

Semi-Supervised Ensemble Clustering

Ruchi Jain, Prof. L.A.Deshpande

Abstract - Semi-Supervised cluster has restricted supervising with in the kind of labeled instances to help unattended cluster and considerably improves the performance of cluster. Despite the amount of knowledgeable data on this, existing work isn't designed for handling huge higher dimensional knowledge. There are varied limitations in ancient cluster approach. Victimization of solely previous data provided by supervisor. Good performance in high dimensional datasets. All of the load values of the ensemble members are equal, which ignores totally different contributions from different ensemble members. Not all constraints contribute to the ultimate result. To overcome these limitations we tend to propose Double Weight Semi-Supervised Ensemble cluster.

Keywords - Semi-Supervised Cluster, Cluster ensemble, Adaboost, Constraint Clustering.



IJSER

I. INTRODUCTION

Semi-supervised cluster (SSC) is the downside of cluster unlabeled knowledge with the support of the aspect data provided by a supervisor (who are often associated in nursing knowledgeable data or associate in oracle system). Since its high success in recent years, SSC has received vital attention from researchers. The aspect data has been shown to guide the cluster algorithms towards the required cluster solutions or facilitate the cluster algorithms break loose the native minima effectively. The aspect data will contribute not solely to the performance improvement however conjointly to the quality reduction.

An example is the automobile land distinguishing downside from GPS knowledge however the goal is to cluster knowledge points into totally different lanes. This can be a troublesome cluster downside for the well-known cluster rule KMEANS as a result of the lane clusters have an awfully special form that is extremely elongated and parallel to the road center line. And therefore the KMEANS with constraints has achieved the high accuracy in comparison of accuracy of the KMEANS with no constraints. The works that are done up to now are often classified into one among the subsequent 2 themes: the a-priori scheme, the interactive theme. Within the a-priori theme, the aspect data is given once before applying the SSC rule whereas within the interactive theme, the aspect data is collected iteratively by interacting with the supervisor.

Although 2 glorious surveys by Davidson et al. [16] and Basu et al. [7] have lined main aspects of the apriori theme, there's still no surveys that conjointly the opposite theme. Besides, some recent vital algorithms also are missing from these surveys. This survey involves fill in this want with the hope that it will gift not solely a lot of general read however conjointly a deeper read of this field for brand new researchers. The algorithms given during this paper are classified into common techniques for straightforward comparison. The pseudo-code likewise because the benefits and drawbacks of every rule are going to be given clearly. Additionally, the open problems are going to be conjointly summarized within the survey.

Currently, there are 2 themes for SSC that are the apriori scheme, and therefore the interactive theme.

they're primarily totally different by the approach the aspect data is collected in every theme. within the initial theme, all aspect data is given once before the SSC rule is dead whereas within the second theme, aspect data is collected iteratively by interacting with the supervisor.

A Priori theme

In the a priori theme, the SSC rule reads all aspect data once and uses these data to enhance the cluster performance. Several works following this theme are worn out literature and split into differing kinds of aspect data like labeled knowledge, instance-level or cluster-level constraints.

Interactive theme

In this interactive theme, the SSC rule presents the cluster result Associate in nursing a question to a supervisor who are often users or an oracle system. Then the supervisor studies the result and pro- vides feedback to the SSC rule. The SSC

rule successively analyses the feedback and adapts this data to bias the cluster method. The interaction between the SSC rule and therefore the supervisor is stopped once some convergence condition is glad. The feedback are often collected in 2 following ways in which supported the role of the supervisor and therefore the SSC rule. If the supervisor plays the active role, then he/she actively provides the constraints to the SSC rule. With in the case that the SSC rule is the active role, the SSC rule can create queries to the supervisor and therefore the supervisor is meant to answer these queries. The second approach has been shown to beat the primary approach in literature. The primary approach needs the supervisor should apprehend that the foremost informative constraints to produce for the SSC rule whereas within the second approach, this troublesome task is on the aspect of the SSC rule, and it's higher if the SSC rule is allowed to raise what it's not clear than passively receives inapplicable feedback from the supervisor. The algorithms within the initial approach are going to be referred as the passive SSC algorithms, whereas those within the second approach are going to be referred as the active SSC algorithms. So far, solely few works are worn out this theme. An interactive SSC rule integrates the constraints by everchanging the gap metric and alternative active SSC algorithms that uses the farthest distance, data gain, density and coassociation confidence to pick out the foremost informative constraints.

