ACSIJ Advances in Computer Science: an International Journal, Vol. 4, Issue 5, No.17 , September 2015
ISSN : 2322-5157
www.ACSIJ.org

MLENN-KELM: a Prototype Selection Based Kernel Extreme Learning Machine Approach for Large-Scale Automatic Image Annotation

Hamid Kargar-Shooroki[1], Mohammad Ali Zare Chahooki[2*] and Shima Javanmardi [3]

[1] **Electrical and Computer Engineering Department, Yazd University**
Yazd, Iran
hamidkargar@stu.yazd.ac.ir

[2] **Electrical and Computer Engineering Department, Yazd University**
Yazd, Iran
chahooki@yazd.ac.ir

[3] **Electrical and Computer Engineering Department, Yazd University**
Yazd, Iran
sh.javanmardi@stu.yazd.ac.ir

Abstract

With the fast growth of digital images in web, large-scale Automatic Image Annotation (AIA) dealt with some of critical challenges. The most important of them are system scalability and annotation performance. On the other hand, learning methods in the large-scale systems with the large number of training instances cannot correctly perform and deal with memory and learning time restrictions. In this paper in order to solve the performance of large-scale AIA systems, and limitations of employing learning methods in these systems, MLENN-KELM approach has been proposed. In the proposed approach, first, most effective instances are selected from training set by Prototype Selection (PS) methods. The basic assumption of selecting effective instances is reducing the size of training set and solving memory restrictions in large-scale AIA systems and learning methods. Then, annotation process is done by Kernel Extreme Learning Machine algorithm (KELM). The main advantage of using KELM algorithm is the improvement of annotation performance than other learning methods. Experimental results on NUS-WIDE-Object image set demonstrate the good performance of proposed approach in solving large-scale AIA challenges and also capability improvement of KELM algorithm in large-scale applications.

Keywords: *Automatic Image Annotation, Kernel Extreme Learning Machine, Large-Scale Learning Context, Prototype Selection.*

1. Introduction

AIA is the process of automatic assignment of text descriptions into digital images based on their visual contents [1]. This process is somewhat similar to object recognition. In AIA, the position of objects is not important while in object recognition it is substantial [2]. AIA as a pre-processing step can improve performance of text-based image retrieval (TBIR) systems. By using advantages of content-based image retrieval (CBIR) and improving retrieval accuracy, the process of searching a query image converts to text matching and retrieval speed will be increased [3].

With increasing the number of images and conceptual annotation words, AIA systems deal with some of major challenges. Some of main challenges are system scalability or memory restrictions and annotation performance [4]. Since AIA systems in real environments should be online with high performance results, challenge solving is critical. In recent years, extensive researches have been done to address mentioned challenges. In [3], [5] and [6] scalability challenge and in [2] and [7] large-scale image annotation performance is addressed.

Because AIA is a multi-label classification problem, to predict multiple labels for each query image, label prediction and learning methods must be employed [8]. On the other hand, with the large number of images and annotation words, some of learning methods like Lazy, SVM and Neural Networks deal with learning time and memory limitations. So performance of large-scale AIA systems is extremely decreased due to the large number of instances.

As previously stated, the main challenge of learning methods especially for their kernel versions in large-scale learning context is memory limitations in constructing kernel matrix and some of its operations. These limitations for some kernel learning methods like kernel SVM and KELM is pretty serious that these methods cannot be used in large-scale learning context [9]. In recent years, numerous researches have been done to overcome the limitations of using kernel methods [10–13]. A set of approaches try to reduce training set size and construct a subset with lower instances. Proposed approaches in [9], [12] and [13] select some of instances from original set in a random way to construct reduced subset. The main disadvantage of random selection is the lack of attention to effective instances and tacking into account same value for instances. In subset construction procedure, selecting effective instances is important and lead to performance improvement.

In this paper, in order to solve above mentioned challenges of large-scale AIA systems and also memory restrictions of kernel methods in large-scale learning context, **M**ulti-**L**abel **E**dited **N**earest **N**eighbor based **K**ernel **E**xtreme **L**earning **M**achine (MLENN-KELM) has been proposed.

In the proposed MLENN-KELM approach, at firstly and to scalable up AIA systems, large-scale training set is pre-processed. Pre-processing step means using PS approaches and selecting effective instances and removing noisy and implicitly duplicated ones. AIA is a multi-label classification problem, while most of PS methods are based on single label instances [14,15]. So in the proposed approach, MLENN algorithm [16] has been employed. This method is proposed for balancing multi-label datasets and is based on ENN [17] that is one of the famous PS algorithms.

