International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 3 (2015) pp. 6419-6430

http://www.ripublication.com

Sparse Representation in Active Learning Methods for Remote Sensing Image Classification –A Survey

Shivakumar G.S[1] ,Dr.S.Natarajan[2] and Dr. K.Srikanta Murthy[3]

[1]Faculty, Department of Computer Science and Engineering,
Srinivas Institute of Technology-Mangalore,India
[2]Professor,Department of Information Science,PESIT,Bangalore.
[3]Professor,Department of Computer Science and Engineering ,PESSE,Bangalore
E-mail: shivakumar_gs1@yahoo.co.in

ABSTRACT

This paper is an endeavour to investigate the latest active learning methods with the Sparse Representation in the feature extraction phase of active learning method. Active learning has been realized in classification algorithms for over a decade. This paper examines the remote sensing image classification with the sparse representation perspective. The review of literatures involving the sparse representation, Kernel Sparse representation (KSR) and Multikernel Sparse representation is carried out. Active learning with sparse representation as the feature extraction method has been dealt in detail. Kernel trick has the quality of extracting the similarity between nonlinear features, which would assist in finding the sparse representation of the non-linear features, thus motivating to go forward with Kernel Sparse representation. MultiKernel Sparse representation also known as Multikernel Fusion is an improvement in KSR that would take into account many sophisticated object features for representation. The variants of the Support Vector Machine for different application and the Multiclass SVM based literatures are reviewed. The future suggestions and perspective for the remote sensing image classification methods have been highlighted along with the review.

Keywords: Active Learning, Sparse Representation, Remote Sensing and Classification.

INTRODUCTION

The remote sensing i m a g e classification algorithms are categorized into Unsupervised,

Supervised and object based image analysis. The unsupervised classification used clustering algorithm for grouping the pixels and user identifies the number of bands to use for classification. Supervised classification would take up the learning algorithms to make the classification learnt and act as a black box of classification. Active Learning, which is one of the supervised classification algorithms for remote sensing image classification, has long been a research area and has grown exponentially for the past decade. Support Vector Machine has been plentifully used for the Active Learning method in most of the literature in the past. Feature extraction methods are combined with the Active Learning Methods for image classification. Different feature extraction methods in active learning implementation are not much concentrated in this survey taking into consideration of the amount of data covered. In search of the sparser representation the improvement of the wavelet transform and the curvelet transform were developed and tested. The curvelet transform has been considered as more sparser compared to the wavelet transform [68].The Ortho normal basis to extract Sparse Representation from any signal would not be optimal, and there is no single sparse transform that can be applied on all the signals. Adapting Sparse representation on signal properties and deriving efficient processing operators is must to work different signals. Sparse representation would be a signal specific basis pursuit method that would be a perfect representation for any non-linear signal. The deeper insight of sparse representation into the data structure would make it eligible for the classification algorithms of remote sensing. The sparse representation is the most compact representation by using the linear combination of the building blocks of the data. Remote sensing image classification using sparse implementation and active learning method is taken for a literature review in this paper. The active learning method is the process where the data under learning has to be decided by the human intervention in order to have higher redundancy. Once the signal or data is learnt by the active learning method that would become a black box used for classification. Support Vector Machines (SVMs) proposed by Vapnik [1] are a set of related supervised learning methods used for classification, regression and ranking. SVMs are classification prediction tools that use Machine Learning theory as a principled and very robust method to maximize predictive accuracy for detection and classification. SVMs can be considered as techniques which use hypothesis space of linear separators in a high dimensional feature space, trained with a learning algorithm from optimization theory that makes a learning bias derived from statistical learning theory [1, 2]. Support vector machine is one of the important methods for use in the active learning method. The classification was binary but now the multiclass Support Vector Machine has been introduced for better classification.

This paper is organized as follows, Section II will have survey about the Active Learning (AL) Methods, Section III deals about AL methods in Remote Sensing Image Classification, Section IV puts light on the Sparse representation based AL methods on remote sensing image classification, Section V on the Kernel Sparse representation and MultiKernel based Sparse Representation, and the conclusion follows.

ACTIVE LEARNING METHODS

Active Learning method is the supervised learning method that needs human intervention for deciding what needs to be learnt and what not. For a prior knowledge regarding supervised learning reference [3] is highly recommended in some of the previous literatures. The supervised learning would reduce the time consumption that would be usually taken for machine learning by making the algorithm learn itself.

