$-th supply only to its $K$ nearest demands instead of all the $D$ demands. The $K$ nearest demands of $i$-th supply is defined by using the ground distances. In this way, we
reduce the transported mass variables for each supply of each sample from $D$ to $K$, and usually $K \ll D$, thus the total transported amount is reduced from $N \times D \times D$ to $N \times D \times K$.

2.3. Algorithm

We summarize the iterative basic histograms and coding vectors learning algorithm in Algorithm 1. In each iteration, the sparse coding vector for each sample is first learned sequentially, and the basic histograms are then updated based on the learned sparse coding vectors. The iterations will be repeated for $T$ times. When a novel sample comes with its histogram in the test procedure, we simply solve (9) to obtain its sparse coding vector.

**Algorithm 1 ScEMD Algorithm.**

Input: Histograms of training samples $X = \{x_i, \cdots, x_N\}$;
Input: Number of basic histograms $M$;
Initialize the basic histogram matrix $U^0_0 = u_{0,1}, \cdots, u_{0,M}$
for $t = 1, \cdots, T$ do
    for $n = 1, \cdots, N$ do
        Update the sparse coding vector $v_t^n$ for the $n$-th sample by solving (9) while fixing $U^{t-1}$;
    end for
    Update the basic histogram matrix $U^t$ by solving (10) while fixing $V^t$;
end for
Output: The basic histogram matrix $U^T$ and the sparse coding matrix $V^T$.

3. Experiments

In this section, we evaluate the proposed method on two multi-instance learning problems, where the feature vector is a histogram for each sample.

3.1. Experiment I: Abnormal Image Detection in Wireless Capsule Endoscopy Videos

Wireless Capsule Endoscopy (WCE) has been used to detect the mucosal abnormalities in the gastrointestinal tract, including blood, ulcer, polyp, etc [18]. However, usually only a few frames from a large number of WCE video contain abnormalities, thus a would medical clinician spends long time to fine the abnormal frames from a WCE video. In this situation, it is very necessary to develop a system to automatically discriminate abnormal frames from the normal ones. In this experiment, we will evaluate the proposed method as image representation method for the task of abnormal image detection in wireless capsule endoscopy videos.
3.1.1. Dataset and Setup

We construct the dataset for the experiment by collecting 170 images of WCE videos belonging to three abnormal classes and one normal class. The dataset contains 50 normal images, 40 polyp images, 40 ulcer images, and 40 blood images. Given an image of WCE video, the task of abnormal image detection is to classify it to one of the four classes. To this end, each image will be split into many $8 \times 8$ small patches, and each patch is treated as an instance, thus the image will represented as a bag of instances under the framework of multi-instance learning. To extract features from each instance, we calculate the mean, standard deviation, and skewness from the HIS color space of the patch, and the texture features by applying the Gabor filters to the patch. The color and texture features will be concatenated as the feature vector of the instance. Then the instances will be quantized into a pool of instance prototypes and the quantization histogram will be used as the visual feature of the image \[19\]. The histogram will further be represented using the proposed Sc-EMD algorithm as the sparse coding vectors, and the coding vectors will be used to train a support vector machine (SVM) \[20, 21\] to classify the images into one of the four image types.

To conduct the experiment, we employ the 10-fold cross-validation \[22\]. The entire dataset will be split into 10 non-overlapping folds randomly. In each fold, there will be 5 normal images, 4 polyp images, 4 ulcer images, and 4 blood images separately. Each fold will be used as the test set in turns, and the remaining 9 folds will be used as training set. After the images in the training set are represented as histograms under the multi-instance learning framework, we perform the ScEMD algorithm to them and obtain the basic histograms and the sparse coding vectors for the training images. Then we train the SVM classifier using these sparse coding vectors. Based on the basic histograms learned from training histograms, we represent the test images and obtain the sparse coding vectors, and finally input them into the learned SVM classifier to have the final classification results. Please notice that the parameters are turned using only the training set while excluding the test set. To handle the multi-class problem, we have used the one-against-all protocol to train the classifier. A SVM classifier is trained for each class, using the images of this class as positive samples while all other images as negative samples.