To overcome annotation accuracy challenge and improving quality of predicted labels, kernel version of ELM has been employed [18]. ELM is new extension of Back Propagation algorithm [19]. ELM is a powerful method and it is led to performance improvement in various applications such as image understanding and video analysis [20–23]. As before stated, because of the large number of instances, this powerful method cannot be used in large-scale AIA. So this research tries to employ KELM in large-scale AIA with combining with a PS step.

The contributions of this paper are listed as follows. (1) Scalability challenge in large-scale AIA is addressed by selecting effective instances from original training set. (2) By employing this idea, memory limitations of some kernel learning methods like KELM in large-scale learning context is resolved. More precisely, KELM can be applied in large-scale applications with this idea. (3) Annotation performance and quality of predicted labels by using MLENN-KELM in large-scale AIA is more improved than other state-of-the-art methods. Fig. 1 shows the overview of proposed MLENN-KELM approach.

The remainder of this paper is as follows. Section 2 reviews ENN and MLENN that are base algorithms in prototype selection step. Section 3 describes Kernel version of ELM. Proposed MLENN-KELM is presented in section 4. Section 5 is about experimental setup and results. Finally section 6 concludes this paper.

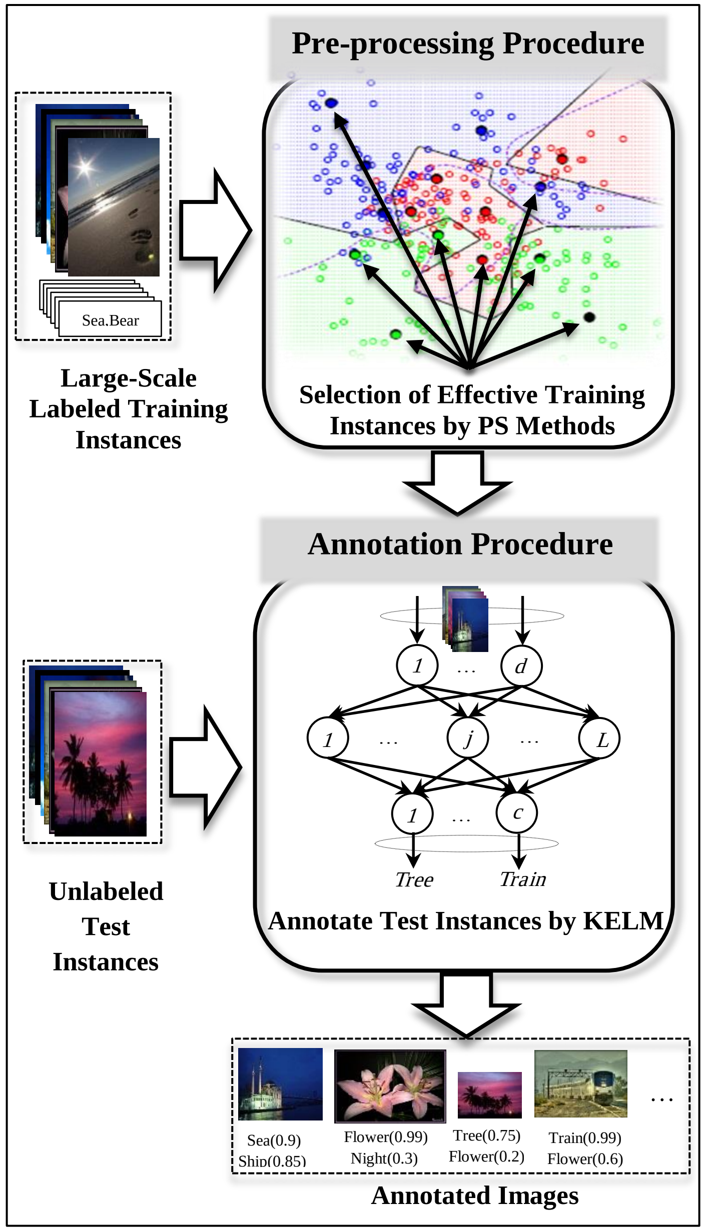

Fig.1 Illustration of proposed MLENN-KELM approach. Large-scale labeled training instances are sent to pre-processing step and the process of selection of effective training instances by PS methods is done. Then the output of reduction phase i.e. reduced training set and unlabeled test instances are sent to annotation step. Annotation procedure is done by using KELM method.