Active learning systems attempt to overcome the labelling bottleneck by asking queries in the form of unlabeled instances to be labelled by an oracle (e.g., a human annotator). In this way, the active learner aims to achieve high accuracy using as few labelled instances as possible, thereby minimizing the cost of obtaining labelled data. Active learning is well motivated in many modern machine-learning problems where data may be abundant but labels are scarce or expensive to obtain. Note that this kind of active learning is related in spirit, though not to be confused, with the family of instructional techniques by the same name in the education literature [4].

In [5] the first active learning circumstances were probed in learning through membership queries that takes random examples from the whole dataset [5]. Instead of taking the unlabeled instances from the dataset generated the learner generates the set of unlabeled instance to be learnt in the membership queries method. The number of queries needed to be included and that must be excluded are demarcated in [6], and also defines that an efficient query synthesis makes the active learning more tractable. The extension of these query synthesize method to regression learning task was carried out for coordinate prediction of the robotic hand by getting the angles of the joints as inputs [7]. If the oracle is a human annotator, then there will be lot of problems applying query synthesis by choosing the arbitrary samples. This problem was reflected in [8] while training the Neural Network [NN] with the help of human oracles for classifying hand written characters. The unforeseen problem that they faced was that there were no identifiable symbols in many of the query images generated by the learner. To get rid of this hitch the stream based and pool-based methods that are taken into account in the coming section is taken up.

Selective sampling replaced query synthesizing in literature [9]. As the name indicates in selective sampling each of the unlabeled instances is taken into consideration, so that it would be decided by the learner whether to involve it for labelling. This approach is also called as stream-based or sequential active learning.

This approach is sometimes called stream-based or sequential active learning, as each unlabeled instance is typically drawn one at a time from the data source, and the learner must decide whether to query or discard it. Selective sampling methodology would act like a membership query learning technique where the membership oracle acts like the function of the uniform distribution, meaning that the precondition is that the sampling must be uniform. Even if there is a non-uniform distribution queries will still be sensible, as they are coming from the actual distribution. The application of this stream-based approach were considered in several literatures like in Grammatical Tagging [10] in which the text is analyzed to discover the context of occurrence, by putting into test the adjacent words along with word under scrutiny.

Other applications like the sensor scheduling in networking application, which selects the network, which has to sense at any instance from a group of network sensors, which we find in [11]. In [12-13] this approach is used in the search engine for information retrieval, where the query under scrutiny is ranked for prioritizing the query sample. The overall effort for the annotation gets reduced in the stream-based sampling method, which helps in increasing the performance of the classification algorithm.

Due to the abundant real world data, the active learning using the stream-based sampling would become a tedious process bringing into play a method called pool-based sampling [14]. It is assumed that there is a small set of labelled data L and a large pool of data U, from which a query is selected at random and involved in active learning. The pool-based sampling had a broad application space wherein it was applied in text classification [14-17], information extraction [20,21], video classification and retrieval [22,23] speech recognition, and in cancer diagnosis [24].

When some real world application involves larger dataset, pool-based active learning would be suitable since it reduces the learning time by ranking the query before selecting it for learning. However, the stream-based sampling would consume more time in sequentially selecting the query and training. Stream-based method can be used where very less dataset is used like in mobile devices and in embedded processors.

ACTIVE LEARNING METHODS IN REMOTE SENSING CLASSIFICATION

Remote sensing classification methods have the accountability of automatically generating the land cover maps, which is largely accomplished by the supervised learning based classification techniques. The accuracy of obtaining the best land cover map depends on the number and quality of the reference samples taken for this purpose [25]. Consistent set of labelled samples selection is a long process and also costly in real time applications, which affects the accuracy of classification [25].In order to get rid of this cost and time issues the Active Learning (AL) methods are exclusively used in the remote sensing classification literature [25]-[33]. AL method needs interaction between the human experts or the oracle and the automatic classification system for expanding the initial training set to the fully labelled set. Information abundant samples from the unlabeled samples are taken at each iteration, and the classification system interacts with the oracle to get the true class label of the sample. The supervised Active learning algorithm would train the newly added labelled samples to complete the process[25]-[33]. The process is done in iterative manner in order to effectively train the minimum number of the training samples that are considered. Only few of the literatures are putting light on the cost incurred for the query analysis and labelling it in the AL processes.[31]-[33]. The total travelling time between the samples are taken as the measure to assess the labelling cost of any set [31], [32].The optimization techniques are used to find the cost effective active learning method in [34]. The Cost Sensitive Active Learning (CSAL) [34] method targets to decrease the labelling expenses with respect to both samples availability and travelling time required to visit the samples, however obtaining precise classification maps. Regulating the travelling area of the supervisor by choosing the little study quantity of the image; and optimizing the selection of informative and cost efficient samples by a genetic algorithm optimization method can provide an optimal solution to attain this.