II. RELATED WORK

Z. Yu, L. Li, Wong H-S, You J., G. Han, Y. Gao, G. Yu, "Probabilistic Cluster Structure Ensemble", In this paper, there is a unique probabilistic cluster structure ensemble framework designed, stated as Guassian mixture model based cluster structure ensemble framework (GMMSE), to spot the foremost representative cluster structure from the dataset. Specifically, GMMSE initially applies the KMeans approach to supply a collection of variant datasets. Then, a collection of Gaussian mixture models accustomed capture the underlying cluster structures of the datasets. GMMSE applies K-means to initialize the values of the parameters of the Gaussian mixture model, and adopts the Expectation Maximization approach (EM) to estimate the parameter values of the model. Next, the elements of the Gaussian mixture models are viewed as new knowledge samples that are accustomed construct the representative matrix capturing the relationships among elements[1]

Z. Yu, X. Zhu, H. S. Wong, J. You, J. Zhang, G. Han, "Distribution- Based Cluster Structure Selection", This paper investigates the matter of a way to choose the acceptable cluster structures within the ensemble which is able to be summarized to a lot of representative cluster structure. Specifically, the cluster structure is initially drawn by a combination of Gaussian distributions, the parameters of that are calculable victimization the expectation maximization rule. Then, many distribution primarily based on distance functions are designed to judge the similarity between 2 cluster structures. supported the similarity comparison results, we tend to propose a brand new approach, that is stated because the distribution-based cluster structure

ensemble (DCSE) framework, to search out the foremost representative unified cluster structure. They tend to then style a brand new technique, the distribution-based cluster structure choice strategy, to pick out a set of cluster structures.

They propose employing a distribution based normalized hyper graph cut rule to come up with the ultimate result[2]

Z. Yu, L. Li, J. Liu, J. Zhang, G. Han, "Adaptive Noise Immune Cluster Ensemble Using Affinity

Propagation", the target of cluster ensemble is to mix multiple cluster solutions in an exceedingly appropriate thanks to improve the standard of the cluster result. during this paper, we tend to style a

brand new noise immune cluster ensemble framework named as AP2CE to tackle the challenges raised by buzzing datasets. AP2CE not solely takes advantage of the affinity propagation rule (AP) and therefore the normalized cut rule (Ncut), however conjointly possesses the characteristics of cluster ensemble. Compared with ancient cluster ensemble approaches, AP2CE is characterized by many properties. (1) It adopts multiple distance operates rather than one geometer distance operate to avoid the noise associated with the gap function. (2) AP2CE applies AP to prune buzzing attributes and generate a collection of recent datasets within the subspaces consists of representative attributes obtained by AP. (3) It avoids the express specification of the quantity of clusters. (4) AP2CE adopts the normalized cut rule because the accord operate to partition the accord matrix and procure the ultimate result[3]

K. Tademir, Y. Moazzen and I. Yildirim, "An Approximate Spectral Clustering Ensemble for High Spatial Resolution Remote-Sensing Images", unattended cluster of high abstraction resolution remote-sensing pictures plays a big role in elaborate land cowl identification, particularly for agricultural and environmental watching. A recently promising technique is approximate spectral cluster (SC) that permits spectral partitioning for big datasets to extract clusters with distinct characteristics while not a constant model. It conjointly facilitates the employment of assorted data varieties via advanced similarity criteria. However, it needs Associate in Nursing empirical choice of a similarity criterion optimum for the corresponding application. to handle this challenge, we tend to propose Associate in Nursing approximate SC ensemble (ASCE2) that fuses partitioning obtained by totally different similarity representations.[4]