2. ENN and MLENN

Wilson's Edited Nearest Neighbor (ENN) has been proposed in 1972 [17]. This method finds effective instances by using nearest neighbor rule. For each training instance it is removed, if its label vector disagrees with

ACSIJ Advances in Computer Science: an International Journal, Vol. 4, Issue 5, No.17 , September 2015
ISSN : 2322-5157
www.ACSIJ.org

label vector voted from its K nearest neighbors. This process is executed once and for all instances. At the end of this process, remaining instances are prototypes and they can be used as new training set in learning methods.

In order to improve performance of applying learning methods on multi-label datasets, MLENN approach has been proposed [16]. The assumption of MLENN in balancing of imbalanced dataset is identification and preservation of those instances which lead to creation of imbalance dataset. Thus, MLENN identifies all these instances. Then ENN rule applies on remaining instances. Instances which lead to unbalancing in data are identified with two *LblIR* and *DSIR* factors. *LblIR* shows imbalance level of each label. Eq. (1) shows the formula for calculating it.

$$LblIR(y) = \frac{\underset{y'=y_1}{\overset{y_c}{argMax}}\left(\sum_{i=1}^{size(D)} k(y',Y_i)\right)}{\sum_{i=1}^{size(D)} k(y,Y_i)},$$
$$k(y',Y_i) = \begin{cases} 1 & if\ y \in Y_i \\ 0 & otherwise \end{cases} \tag{1}$$

In Eq. (1), y is arbitrary label from label set (Y) and Y_i is label vector of i-th instance. Higher *LblIR* indicates higher imbalance ratio for each label. On the other hand, *DSIR* shows imbalance level of dataset and it is calculated by averaging among all *LblIR*s. Eq. (2) shows its calculation.

$$DSIR = \frac{\sum_{y=y_1}^{y_c}(LblIR(y))}{Size(Y)} \tag{2}$$

The main idea of MLENN in balancing multi-label datasets is preserving those instances that have a label with *LblIR* greater than *DSIR*. Then ENN rule is applied on remaining instances and those instances that do not have ENN condition are removed. With increasing the number of images and annotation words, imbalance property of large-scale image sets is increased. So, in the proposed approach at first by computing *LblIR* and *DSIR* factors, imbalance instances are identified and sent to reduced training set. Then, the ENN is applied on the remaining instances and other effective instances are selected. After selecting effective instances, KELM is applied to predict labels of unlabeled instances.

3. KELM

ELM algorithm is proposed to speed up the learning process in single layer feed forward neural networks. The most important feature of ELM that distinguishes it from back propagation algorithm is random assignments of hidden neurons. In other words, at first and independent from learning process random hidden neuron parameters are assigned and iterative learning process in back propagation algorithm is converted to tuning free learning process [19].

Multiclass and kernel version of ELM have been proposed in [18]. These versions are became popular learning methods in machine learning research area and are used in various applications like image and video processing and understanding [20–23]. As well as the competency of KELM than other learning methods is proved in [18].

The learning process in neural networks is based on finding optimal output weights. Eq. (3) shows output function of an ELM network with L hidden neuron.

$$f(X_i) = \sum_{j=1}^{L} \beta_j h_j(\alpha_j, \beta_j, X_i) = h(X_i)\,\beta\ , i = 1 \ldots N \tag{3}$$

In Eq. (3), $\beta=[\beta_1, \beta_2, \ldots, \beta_L]^T$ is the output weights matrix, $h(X_i)=[h_1(a_1,b_1,X_i), \ldots, h_j(a_L,b_L,X_i)]$ is the network output corresponding to training instance X_i and $h_j(.)$ is a nonlinear or kernel function like RBF. As well as a_j and b_j are the parameters of j-th hidden neuron that is randomly assigned in ELM algorithm. The learning process of ELM network is minimization of $||H\beta-Y||_2$ that H is the hidden layer output matrix and Y is the labels matrix. After solving optimization problem by Karush-Kohn-Tucker (KKT) theorem [24], output weights matrix (β) is obtained based on Eq. (4).

$$\beta = H^T\left(\frac{I}{C} + H^T H\right)^{-1} Y$$
$$= H^T\left(\frac{I}{C} + KM\right)^{-1} Y \tag{4}$$

In Eq. (4), I is the identity matrix and $Km=\{Km(X_i,X_j)\}=\{h(X_i).h(X_j)\}\ ,\ i,j=1 \ldots N$ is the kernel matrix of training instances. After obtaining β in training process, the score vector of each unlabeled instances is calculated by Eq. (3). Then the decision about label assignment based on obtained score values is made. In single label classification, the largest value is considered as label of unlabeled instances. While in multi-label classification the decision making procedure is based on a predefined threshold value.