Most of the literature was dedicated to choose the ambiguous and unlabeled pixels by describing enhanced spectral-spatial standards to select it [35,36]. Increasing the classifier precision in the AL procedure carried out reduction of human interactions. Semi supervised learning methods that jointly targets to learn both the labelled and unlabelled samples are an important alternative to the previous AL methods [37]. Graph theory has been used to label the samples as labelled and unlabeled in this method through the combinatorial Laplacian matrix [38]. After the regularization of the classifier function the final solution is represented in the kernel space. It is worth recalling that this view has been commonly accepted for the support vector machine (SVM) classifiers [38]. Nevertheless, in the AL context, this learning mode will: 1) overburden the classifier; 2) only permit an incomplete misuse of the spatial– contextual knowledge; and 3) not permit a proper exploitation of the power of graph-based optimization methods. The literature [39] fuses the competences of an extreme learning machine (ELM) classifier and graph-based optimization methods to advance the classification accuracy while diminishing the operator collaboration. Compared with the existing learning methods such as the SVM, the ELM classifier is described by several attractive modesties: 1) it has a unified formulation for binary, multiclass, and regression problems, and the solution of these problems is also given in a unified analytical form; 2) the feature mapping could be done either in a known space similar to neural networks or in an infinite space similar to kernel methods; and 3) for multiclass classification, it uses a configuration of multi-output nodes where the number of nodes is equal to the number of classes.

In [40] the technique studies the 1-D output space of the classifier to classify the most uncertain samples whose labelling and inclusion in the training set involve a high probability to advance the classification results. A simple histogram-thresholding algorithm is used to find out the low-density (i.e., under the cluster assumption, the most uncertain) region in the 1-D SVM output space. After training, a histogram is constructed in the 1-D output space of the classifier by considering the output scores of the unlabeled samples in $[-1, +1]$. Since the classifier ranks each sample from the most likely members to the most unlikely members of a class, the samples whose output scores fall in the valley region of the histogram (low-density region of the kernel space) are the most uncertain/ambiguous.

Present strategies articulate the active learning problem merely on the spectral domain. In [41], by discovering the fact that, remote sensing images are intrinsically defined both in the spectral and spatial domains propose a novel active learning approach for support vector machine classification. More specifically, it is recommended that relating spectral and spatial information straight in the iterative process of sample selection. For this purpose, three standards are proposed for the selection of samples aloof from the samples already comprising the existing training set. In the first approach, the Euclidean distances in the spatial domain from the training samples are plainly computed, while the second one is based on the Parzen window technique in the spatial domain. Conclusively, the last criterion involves the concept of spatial entropy.

SPARSE REPRESENTATION BASED ACTIVE LEARNING METHODS FOR REMOTE SENSING IMAGE CLASSIFICATION

Sparse representations from over-complete dictionaries have been highlighted as one of the crucial principles in signal processing. Much of the excitement comes from the discovery that sparse representation can be reduced to linear programming or second conic programming when the solution is sparse enough [42-44]. Following this thread, considerable research has been performed to explore structures to solve sparse coding [45-47], dictionary learning [48-51], compressive sensing [52,53]. Literature [54] present a novel unified framework, i.e. BMSAL (Batch Mode Sparse Active Learning). Based on the existing sparse family of classifiers, it defines thoroughly the corresponding BMSAL family and explores their shared properties, most prominently (approximate) sub-modularity. The first one that stimulates is to optimize the algorithms and conduct experiments comparing with state-of-the-art methods; for reliability, error-bounded algorithms are given, as well as detailed logical deductions and empirical tests for employing sparse in non-linear data sets are done.