J. T. Tsai, Y. Y. Lin, H. Y. M. Liao, "Per-Cluster Ensemble Kernel Learning for Multi-Modal Image Clustering With Group-Dependent Feature Selection", during this paper, we tend to gift a cluster approach, MK-SOM that carries out cluster dependent feature choice, and partitions pictures with multiple feature representations into clusters. This work is impelled by the observations that human visual systems (HVS) will receive varied types of visual cues for deciphering the planet. pictures known by HVS because the same class are usually coherent to every alternative in sure crucial visual cues, however the crucial cues vary from class to class. To account for this observation and bridge the linguistics gap, the projected MK-SOM integrates multiple kernel learning

(MKL) into the coaching method of self organizing map (SOM), and associates every cluster with a learnable, ensemble kernel. Hence, it will leverage data captured by varied image descriptors, and discoveries the cluster-specific characteristics via learning the per-cluster ensemble kernels.[5]

SEMI-SUPERVISED CLUSTER ALGORITHMS WITH LABELS

The problem of SSC with labeled knowledge provided by users as per the aspect data is outlined. Given a dataset X , the goal is to separate this dataset into K disjoint clusters $K_h=1$ such that some objective is decreased (often locally). Let $S \in X$ be the set of knowledge objects and referred to as the seed set. The aspect data is given as follows: for every $x_i \in S$, the label $Y_i = h$ of x_i denotes the cluster X_h that x_i belongs to. The seed set S is partitioned off into L disjoint set $L_h=1$ wherever $L \leq K$. If $L = K$, the seed set is named complete. Otherwise, it's the case of incomplete seeding.

Basu et al. projected 2 versions of KMeans that create use of labeled knowledge as aspect data for improving the KMeans performance [4]. Within the initial rule Seeded-KMeans, the seed set is employed to initialize cluster centers. Every cluster center μ_h is computed because the mean of knowledge objects with the label of h within the seed set. If for a few cluster X_h , there's no labeled knowledge objects there to, its center is initialized by random perturbations of the global center. And so the KMeans rule is applied on the entire dataset as was common. The concept of SeededKMeans is that a decent seed set will guide KMeans towards a decent region of search space.

In the Seeded-KMeans, the cluster memberships of knowledge objects within the seed set are often modified within the assignment step of KMeans. Therefore, to keep these memberships unchanged, the information objects within the seed set should be skipped within the assignment step. This modification results in the Constrained-KMeans rule. Once the seed set is noise-free or the user doesn't need the modification within the labels of the seed set, Constrained-KMeans is a lot of appropriate than Seeded-KMeans. However, if the seed set is buzzing, Seeded-KMeans is meant to be higher as a result of it doesn't have to be compelled to keep the labels unchanged and so the buzzing labels are often removed by KMeans.

SEMI-SUPERVISED CLUSTER ALGORITHMS WITH CONSTRAINTS

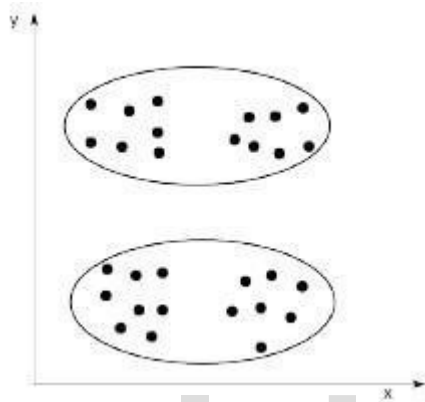
In several applications, the labeled knowledge isn't obtainable whereas the constraints between instances or the constraints on clusters are easier to gather. Constraints are often divided into instance-level and cluster-level constraints. Instance-Level Constraints

Instance-level constraints, conjointly referred to as pairwise constraints, the constraints between knowledge objects. There are 2 styles of instance-level constraints that are must-link and cannot-link introduced by Wagstaff. A must-link $c=(x, y)$ or a cannot-link $c_6=(x, y)$ constraint between 2 objects x and y means these 2 objects should or should not be within