4. Proposed MLENN-KELM approach

As previously stated, the main idea of MLENN-KELM approach is solving limitations of employing KELM in large-scale AIA. A large-scale training set is sent to MLENN-KELM algorithm. At first and by employing Eq. (1) and (2), all of imbalanced instances are identified and added to reduced training set.

After adding all imbalanced instances to reduced set, the decision about all other remaining instances are made. So,

ACSIJ Advances in Computer Science: an International Journal, Vol. 4, Issue 5, No.17 , September 2015
ISSN : 2322-5157
www.ACSIJ.org

K nearest neighbors of each instance is found. If the voted label vector of nearest neighbors is disagree with original label vector, instance is removed. Decision about disagreement level of label vectors is based on Hamming dissimilarity measure [16]. At the end of removing process, all instances that do not have removing condition are added into reduced set. Then, unlabeled instances are annotated by KELM algorithm. The main argument of using KELM is competency of performance of KELM than other version of ELM [9].

To classify unlabeled test instances, KELM is learned on reduced training set with tuned parameters and the scores matrix is obtained. Then the multiple label assignment process is done. Various methods have been proposed in the literature for assigning labels in multi-label classification [8]. In this paper an adaptive thresholding mechanism has been employed. This mechanism assigns 1 to all of the scores that are θ percent of sum of all scores and assigns 0 to remains. The output of algorithm is the predicted label vector of all unlabeled instances. Fig. 2 shows pseudo code of proposed MLENN-KELM algorithm.

MLENN-KELM algorithm

Inputs:

Trn → Large-scale Labeled Train dataset
Tst → Large-scale unlabeled Test dataset
K → Number of nearest neighbor
C → Regularization coefficient
θ → Threshold parameter for assigning labels according to scores

Outputs:

Tst → Labeled test dataset

Algorithm:

Compute *LblIR* for each label in *Trn* from Eq. (1)
Compute *DSIR* for *Trn* dataset from Eq. (2)
for all Instance X_i in *Trn*
    if X_i has a label with *LblIR* > *DSIR*
        add X_i into *S*//reduced train instances
    else
        Find *K*-NN of X_i in *Trn*
        Y_{Pred}=majority voting of *K*-NN
            if Y_{Pred} and Y_i is agree
                add X_i into *S*
            end of if
    end of else
end of main for
for all Instance X_i in *Tst*
    SC_i = KELM(X_i, *S* ,*C*)
end of for
for all $score_i$ in SC_i
    Tst_i= Label θ percent of summation of SC_i
end of for
Return *Tst*

Fig. 2 Pseudo code of proposed MLENN-KELM algorithm.

5. Experiments

In this section, experimental results of proposed MLENN-KELM approach are described. So, large-scale image annotation dataset descriptions, evaluation criteria and parameter tuning are described in the experimental setup sub-section. As well as the results of applying proposed approach is presented in the results sub-section.

5.1 Experimental Setup

In this paper, NUS-WIDE-Object large-scale image annotation set has been used [25]. This set includes 30000 images crawled from Flickr[1] in 31 object categories. From 30000 images, 17928 images are considered as training set and 12072 images are considered as test set. In all experiments for representing each image three groups of global features are used [3,26]. These features are 144-dimensional HSV color correlogram, 73-dimensional edge direction histogram and 128-dimensional wavelet texture. So each image represents by a 345-dimensional vector.

To evaluate proposed MLENN-KELM approach, precision, recall and F1 classification measures have been used. As well as and to compute these criteria in multi-label learning context, macro and micro averaging strategies have been employed [8,27]. So, for evaluating classification performance, micro precision, macro precision, micro recall, macro recall, micro F1 and macro F1 are calculated.

In MLENN-KELM algorithm, the number of nearest neighbors (*K*), threshold value (θ) and regularization factor in KELM (*C*) must be tuned. These parameters are adjusted by 2-fold cross validation strategy on train and validation sets. The optimal values of *K*, θ and *C* are initialized to 25, 0.18 and 4 respectively. As well as RBF (sig=4.5) is used as kernel function in KELM [9].

5.2 Results

In this sub-section, experimental results of applying proposed MLENN-KELM approach on NUS-WIDE-Object are presented. MLENN-KELM is compared with Base KNN (K=1) [25], ELM [18], Baselines approach [28], RKELM [9] and MLENN [16]. Table 1 shows comparison of annotation performance between different approaches on NUS-WIDE-Object.