KERNEL BASED ACTIVE LEARNING AND SPARSE REPRESENTATION

In [55] the problem considered is the problem of dictionary learning and sparse coding, where the task is to find a brief set of basis vectors that precisely represent the observation data with only small numbers of active bases. Characteristically formulated as an L1-regularized least-squares problem, the problem experiences computational difficulty originating from the no differentiable objective. Contemporary approaches to sparse coding thus have mainly attentive on quickening of the learning algorithm. It is stretched to nonlinear sparse coding using kernel trick by showing that the represented theorem holds for the kernel sparse coding problem. Presently, kernel methods in general and support vector machines (SVMs) in specific dictate the field of discriminative data classification models [56]. Through the last few years, the approaches have been effectively familiarized in the field of remote sensing image classification [57], [58]. Kernel methods deal efficiently with low-sized datasets of potentially high dimensionality, as in the case of hyper spectral images. The use of the kernel trick [59], as is known in the literature, allows kernel methods to work in higher dimensional (possibly infinite-dimensional) spaces requiring the knowledge of only a kernel function, which calculates an inner product in the new space using the original data. Also, since kernel methods do not adopt an explicit prior data distribution but are inherently non-parametric models, they cope well with remote sensing data specificities and complexities. Alternative Bayesian approaches to remote sensing processing problems also exist and have been introduced as well to

Earth observation applications. For example, the relevance vector machine (RVM) [60] assumes a Gaussian prior over the weights to enforce sparsity and uses expectation-maximization to infer the parameters. In [61],[62], the RVM was used for multispectral image segmentation and landmine detection using ground penetrating radar, while in the model was used for adaptive biophysical parameter retrieval. Lately, Gaussian Processes [64] have received much attention in the field of machine learning, and some applications and developments have been introduced in remote sensing data processing as well, both for classification [65]and parameter retrieval settings. Motivated by the fact that kernel trick can capture the nonlinear similarity of features, which helps in finding a sparse representation of nonlinear features, [66] it proposes kernel sparse representation (KSR). Essentially, KSR is a sparse coding technique in a high dimensional feature space mapped by an implicit mapping function. KSR is applied to feature especially when a small fraction of the data is used for coding in image classification and face recognition. Furthermore, in [67] to overcome the vulnerability of single feature in object description, it is proposed a multikernel fusion method for multifeature integration.

The details or main advantages of our method are summarized as follows.

1) The utilization of the kernel method permits us to introduce some complex features in SR, such as spatial-color and spatial-gradient histograms. Compared with the raw template used in the traditional SR-based tracking methods, the histogram features are less sensitive to partial occlusion, illumination variation, and object deformation. Thus, our tracking results can be more precise.
2) Due to the kernel method, it is prevented to introduce a large number of unimportant templates. Therefore, this KSR-based tracking algorithm speeds up at least four times as compared with the traditional SR-based methods.
3) The adoption of multikernel fusion makes it necessary to incorporate multifeature fusion to achieve a balancing effect in object representation. Experimental results indicate that the method can outperform the state-of-the art approaches. The idea of this multikernel fusion method can be implemented in the sparse based remote sensing image classification for reduced computational complexity and higher accuracy in dynamic environment.

CONCLUSION

The remote sensing image classification methods have been exclusively reviewed with the sparse representation perspective and future techniques for implementation are also suggested in this review paper. The way to reduce the computational complexity and increasing the accuracy has been discussed with the support of different literatures. Kernel Sparse representation and the Multikernel Sparse representation methods are taken into consideration from the other domains of the image processing and signal processing in order to put a light on future remote sensing classification method perspective. The Multiclass Support Vector Machine also has been discussed with the application perspective and the review tells that the multiclass SVM would be a future prospective in the classification field with more dynamic environment like the hyper spectral remote sensing image classification which has been in use.

REFERENCES

[1] V. Vapnik, “The natural of statistical learning theory,” Springer, New York, 1995.

[2] B. Scholkopf, and A. Smola, “Learning with kernels,” MIT Press, Cabridge, MA, 2002.

[3] T. Mitchell.”Machine Learning”. McGraw-Hill, 1997.

[4] C. Bonwell and J. Eison. “ Active Learning: Creating Excitement in the Classroom” Jossey-Bass, 1991.