The classification results are measured by the recall-precision curve, receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) value for each class.

3.1.2. Results

In the experiments, we compare our Sc-EMD algorithm as a data representation mother to the traditional sparse coding method using the $l_2$ norm distance (denoted as Sc-$l_2$ norm), and also to the original histogram as representation (denoted as Histogram). The recall-precision curves for four different classes are given in Figure \[1\]. In these figures, it is shown clearly that with the proposed Sc-EMD, the classification performance for all four classes is improved significantly, even more so for the last three classes. The performance
improvement is particularly dramatic for the Polyp and Normal classes. Sc-$l_2$ norm could improve the original histogram features somehow, however, due to the reason that it employs the $l_2$ norm distance as loss function, which is not suitable for the histogram data type, the improvement is limited. In particular, Figure 1(a) shows that an increase in classification performance is obtained by Sc-$l_2$ norm against both original histogram and Sc-EMD. The results validate the importance of performing sparse coding with appropriate loss function to the histogram data type.

![Recall-Precision: Blood](image1.png)

(a) Blood

![Recall-Precision: Polyp](image2.png)

(b) Polyp

![Recall-Precision: Ulcer](image3.png)

(c) Ulcer

![Recall-Precision: Normal](image4.png)

(d) Normal

Figure 1: The recall-precession curves of different classes using different histogram representation methods on the WCE images database.

The ROC curves of different classes are shown in Figure 2. Moreover, the AUC values are given in Figure 3. As shown in Figure 2 and Figure 3, our Sc-EMD algorithm clearly outperforms the original histogram feature and Sc-
The advantage is particularly significant on the more challenging Normal class. This result highlights the importance of using the EMD measure for histograms rather than $l_2$ norm distance.

3.2. Experiment II: Protein binding site retrieval

Searching geometrically similar protein binding sites is significantly important to understand the functions of protein and also to drug discovery [23]. Pang et al. [23] presented the protein binding sites as a histogram using the multi-instance learning framework for the protein binding site retrieval problem. In
this experiment, we will evaluate the proposed algorithm for the representation of histogram features of protein binding sites.

3.2.1. Dataset and Setup

In this experiment, we use a protein binding site dataset reported by Pang et al. [23]. In this nonredundant dataset, there are totally 2,819 protein binding sites, belonging to 501 different classes. The number of sites in each class varies from 2 to 58. To conduct the 4-fold cross-validation, we have selected 2,226 binding sites randomly to construct our dataset. The selected dataset contains sites of 249 classes, and the number of sites for each class is from 4 to 58, so that we could guarantee that when the 4-fold cross-validation is performed, in each fold there will be at least one site from every class. The numbers of sites for all the selected classes are shown in Figure 4.

![Figure 4: The number of sites in different classes of the protein binding site dataset.](image)

Given a query binding site and a protein binding site database, the protein binding site retrieval problem is to rank the database sites according to their similarity to the query in a decerning order, so that the database sites belonging to the same class as the query will be ranked at the top of the returned list. To this end, we first represent each binding site as a bag of feature points selected from the binding site surface, and for each point the geometric features are extracted [23]. In the multi-instance learning framework, a binding site is refereed as a bag, and each feature point is refereed as a instance. Then all the feature points are quantized into a set of prototype points, and a histogram is
generated as the bag-level feature of the binding site [23]. Using the proposed Sc-EMD algorithm the histograms are represented as sparse codes for the final ranking. The ranking performances are evaluated by the recall-precision and the ROC curves. AUC values of the ROC curve are also reported as a single performance measure of the ranking results.

3.2.2. Results

The recall-precision and ROC curves of different histogram representation methods are given in Figure 5. From the results in Figure 5 we can find that the Sc-EMD measures performs the best in terms of both recall-precision and ROC curves. It proves that the EMD based method could discover the best distance measure. The AUC values of the ROC curves are also given in Figure 6. These protein binding site retrieval system with Sc-EMD representation method achieves an AUC value of 0.9466, compared to an AUC value of 0.9282 using Sc-$l_2$ norm and 0.9114 using the original histogram.

