A Survey on Unsupervised Data Learning Approaches

Parul Agrawal[1] and Dr. S. S. Sane[2]
[1]ME student,[2]Head computer department,
[1]Department of Computer Engieering,
[1]K. K. Wagh Institute of Engineering Education & Research, Nashik,
Savitribai Phule Pune University, Maharashtra, India

Received: December 08, 2018 Accepted: January 08, 2019

ABSTRACT: *Researchers from machine learning community are gaining interest in the development of kernel-based learning algorithms in the recent years. The performance of algorithms totalyrelies on the choice of the kernel function. Kernel function absolute presents the inner product among a pairof points of a dataset in a higherdimensional space. The selection of an appropriate kernel for agiven dataset is the crucial and important dispute in kernel-based learning. Extracting knowledge using unsupervised learning techniques from a large amount of unlabeled heterogeneous data is a big challenge in big data these days. In this paper we present a deep literature survey on unsupervised learning methods for knowledge extraction.*

Key Words: *Unsupervised learning, heterogeneous data, clustering, extreme learning machine, Kernel-based algorithms*

I. Introduction

Kernel-based algorithms have recentlygained a significant attention in machine learning community. Supervised algorithms such assupport vector machine (SVM) and kernel discriminant analysis as well as unsupervised algorithms like kernel principle component analysis (kernel-PCA) and support vector clustering have been successfully applied to various real-world problems. Due to lower error rate compared toother learning methods, relatively fast training time and elegant compatibility with highdimensional data these algorithms are the potential solutions to many problems.In most real-world data analytics problems, a huge amount of data are collected from multiple sources without label information,which is much with various types, structures, namely heterogeneous data.

II. Literature Survey

Mostly all the existing system uses two learning paradigms. The one is unsupervised deep learning that utilizes deep models to handle large data complexities. The other one is unsupervised multiple view learning that leverages heterogeneous information from multiple views/modes. More recently, unsupervised MKL [1]-[3] has been studied to tackle the heterogeneous data learning without supervised labels. Similar to MKL, unsupervised MKL also uses multiple kernels to distill information from various sources. To enable the learning without supervised labels, it introduces a kernel-based unsupervised learning objective, e.g. kernel *k*-means [4], to learn the optimal kernel combination coefficients. Although unsupervised MKL achieves remarkable performance in unsupervised heterogeneous data learning, most of the current unsupervised MKL methods are with a slow learning speed.

2.1 Unsupervised Deep Learning

Deep learning[5]allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have enhanced the state-of-the-art in speech recognition, visual object recognition, object detection & many other domains such as drug discovery and genomics. Deep learning discovers complex structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in every layer from the representation in the previous layer. The unsupervised deep learning method combines an unsupervised objective and deep neural networks to learn a powerful data representation [6].

Many efforts try to learn unsupervised data representation in adversarial approaches [7]-[9],which simultaneously take the advantages of both deep generator and deep discriminator. Authors in [8] proposes a class of CNNs called deep convolutional generative adversarial networks (DCGANs), that have certain architectural constraints, anddemonstrate that they are a strong candidate for unsupervised learning. Trainingon various image datasets, they have also show convincing evidence that deep convolutional adversarial pair learns a hierarchy of representations from object parts toscenes in both the generator and discriminator.

Nowadays, deep learning is one of the keenest research trends inmachine learning field. In contrast to most traditional learning methods, which are believedusing shallow-structured learning architectures, deeplearning primarily uses supervised &/or unsupervisedstrategies in deep architectures to mechanically learnhierarchical representations [33]. Deep architecturescan catchmore complicated, hierarchicallylaunched statistical patterns of inputs for attaining tobe adjustive to new areas than traditional learningtechniques& often outperform state of the artachieved by hand-made features [34]. Deep beliefnetworks (DBNs) [33,35] & convolutional neuralnetworks (CNNs) [36,37] are two mainstream deeplearning methods & research directions proposedover the past decade, which have been well establishedin the deep learning field and shown great promise forfuture work [13].

In the recent years because of the state-of-the-art performance of deeplearning, it has attracted much attention from theacademic community such as speechrecognition, computer vision, language processing,and information retrieval [33,38–40]. As the datakeeps getting bigger, deep learning is coming to play apivotal role in providing predictive analytics solutionsfor large-scale data sets, particularly with theincreased processing power and the advances ingraphics processors [13]. For example, IBM'sbrain-like computer [22] and Microsoft'sreal-timelanguage translation in Bing voice search [41]have used techniques like deep learning toleverage big data for competitive advantage.