[1] www.flickr.com

Table 1: Performance comparison of MLENN-KELM and state-of-the-art methods on NUS-WIDE-Object.

Method	*Macro Precision (%)*	*Macro Recall (%)*	*Macro F1 (%)*	*Micro Precision (%)*	*Micro Recall (%)*	*Micro F1 (%)*
Base KNN (K=1) [25]	23.94	24.89	24.41	31.34	30.80	31.07
MLENN [16]	25.59	23.27	24.37	32.58	30.09	31.29
Baselines [28]	25.78	24.54	25.14	32.31	19.73	24.5
ELM [18]	15.13	32.03	20.55	32.56	50.34	39.54
RKELM [9]	16.74	26.83	20.62	31.08	39.03	34.60
MLENN-KELM	**34.27**	**42.32**	**37.87**	**43.99**	**44.62**	**44.30**

As it can be seen from Table 1, the proposed MLENN-KELM approach has better annotation performance than other methods. The improvements of MLENN-KELM in macro F1 compared to KNN, MLENN, Baselines approach, ELM and RKELM are 13.46%, 13.5%, 12.73%, 17.32% and 17.28% respectively. As well as the significant improvements can be seen in terms of micro F1. The improvements of MLENN-KELM in micro F1 compared to KNN, MLENN, Baselines approach, ELM and RKELM are13.23%, 13.01%, 19.8%, 4.76% and 9.7% respectively.

The good annotation performance of MLENN-KELM than other methods in large-scale AIA shows the importance of using KELM in large-scale AIA. But this method cannot be used on training set with the large number of instances and the large-scale set must be reduced. However, every reduced training set is not suitable for using in KELM and optimal reduced training set should be found.

6. Conclusions

In this research, to resolve challenges of large-scale AIA systems and also KELM memory restrictions in such systems, reducing training set and selecting effective prototypes approach is employed. In the proposed KELM-MLENN approach at first, training set size is reduced by MLENN algorithm. The reduced training set is considered as a new training set in KELM algorithm and unlabeled test instances are annotated. By proposing this approach, kernel methods like KELM that have memory limitations in large-scale learning context can be used in these domains. So, in this paper the challenges of using KELM in large-scale AIA systems are addressed. Experimental results on a large-scale image set showed effectiveness and efficiency of our proposed MLENN-KELM approach.

In other future researches, we are going to improve the annotation performance of large-scale AIA systems by combining prototype and feature selection methods. So, effective instances by intrinsic features are selected and sent to KELM algorithm for annotating unlabeled instances.

References

[1] F. Wang, "A survey on automatic image annotation and trends of the new age", in: Procedia Eng., 2011: pp. 434 – 438. doi:10.1016/j.proeng.2011.11.2526.

[2] X. Ke, S. Li, G. Chen, "Real web community based automatic image annotation", Comput. Electr. Eng. 39 (2013) 945–956.

[3] Y. Yang, F. Wu, F. Nie, H.T. Shen, Y. Zhuang, A.G. Hauptmann, "Web and Personal Image Annotation by Mining Label Correlation With Relaxed Visual Graph Embedding", IEEE Trans. Image Process. 21 (2012) 1339–1351.

[4] R. Li, J. Lu, Y. Zhang, Z. Lu, W. Xu, "A framework of large-scale and real-time image annotation system", in: Artif. Intell. 2009. JCAI'09. Int. Jt. Conf., 2009: pp. 576–579.

[5] C. Shi, Q. Ruan, G. An, "Sparse feature selection based on graph Laplacian for web image annotation", Image Vis. Comput. 32 (2014) 189–201.

[6] Y. Feng, J. Xiao, Y. Zhuang, X. Liu, "Adaptive Unsupervised Multi-view Feature Selection for Visual Concept Recognition", Lect. Notes Comput. Sci. 7724 (2013) 343–357.

[7] F. Angiulli, "Fast nearest neighbor condensation for large data sets classification", Knowl. Data Eng. IEEE Trans. 19 (2007) 1450–1464.

[8] M.-L. Zhang, Z.-H. Zhou, "A review on multi-label learning algorithms", Knowl. Data Eng. IEEE Trans. 26 (2014) 1819–1837.

[9] W. Deng, Q. Zheng, K. Zhang, "Reduced kernel extreme learning machine", in: Proc. 8th Int. Conf. Comput. Recognit. Syst. CORES 2013, 2013: pp. 63–69.