![Recall-Precision](image1.png) ![ROC](image2.png)

(a) Recall-Precision  (b) ROC

Figure 5: The recall-precision and ROC curves different histogram representation methods on the protein binding site database.

4. Conclusion

A new type of sparse coding method, sparse coding with EMD metric, is proposed in this paper for the representation of histogram data type. The
objective function is composed of an EMD term between the original histogram and the reconstructions result from a pool of basic histograms, and a \( l_\infty \) term for the regularization of the coding vector. The optimization problem is solved as a linear programming problem in an iterative algorithm. Algorithms based on the proposed Sc-EMD outperformed previous \( l_2 \) norm based sparse coding algorithm in two challenging multi-instance learning tasks. In the future, we will apply the proposed method to cyber security and malware detection problems [24, 25, 26, 27, 28, 29, 30].

References

[1] Shengping Zhang, Hongxun Yao, Xin Sun, Xiusheng Lu, Sparse coding based visual tracking: Review and experimental comparison, Pattern Recognition 46 (2013) 1772–88.

[2] Z. Lu, Y. Peng, Image annotation by semantic sparse recoding of visual content, in: ACM Multimedia, 2012, pp. 499–508.

[3] Z. Lu, Y. Peng, Latent semantic learning by efficient sparse coding with hypergraph regularization, in: AAAI, 2011.

[4] H. Lee, A. Battle, R. Raina, A. Ng, Efficient sparse coding algorithms, in: Advances in Neural Information Processing Systems, 2007, pp. 801–808.

[5] S. Gao, I.-H. Tsang, L.-T. Chia, P. Zhao, Local features are not lonely - laplacian sparse coding for image classification, 2010, pp. 3555–3561.

[6] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, D. Cai, Graph Regularized Sparse Coding for Image Representation, IEEE TRANSACTIONS ON IMAGE PROCESSING 20 (5) (2011) 1327–1336. doi:{10.1109/TIP.2010.2090535}

[7] S. Gao, I. W.-H. Tsang, L.-T. Chia, Laplacian Sparse Coding, Hypergraph Laplacian Sparse Coding, and Applications, IEEE TRANSACTIONS ON
[8] X. Kang, D. Li, S. Wang, A multi-instance ensemble learning model based on concept lattice, Knowledge-Based Systems 24 (8) (2011) 1203–1213.

[9] Z. Lu, Y. Peng, Exhaustive and efficient constraint propagation: A graph-based learning approach and its applications, International Journal of Computer Vision 103 (3) (2013) 306–325.

[10] Zhizheng Liang, Youfu Li, ShiXiong Xia, Adaptive weighted learning for linear regression problems via Kullback-Leibler divergence, Pattern Recognition 46 (2013) 1209–19.

[11] J. Tian, D. Yu, B. Yu, S. Ma, A fuzzy TOPSIS model via chi-square test for information source selection, KNOWLEDGE-BASED SYSTEMS 37 (2013) 515–527. doi:{10.1016/j.knosys.2012.09.010}

[12] Z. Lu, An iterative algorithm for entropy regularized likelihood learning on gaussian mixture with automatic model selection, Neurocomputing 69 (2006) 1674–1677.

[13] Y. Lin, Y. Y. Tang, B. Fang, Z. Shang, Y. Huang, S. Wang, A Visual-Attention Model Using Earth Mover’s Distance-Based Saliency Measurement and Nonlinear Feature Combination, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 35 (2) (2013) 314–328. doi:{10.1109/TPAMI.2012.119}

[14] Z. Lu, H. Ip, Spatial markov kernels for image categorization and annotation, IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics 41 (4) (2011) 976–989.