2.2 Unsupervised Multiple View Learning

Conventional machine learning algorithms, such as support vector machines, discriminant analysis, kernel machines, and spectral clustering, concatenate all multiple views intoone single view to adapt to the learning setting. Still, this chain causes overfitting in the case of a small size training sample & is not physically significant as each view has a specific statistical property. In contrast tosingle view learning, multi-viewlearning as a new prototypebrings in one function to model aparticular view & jointlyoptimizes all the functions to exploit the redundant views of the same input data &improve the learning performance. Therefore, multi-view learning has been receiving more attention.

Authors in [21] presents a new optimized kernel k-means algorithm (OKKC) which combines duple data sources for clusteringanalysis. An alternating minimization framework is used to modify the cluster membership & kernel coefficients as a non convex issue. The problem to reduce the cluster membership & the kernel coefficients are all based on the same Rayleigh quotient objective. OKKC has a simpler procedure & lower complexity. After that, the work in [2] proposes a Heterogeneous mEtric Learning with hIerarchical Couplings (HELIC for short) for this type of categorical data. HELIC catches both low-level value-to-attribute & high-level attribute-to-class hierarchical couplings, & reveals the intrinsic heterogeneities embedded in each level of couplings. Authors in [1] present a robust kmeans using `2;1-norm in the feature space and then extend it to the kernel space. To recap the powerfulness of kernel methods, author further propose a novel robust multiple kernel k-means (RMKKM) algorithm that simultaneously finds the best clustering label, the cluster membership and the optimal combination of multiple kernels. An alternating iterative schema is developed to find the optimal value.

A novel MKC algorithm with a "local" kernel alignment[3] , which only requires that the similarity of a sample to its k-nearest neighbours be aligned with the ideal similarity matrix. Such an alignment helps the clustering algorithm to focus on closer sample pairs that shall stay together and avoids involving unreliable similarity evaluation for farther sample pairs. Also a new optimization problem to implement this idea, and design a two-step algorithm to efficiently solve it is proposed.The proposed model[22] introduces the concept of mapping function to make the different patterns from different pattern spaces comparable and hence an optimal pattern can be learned from the multiple patterns of multiple representations. Under this model, we formulate two specific models for two important cases of unsupervised learning, clustering and spectral dimensionality reduction; an iterating algorithm is derived for multiple view clustering, and a simple algorithm providing a global optimum to multiple spectral dimensionality reduction. They also extend the proposed model and algorithms to evolutionary clustering and unsupervised learning with side information.

Table 1 COMPARISON OF MACHINE LEARNING TECHNOLOGIES[22]

Learning Types	Data processing task	Distinct form	Learning algorithm
Supervised learning	Classification/Regression/Estimation	Computational classifiers	Support vector machine[23]

		Statistical classifiers	Naïve Bayes[24] Hidden Markov model[25] Bayesian networks[26]
		Connectionist classifiers	Neural networks[27]
Unsupervised learning	Clustering/Prediction	Parametric	K-means[28] Gaussian mixture model[29
		Nonparametric	Dirichlet process mixture model[29] X-means[28]

III. Conclusion

In this paper a deep literature survey is presented on unsupervised data learning approaches we have therefore reviewed several currenttrends of multi-view learning and deep learning. Through analyzing these different approaches to the integration of multiple views, we observe that they mainly dependon either the consensus principle or the complementary principle to ensure their success.Although unsupervised MKL achieves remarkable performance in unsupervised heterogeneous data learning, most of the current unsupervised MKL methods are with a slow learning speed. So there is a need of to develop a system adaptively adjust the base kernels to fit the dynamic heterogeneousdata distributions.