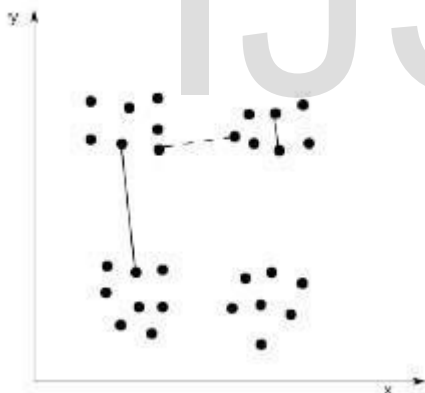
the same cluster, re- spectively. The must-link constraint is Associate in nursing equivalence relation as a result of it's reflexive, radially symmetrical and transitive. Besides, cannot-link constraints are often entailed from connected elements CC_i wherever every connected element CC_i could be a utterly connected subgraph by must-link constraints.

Constraint-Based cluster

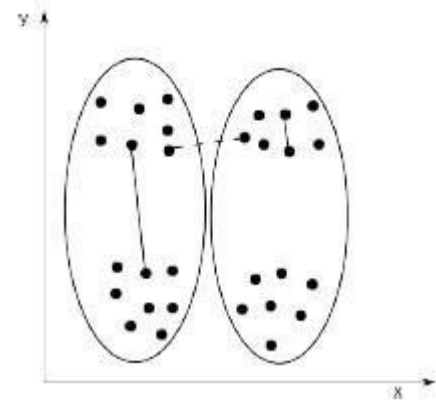
In this approach, the first cluster rule is changed to integrate the constraints so the search strategy is biased towards the solutions that respect these constraints as several as attainable. These constraints are often reverred strictly or partly on the various cluster algorithms.



a. Input instances without Constraints



b. Input instances and constraints. Mustlink constraints are denoted as solid lines, and cannot-link constraints are denoted as dashed lines.



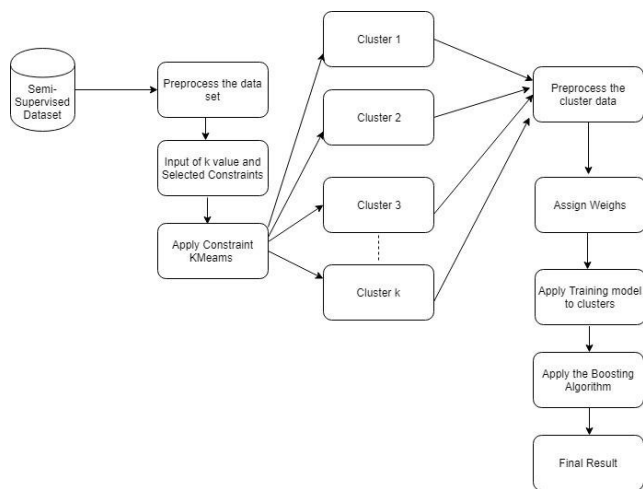
c. A clustering solution that satisfies all constraints

The first part of this section will present the agglomerated hierarchical cluster algorithms. Hierarchical cluster (HC) is wide employed in several areas of science to explain the data structure of knowledge. The goal of HC is to construct a cluster hierarchical or a tree of clusters, conjointly called dendrogram, from knowledge objects. HC algorithms are chiefly classified into: agglomerative (bottom-up) and divisive (top-down) approach. The agglomerative approach starts with singleton clusters (each singleton cluster could be a knowledge object) and recursively merge the 2 most similar clusters to larger clusters till the required range of clusters is achieved. In distinction, the divisive approach starts with a cluster consisting all knowledge objects, and in turn splits every cluster into little clusters till a stopping condition is met. Until now, solely the agglomerative cluster is tailored to figure with aspect data.

III. PROPOSED SYSTEM

Previous authors processed sixteen real-world datasets from UCI machine learning repository. For analysis, heap of cases used micro-precision to live the accuracy of the cluster with reference to actuality labels. In our experiments, the constraints are generated as follows: for every constraint, one try of knowledge points are picked out indiscriminately from exemplars of the input file sets (the labels of which are obtainable for analysis purpose however unavailable for clustering). If the pair of labels are same, then a ML constraint is generated. If the labels are totally different, a CL constraint is generated. The amounts of constraints are determined by the scale of input file.