[15] G. R. Amin, A comment on modified big-M method to recognize the infeasibility of linear programming models, KNOWLEDGE-BASED SYSTEMS 23 (3) (2010) 283–284. doi:{10.1016/j.knosys.2010.01.002}

[16] M. Soleimani-damaneh, Modified big-M method to recognize the infeasibility of linear programming models, KNOWLEDGE-BASED SYSTEMS 21 (5) (2008) 377–382. doi:{10.1016/j.knosys.2008.02.004}

[17] Lu Zhang, F. Coenen, P. Leng, Formalising optimal feature weight setting in case based diagnosis as linear programming problems, Knowledge-Based Systems 15 (2002) 391–8.

[18] Sae Hwang, Bag-of-Visual-Words Approach to Abnormal Image Detection in Wireless Capsule Endoscopy Videos, in: G. Bebis, R. Boyle, B. Parvin, D. Koracin, Song Wang, Kim Kyungnam, B. Benes, K. Moreland, C. Borst, S. DiVerdi, Chiang Yi-Jen, Jiang Ming (Eds.), Advances in Visual Computing. Proceedings 7th International Symposium, ISVC 2011, 2011, pp. 320–7, Advances in Visual Computing. 7th International Symposium, ISVC 2011, 26-28 Sept. 2011, Las Vegas, NV, USA.
[19] A. M. Denton, J. Wu, Data mining of vector-item patterns using neighborhood histograms, KNOWLEDGE AND INFORMATION SYSTEMS 21 (2) (2009) 173–199. doi:{10.1007/s10115-009-0201-7}

[20] S. Koeknar-Tezel, L. J. Latecki, Improving SVM classification on imbalanced time series data sets with ghost points, KNOWLEDGE AND INFORMATION SYSTEMS 28 (1) (2011) 1–23. doi:{10.1007/s10115-010-0310-3}

[21] B. X. Wang, N. Japkowicz, Boosting support vector machines for imbalanced data sets, KNOWLEDGE AND INFORMATION SYSTEMS 25 (1) (2010) 1–20. doi:{10.1007/s10115-009-0198-y}

[22] J. Neville, B. Gallagher, T. Eliassi-Rad, T. Wang, Correcting evaluation bias of relational classifiers with network cross validation, KNOWLEDGE AND INFORMATION SYSTEMS 30 (1) (2012) 31–55. doi:{10.1007/s10115-010-0373-1}

[23] B. Pang, N. Zhao, D. Korkin, C.-R. Shyu, Fast protein binding site comparisons using visual words representation, BIOINFORMATICS 28 (10) (2012) 1345–1352. doi:{10.1093/bioinformatics/bts138}

[24] L. Xu, Z. Zhan, S. Xu, K. Ye, Cross-layer detection of malicious websites, in: Proceedings of the third ACM conference on Data and application security and privacy, ACM, 2013, pp. 141–152.

[25] S. Xu, W. Lu, Z. Zhan, A stochastic model of multivirus dynamics, Dependable and Secure Computing, IEEE Transactions on 9 (1) (2012) 30–45.

[26] Z. Zhan, M. Xu, S. Xu, Characterizing honeypot-captured cyber attacks: Statistical framework and case study, Information Forensics and Security, IEEE Transactions on 8 (11) (2013) 1775–1789.

[27] S. Xu, W. Lu, L. Xu, Z. Zhan, Adaptive epidemic dynamics in networks: Thresholds and control, ACM Transactions on Autonomous and Adaptive Systems (TAAS) 8 (4) (2014) 19.

[28] L. Xu, Z. Zhan, S. Xu, K. Ye, An evasion and counter-evasion study in malicious websites detection, in: Communications and Network Security (CNS), 2014 IEEE Conference on, IEEE, 2014, pp. 265–273.

[29] S. Xu, H. Qian, F. Wang, Z. Zhan, E. Bertino, R. Sandhu, Trustworthy information: concepts and mechanisms, in: Web-Age Information Management, Springer, 2010, pp. 398–404.

[30] Z. Zhan, M. Xu, S. Xu, A characterization of cybersecurity posture from network telescope data, in: Proceedings of The 6th International Conference on Trustworthy Systems, InTrust, Vol. 14.
Recall–Precision

- Sc–EMD
- Sc–$l_2$ norm
- Histogram