[5] D. Angluin. “ Queries and concept learning” Machine Learning, 2:319–342, 1988.

[6] D. Angluin. “Queries revisited. In Proceedings of the International Conference on Algorithmic Learning Theory”, pages 12–31. Springer-Verlag, 2001.

[7] D. Cohn, Z. Ghahramani, and M.I. Jordan. “ Active learning with statistical models”. Journal of Artificial Intelligence Research, 4:129–145, 1996.

[8] K. Lang and E. Baum. “Query learning can work poorly when a human oracle is used” In Proceedings of the IEEE International Joint Conference on NeuralNetworks, pages 335–340. IEEE Press, 1992.

[9] D. Cohn, L. Atlas, R. Ladner, M. El-Sharkawi, R. Marks II, M. Aggoune, and D. Park. “Training connectionist networks with queries and selective sampling”, Advances in Neural Information Processing Systems (NIPS). Morgan Kaufmann,1990.

[10] I. Dagan and S. Engelson. “ Committee-based sampling for training probabilistic classifiers”. Proceedings of the International Conference on Machine Learning (ICML), pages 150–157. Morgan Kaufmann, 1995.

[11] V. Krishnamurthy. “ Algorithms for optimal scheduling and management of hidden markov model sensors”. IEEE Transactions on Signal Processing, 50(6):1382–1397, 2002.

[12] H. Yu. SVM “Selective sampling for ranking with application to data retrieval”,Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD), pages 354–363. ACM Press, 2005.

[13] A. Fujii, T. Tokunaga, K. Inui, and H. Tanaka. “ Selective sampling for example based word sense disambiguation”. Computational Linguistics, 24(4):573–597,1998

[14] D. Lewis and W. Gale. “A sequential algorithm for training text classifiers”. Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval, pages 3–12. ACM/Springer, 1994.

[15] A. McCallum and K. Nigam. “Employing EM in pool-based active learning for text classification”, Proceedings of the International Conference on Machine Learning (ICML), pages 359–367. Morgan Kaufmann, 1998.

[16] S. Tong and D. Koller. “ Support vector machine active learning with applications to text classification”, In Proceedings of the International Conference on Machine Learning (ICML), pages 999–1006. Morgan Kaufmann, 2000.

[17] S.C.H. Hoi, R. Jin, and M.R. Lyu. “Large-scale text categorization by batch mode active learning”, Proceedings of the International Conference on the World Wide Web, pages 633–642. ACM Press, 2006a.

[18] C.A. Thompson, M.E. Califf, and R.J. Mooney.”Active learning for natural language parsing and information extraction”, Proceedings of the International Conference on Machine Learning (ICML), pages 406–414. Morgan Kaufmann, 1999.

[19] B. Settles and M. Craven. “ An analysis of active learning strategies for sequence labelling tasks” ,Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1069–1078. ACL Press, 2008.

[20] S. Tong and E. Chang. “ Support vector machine active learning for image retrieval”, Proceedings of the ACM International Conference on Multimedia, pages 107–118. ACM Press, 2001.

[21] C. Zhang and T. Chen. “ An active learning framework for content based information retrieval”. IEEE Transactions on Multimedia, 4(2):260–268, 2002.

[22] R. Yan, J. Yang, and A. Hauptmann. “Automatically labelling video data using multi-class active learning”, Proceedings of the International Conference on Computer Vision, pages 516–523. IEEE Press, 2003.

[23] G. T¨ur, D. Hakkani-T¨ur, and R.E. Schapire. “ Combining active and semisupervised learning for spoken language understanding”. Speech Communication, 45(2):171–186, 2005.

[24] Y. Liu. “ Active learning with support vector machine applied to gene expression data for cancer classification”, Journal of Chemical Information and Computer Sciences, 44:1936–1941, 2004.

[25] B. Demir, C. Persello, and L. Bruzzone, “Batch mode active learning methods for the interactive classification of remote sensing images”, IEEE Trans. on Geoscience and Remote Sensing, vol. 49, no.3, pp. 1014-1031, March 2011.

[26] P. Mitra, B. U. Shankar, and S. K. Pal, “Segmentation of multispectral remote sensing images using active support vector machines”, Pattern Recognition. Lett., vol. 25, no. 9, pp. 1067-1074, Jul. 2004.