Then the clusters formed are passed to the boosting algorithm (Adaboost) for further enhancement of the features in the existing cluster.



Flow of Algorithm

IV. CONSTRAINT K-MEANS

Constrained data clustering produces clusters by using 2 types of pair constraints: mustlink and cannotlink. Constraints are a means of supervisor that constrain a pair of points that belong to the similar cluster (must-link) or dissimilar clusters (cannot-link) or that simply describe whether a pair of data points are same or different. The clusters are hence formed by means of these constraints. Further these clusters are given to the boosting algorithm Adaboost

V. ADABOOST

AdaBoost is for binary classification. AdaBoost is used to boost the performance of decision trees on binary classification. Each instance in the training dataset is weighted. The initial weight is set to:

$$wgt(xi) = 1/n$$

Where xi is the i'th training instance and n is the number of training instances

A decision stump is prepared on the train data using the weighted samples. Only binary (two-class) classification problems are supported, so each decision stump makes one decision on one input variable and outputs a +1.0 or -1.0 value for the first or second class value.

The misclassification rate is calculated for the trained model. Traditionally, this is calculated as:

$$err = (correct - N) / N$$

Where error is the misclassification rate, correct are the number of training instance predicted correctly by the model and N is the total number of training instances.

Modification to use the weighting of the training instances:

$$err = \sum(w(i) * terr(i)) / \sum(w)$$

Which is the weighted sum of the misclassification rate, where w is the weight for training instance i and terr is the prediction error for training instance i which is 1 if misclassified and 0 if correctly classified. A stage value is calculated for the trained model which provides a weighting for any predictions that the model makes. The stage value for a trained model is calculated as follows:

$$stage = \ln((1-err) / err)$$

Where stage is the stage value used to weight predictions

from the model, ln() is the natural logarithm and error is the misclassification for the model. The effect of the stage weight is that more accurate models have more weight or contribution to the final prediction.

The training weights are updated giving more weight to incorrectly predicted instances, and less weight to correctly predicted instances.

For example, the weight of one training instance (w) is updated using:

$$w = w * \exp(stage * terr)$$

Where w is the weight for a specific training instance, exp() is the numerical constant e raised to a power, stage is the misclassification rate for the weak classifier and terr is the error the weak classifier made predicting the output variable for the training instance, evaluated as:

$$terr = 0 \text{ if}(y == p), \text{ otherwise } 1$$

Where y is the output variable for the training instance and p is the prediction from the weak learner. This is the effect of not changing the weight if the training instance was classified correctly and making the weight slightly larger if the weak learner misclassified the instance.

VI. DATASET

Data set includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms. Each species is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended. Number of Instances in data set are 8124.

VII. EXPERIMENTAL RESULTS

Semi-Supervised Clusters based on Selected constraints which further acquires high accuracy in comparison to existing models. Uses prior knowledge for selecting the constraints. Each link contributes to formation of better clusters. Handles high-dimensional data. Applies different weights to different instances so that the contribution is not ignored. And overcome the various limitations of ancient cluster ensemble approaches. The system hence has planned double coefficient semi-supervised ensemble cluster supported selected constraint projection that applies constraint coefficient and ensemble member coefficient to deal with ancient limitations.