[10] S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, K.R.K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design", Neural Comput. 13 (2001) 637–649.

[11] G. Wu, Z. Zhang, E.Y. Chang, "Kronecker Factorization for Speeding up Kernel Machines.", in: SDM, 2005: pp. 611–615.

[12] K.-M. Lin, C.-J. Lin, "A study on reduced support vector machines, Neural Networks", IEEE Trans. 14 (2003) 1449–1459.

[13] Q. Liu, Q. He, Z. Shi, "Extreme support vector machine classifier", in: Adv. Knowl. Discov. Data Min., Springer, 2008: pp. 222–233.

[14] S. García, J. Luengo, F. Herrera, Data Preprocessing in Data Mining, Springer, 2015.

[15] S. Garcia, J. Derrac, J.R. Cano, F. Herrera, "Prototype Selection for Nearest Neighbor Classification : Taxonomy and Empirical Study", Pattern Anal. Mach. Intell. IEEE Trans. 34 (2012) 417–435.

[16] F. Charte, A.J. Rivera, M.J. del Jesus, F. Herrera, "MLeNN: A First Approach to Heuristic Multilabel Undersampling", in: Intell. Data Eng. Autom. Learn. 2014, Springer, 2014: pp. 1–9.
[17] D.L. Wilson, "Asymptotic properties of nearest neighbor rules using edited data", Syst. Man Cybern. IEEE Trans. (1972) 408–421.
[18] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, "Extreme learning machine for regression and multiclass classification", Syst. Man, Cybern. Part B Cybern. IEEE Trans. 42 (2012) 513–529.
[19] Z. Lin, J. Cao, T. Chen, Y. Jin, Z.-L. Sun, A. Lendasse, "Extreme Learning Machine on High Dimensional and Large Data Applications", Math. Probl. Eng. 501 (2015) 624903.
[20] F. Cao, B. Liu, D.S. Park, "Image classification based on effective extreme learning machine", Neurocomputing. 102 (2013) 90–97.
[21] M. Luo, K. Zhang, "A hybrid approach combining extreme learning machine and sparse representation for image classification", Eng. Appl. Artif. Intell. 27 (2014) 228–235.
[22] R. Minhas, A. Baradarani, S. Seifzadeh, Q.M.J. Wu, "Human action recognition using extreme learning machine based on visual vocabularies", Neurocomputing. 73 (2010) 1906–1917.
[23] B. Lu, G. Wang, Y. Yuan, D. Han, "Semantic concept detection for video based on extreme learning machine", Neurocomputing. 102 (2013) 176–183.
[24] R. Fletcher, Practical methods of optimization, John Wiley & Sons, 2013.
[25] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, Y. Zheng, "NUS-WIDE: a real-world web image database from National University of Singapore", in: Proc. ACM Int. Conf. Image Video Retr., 2009: p. 48.
[26] Z. Ma, F. Nie, Y. Yang, J.R.R. Uijlings, N. Sebe, S. Member, "Web Image Annotation Via Subspace-Sparsity Collaborated Feature Selection", IEEE Trans. Multimed. 14 (2012) 1021–1030.
[27] A. Özgür, L. Özgür, T. Güngör, "Text categorization with class-based and corpus-based keyword selection", in: Comput. Inf. Sci. 2005, Springer, 2005: pp. 606–615.
[28] A. Makadia, V. Pavlovic, S. Kumar, "Baselines for image annotation", Int. J. Comput. Vis. 90 (2010) 88–105.

First Author received his B.Sc. degree in computer engineering, from Birjand University, Birjand, Iran in 2013. Currently, he is a M.Sc. student in the Faculty of Electrical and Computer Engineering in Yazd University, Yazd, Iran. His research interests include machine learning, computer vision, image annotation and retrieval.

Second Author received his B.Sc. in computer engineering from Shahid Beheshti University, Tehran, Iran in 2000 and his MS and PhD in software engineering from Tarbiat Modares University, Tehran, Iran in 2004 and 2013, respectively. Currently, he is an assistant professor in the Faculty of Electrical and Computer Engineering in Yazd University, Yazd, Iran. His research interests include machine learning, computer vision, and software engineering.

Third Author received her B.Sc. degree in computer engineering, from Jahrom University, Fars, Iran in 2011. Currently, she is a M.Sc. student in the Faculty of Electrical and Computer Engineering in Yazd University, Yazd, Iran. Her research interests include machine learning, computer vision, image annotation, retrieval and tag refinement.