[27] S. Rajan, J. Ghosh, and M. M. Crawford, “An active learning approach to hyperspectral data classification”, IEEE Trans. on Geoscience and Remote Sensing, vol. 46, no. 4, pp. 1231-1242, Apr. 2008.

[28] D. Tuia, F. Ratle, F. Pacifici, M. Kanevski, and W. J. Emery, “Active Learning methods for remote sensing image classification”, IEEE Trans. on Geo science and Remote Sensing, vol. 47, no. 7, pp. 2218 -2232, Jul. 2009.

[29] S. Patra and L. Bruzzone, “A fast cluster-based active learning technique for classification of remote sensing images”, IEEE Transactions on Geo science and Remote Sensing, vol. 49, no.5, pp.1617-1626, 2011.

[30] S. Patra, L. Bruzzone, “A cluster-assumption based batch mode active learning technique”, Pattern Recognition Letters, vol. 33, 2012, pp.1042-1048.

[31] A. Liu, G. Jun and J. Ghosh, “Spatially cost-sensitive active learning”,In SIAM International Conference on Data Mining, Sparks, Nevada, USA, pp.814-825, 2009.

[32] A. Liu, G. Jun, and J. Ghosh, “Active learning of hyperspectral data with spatially dependent label acquisition costs”, IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, pp. V-256 - V-259, 2009.

[33] B. Demir, L. Minello, L. Bruzzone, “A Cost-Sensitive Active Learning Technique for the Definition of Effective Training Sets for Supervised Classifiers”, International Conference on Geoscience and Remote Sensing Symposium,Munich, Germany, 2012.

[34] Begu□mDemir, Luca Minello, and Lorenzo Bruzzone,” A Genetic Algorithm Based Cost-Sensitive Active Learning Technique for Classification of Remote Sensing Images” , 2012 CNIT Tyrrhenian Workshop

[35] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W. J. Emery, “SVM active learning approach for image classification using spatial information,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 4, pp. 2217–2233, Apr. 2014.

[36] M.M. Crawford, D. Tuia, and H. L. Yang, “Active learning: Any value for classification of remotely sensed,” Proc. IEEE, vol. 101, no. 3, pp. 593–608, Jul. 2013.

[37] O. Chappele, B. Scholkopf, and A. Zienn, “Semisupervised Learning”,Cambridge, MA, USA: MIT Press, 2006.

[38] M. Belkin, P. Niyogi, and V. Sindhwani, “Manifold regularization: A geometric framework for learning from labeled and unlabeled examples,” J. Mach. Learn. Res., vol. 7, pp. 2399–2434, 2006

[39] Mohamed A. Bencherif, Yakoub Bazi, Abderrezak Guessoum, Naif Alajlan, Farid Melgani and Haikel Al Hichri,” Fusion of Extreme Learning Machine and Graph-Based Optimization Methods for Active Classification of Remote Sensing Images” IEEE transactions in Geoscience and remote sensing, vol. 12, no. 3, March 2015

[40] Swarnajyoti Patra, and Lorenzo Bruzzone,” A Fast Cluster-Assumption Based Active-Learning Technique for Classification of Remote Sensing Images” IEEE Transactions on Geoscience And Remote Sensing, Vol. 49, No. 5, May 2011

[41] EdoardoPasolli, FaridMelgani, DevisTuia, Fabio Pacifici and William J. Emery “SVM Active Learning Approach for Image Classification Using Spatial Information” IEEE Transactions on Geoscience And Remote Sensing, Vol. 52, No. 4, April 2014

[42] D. L. Donoho. “Sparse and redundant representations “Comm. Pure and Applied Math, 59(6):797 – 829, 2006.

[43] E. J. Cand`es and T. Tao. “ Decoding by linear programming” IEEE Transactions on Information Theory, 51(12):4203 – 4215, 2005.