References

1. L. Du et al., ``Robust multiple kernel K-means using `2;1-norm,'' in Proc. Int. Conf. Artif. Intell., 2015, pp. 3476_3482.
2. M. Gönen and A. A. Margolin, ``Localized data fusion for kernel K-means clustering with application to cancer biology,'' in Proc. 7th Int. Conf. Neural Inf. Process. Syst., vol. 1, 2014, pp. 1305_1313.
3. M. Li, X. Liu, L. Wang, Y. Dou, J. Yin, and E. Zhu, ``Multiple kernel clustering with local kernel alignment maximization,'' in Proc. 25th Int. Joint Conf. Artif. Intell., 2016, pp. 1704_1710.
4. I. S. Dhillon, Y. Guan, and B. Kulis, ``Kernel K-means: Spectral clustering and normalized cuts,'' in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2004, pp. 551_556.
5. Y. LeCun, Y. Bengio, and G. Hinton, ``Deep learning,'' Nature, vol. 521, pp. 436_444, May 2015.
6. Y. Bengio, A. Courville, and P. Vincent, ``Representation learning: A review and new perspectives,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798_1828, Aug. 2013.
7. I. Goodfellow et al., ``Generative adversarial nets,'' in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672_2680.
8. A. Radford, L. Metz, and S. Chintala. (2015). ``Unsupervised representation learning with deep convolutional generative adversarial networks.'' [Online]. Available: https://arxiv.org/abs/1511.06434
9. M. Arjovsky, S. Chintala, and L. Bottou, ``Wasserstein generative adversarial networks,'' in Proc. Int. Conf. Mach. Learn., 2017, pp. 214_223.
10. D Yu, L Deng, Deep learning and its applications to signal and information processing. IEEE Signal Proc Mag28 (1), 145 –154 (2011)
11. I Arel, DC Rose, TP Karnowski, Deep machine learning-a new frontier in artificial intelligence research. IEEE Comput Intell Mag 5(4), 13– 18 (2010)
12. Y Bengio, Learning deep architectures for AI. Foundations Trends Mach Learn 2(1), 1–127 (2009)
13. R Collobert, J Weston, L Bottou, M Karlen, K Kavukcuoglu, P Kuksa, Natural language processing (almost) from scratch. J Mach Learn Res 12 , 2493 – 2537 (2011)
14. P Le Callet, C Viard-Gaudin, D Barba, A convolutional neural network approach for objective video quality assessment. IEEE Trans Neural Network17(5), 1316–1327 (2006)
15. GE Dahl, D Yu, L Deng, A Acero, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Trans Audio Speech Lang Proc 20 (1), 30–42 (2012)
16. G Hinton, L Deng, Y Dong, GE Dahl, A Mohamed, N Jaitly, A Senior, V Vanhoucke, P Nguyen, TN Sainath, B Kingsbury, Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Proc Mag 29 (6), 82 – 97 (2012)
17. DC Ciresan, U Meier, LM Gambardella, J Schmidhuber, Deep, big, simple neural nets for handwritten digit recognition. Neural Computing 22 (12), 3207 –3220 (2010)
18. Y Wang, D Yu, Y Ju, A Acero, Voice search, in Language understanding: systems for extracting semantic information from speech (Wiley, New York, 2011)
19. N Jones, Computer science: the learning machines. Nature 505 (7482), 146 – 148 (2014)
20. XW Chen, X Lin, Big data deep learning: challenges and perspectives. IEEE Access 2, 514–525 (2014)
21. S. Yu et al., ``Optimized data fusion for kernel K-means clustering,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 5, pp. 1031_1039, May 2012.
22. Junfei Qiu, Qihui Wu, Guoru Ding*, Yuhua Xu and Shuo Feng "A survey of machine learning for big data Processing", EURASIP Journal on Advances in Signal Processing

23. J Zheng, F Shen, H Fan, J Zhao, An online incremental learning support vector machine for large-scale data. Neural Comput Appl 22(5), 1023–1035 (2013)
24. TM Mitchell, Machine learning (McGraw-Hill, New York, 1997)
25. C Ghosh, C Cordeiro, DP Agrawal, M Bhaskara Rao, Markov chain existence and hidden Markov models in spectrum sensing, in Proceedings of the IEEE International Conference on Pervasive Computing & Communications (PERCOM) (Galveston, 2009), pp. 1–6
26. K Yue, Q Fang, X Wang, J Li, W Weiy, A parallel and incremental approach for data-intensive learning of Bayesian networks. IEEE Trans Cybern 99, 1–15 (2015)
27. X Dong, Y Li, C Wu, Y Cai, A learner based on neural network for cognitive radio, in Proceedings of the 12th IEEE International Conference on Communication Technology (ICCT) (Nanjing, 2010), pp. 893–896
28. A El-Hajj, L Safatly, M Bkassiny, M Husseini, Cognitive radio transceivers: RF, spectrum sensing, and learning algorithms review. Int J Antenn Propag 11(5), 479–482 (2014)
29. M Bkassiny, SK Jayaweera, Y Li, Multidimensional dirichlet process-based non-parametric signal classification for autonomous self-learning cognitive radios. IEEE Trans Wirel Commun 12(11), 5413–5423 (2013)