[44] E. Cand`es, M. Rudelson, T. Tao, and R. Vershynin. " Error correction via linear programming", In FOCS'05, pages 295–308, 2005.
[45] S. S. Chen, D. L. Donoho, and M. A. Saunders"Atomic decomposition by basis pursuit", SIAM Rev., 43(1):129–159, 2001.
[46] J. A. Tropp and A. C. Gilbert. "Signal recovery from random measurements via orthogonal matching pursuit", IEEE Transactions on Information Theory, 53(12):4655–4666, 2007.
[47] D. L. Donoho, Y. Tsaig, I. Drori, and J. Starck. " Sparse solution of underdetermined linear equations by stage wise orthogonal matching pursuit", Technical Report 2006-02, Standford, Department of Statistics,2006.
[48] D. P. Wipf and B. D. Rao." Sparse bayesian learning for basis selection" IEEE Transactions on Signal Processing, 52(8):2153 – 2164, 2004.
[49] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman." Discriminative learned dictionaries for local image analysis", In CVPR'08, 2008.
[50] M. Aharon, M. Elad, and A. Bruckstein." K-svd: An algorithm for designing over complete dictionaries for sparse representation" IEEE Transactions on Signal Processing, 54(11):4311 – 4322, 2006.
[51] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. " Supervised dictionary learning" In NIPS'08, pages 1033–1040, 2008.
[52] P. Indyk"Explicit constructions for compressed sensing of sparse signals", In SODA '08, pages 30–33, 2008.
[53] E. J. Candes and J. Romberg" Quantitative robust uncertainty principles and optimally sparse decompositions", Found. Comput. Math., 6(2):227–254, 2006.
[54] Lixin Shi, Yuhang Zhao," Batch Mode Sparse Active Learning", 2010 IEEE International Conference on Data Mining Workshops
[55] Minyoung Kim, " Efficient Kernel Sparse Coding Via First-Ordersmooth Optimization", IEEE Transactions on Neural Networks And Learning Systems, Vol. 25, No. 8, August 2014.
[56] B. Sch¨olkopf and A. Smola, "Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond", MIT Press Series, Cambridge, MA, USA, 2002.
[57] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," IEEE Trans. on Geoscience and Remote Sensing, vol. 43, no. 6, pp. 1351–1362, June 2005.
[58] G. Camps-Valls and L. Bruzzone, "Kernel Methods for Remote Sensing Data Analysis", John Wiley and Sons, 2009.
[59] C. M. Bishop, " Pattern Recognition and Machine Learning" (Information Science and Statistics), Springer, 2007.
[60] M. E. Tipping, "The relevance vector machine," Journal of Machine Learning Research, vol. 1, pp. 211–244, 2001.
[61] P. Torrione and L.M. Collins, "Texture features for antitank landmine detection using ground penetrating radar," IEEE Trans. on Geoscience and Remote Sensing, vol. 45, no. 7, pp. 2374 –2382, July 2007.

[62] F. A. Mianji and Y. Zhang, "Robust hyperspectral classification using relevance vector machine," IEEE Trans. on Geoscience and Remote Sensing, vol. 49, no. 6, pp. 2100 –2112, June 2011.

[63] G. Camps-Valls, L. G´omez-Chova, J. Vila-Franc´es, J. Amor´os-L´opez, J. Mu˜noz-Mar´ı, and J. Calpe-Maravilla, "Retrieval of oceanic chlorophyll concentration with relevance vector machines," Remote Sensing of Enviroment, vol. 105, no. 1, pp. 23–33, Nov 2006.

[64] C.E. Rasmussen and C.K. Williams, " Gaussian Processes for Machine Learning", MIT Press, NY, 2006.

[65] Y. Bazi and F. Melgani, "Gaussian process approach to remote sensing image classification," IEEE Trans. on Geoscience and Remote Sensing vol. 48, no. 1, pp. 186 –197, January 2010.

[66] ShenghuaGao, Ivor Wai-Hung Tsang, And Liang-Tien Chia," Sparse Representation with Kernels", IEEE Transactions On Image Processing, Vol. 22, No. 2, February 2013

[67] Lingfeng Wang, Hongping Yan, KeLv, and Chunhong Pan,"Visual Tracking Via Kernel Sparse Representation With Multikernel Fusion" IEEE Transactions On Circuits And Systems For Video Technology, Vol. 24, No. 7, July 2014

[68] Jean-Luc Starck and Jerome Bobin, "Astronomical Data Analysis and Sparsity: from Wavelets to Compressed Sensing" Astronomical Data Analysis ,Proceedings of the IEEE Volume:98 , No: 6 , June 2010

[69] https://www.ceremade.dauphine.fr/~peyre/wavelet-tour/05-ch01-p374370.pdf