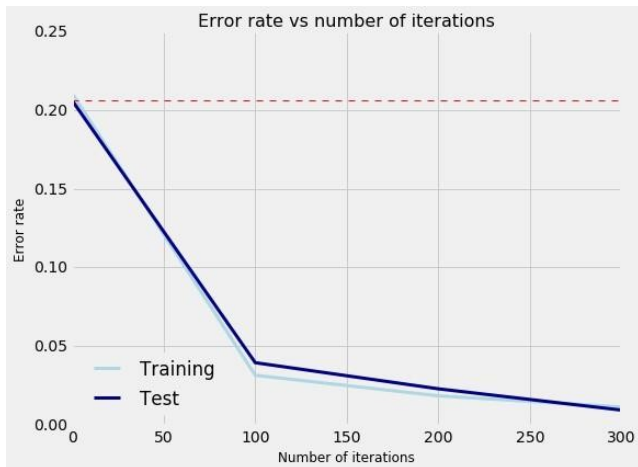
```

2424 2427 2432 2435 2437 2440 2445 2451 2456 2457 2459 2471 2473 2474 2475 2483 2489 2490 2491 2493 2496 2504 2505 2506 2511 2513 2514 2520 2530 2540 2543 2547
2552 2553 2558 2562 2563 2570 2571 2576 2581 2585 2590 2591 2596 2597 2598 2600 2613 2616 2627 2628 2621 2624 2626 2632 2635 2639 2644 2648 2649
2693 2694 2697 2698 2699 2706 2708 2712 2727 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749
2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779
2822 2823 2824 2825 2826 2827 2828 2829 2830 2831 2832 2833 2834 2835 2836 2837 2838 2839 2840 2841 2842 2843 2844 2845 2846 2847 2848 2849 2850 2851 2852 2853
2922 2923 2924 2925 2926 2927 2928 2929 2930 2931 2932 2933 2934 2935 2936 2937 2938 2939 2940 2941 2942 2943 2944 2945 2946 2947 2948 2949 2950 2951 2952 2953
3204 3205 3206 3207 3208 3209 3210 3211 3212 3213 3214 3215 3216 3217 3218 3219 3220 3221 3222 3223 3224 3225 3226 3227 3228 3229 3230 3231 3232 3233 3234
4059 4062 4270 4297 4283 4245 4247 4283 4291 4357 4374 4388 4419 4465 4466 4467 4519 4534 4611 4661 4662 4673 4699 4777 4808 4884 4812 4824 4826 4841 4844 4845
4850 4858 4868 4865 4871 4880 4899 4913 4918 4926 4932 4939 4945 4951 4958 4964 4965 4969 4996 5015 5020 5062 5063 5068 5069 5083 5085 5071 5101 5183 5188 5113
9134 9133 9138 9129 9127 9129 9134 9135 9144 9138 9135 9136 9137 9146 9143 9139 9179 9178 9180 9181 9182 9183 9184 9185 9186 9187 9188 9189 9190 9191 9192 9193
5203 9224 9238 9233 9243 9246 9251 9252 9254 9256 9257 9264 9266 9271 9276 9288 9285 9291 9295 9296 9306 9307 9314 9316 9318 9319 9320 9321 9322 9323 9324
5357 5362 5371 5372 5373 5374 5377 5378 5381 5382 5383 5387 5389 5392 5393 5395 5407 5413 5415 5416 5417 5428 5421 5426 5429 5433 5435 5441 5445 5447 5450 5451 5456
5462 5463 5470 5479 5488 5481 5487 5490 5488 5489 5888 5889 5984 5985 5986 5987 9325 9326 9327 9328 9329 9330 9331 9332 9333 9334 9335 9336 9337 9338 9339 9340 9341 9342
5588 5590 5570 5574 5575 5588 5590 5592 5612 5618 5619 5622 5624 5628 5638 5645 5647 5649 5652 5653 5658 5661 5663 5669 5670 5675 5677 5679 5684 5686 5688 5691
5695 5696 5698 5705 5706 5708 5712 5715 5716 5728 5726 5729 5732 5734 5735 5737 5748 5752 5753 5754 5755 5756 5757 5758 5759 5760 5761 5762 5763 5764 5765 5766 5767
5887 5888 5828 5822 5823 5825 5828 5848 5849 5881 5882 5885 5888 5891 5892 5893 5894 5896 5874 5877 5880 5886 5887 5884 5888 5890 5896 5897 5898 5899 5904 5912 5918 5921
5938 5936 5937 5938 5939 5941 5942 5946 5945 5953 5959 5961 5969 5975 5976 5996 6128 6084 6038 6041 7822 7863 7873 7116 7181 7380 7353 7368 7370 7377 7415 7422 7444 7481
7548 7597 7682 7173 7784 7795 7721 7898 7919 7948 7946 7952 7965 8028
    
```

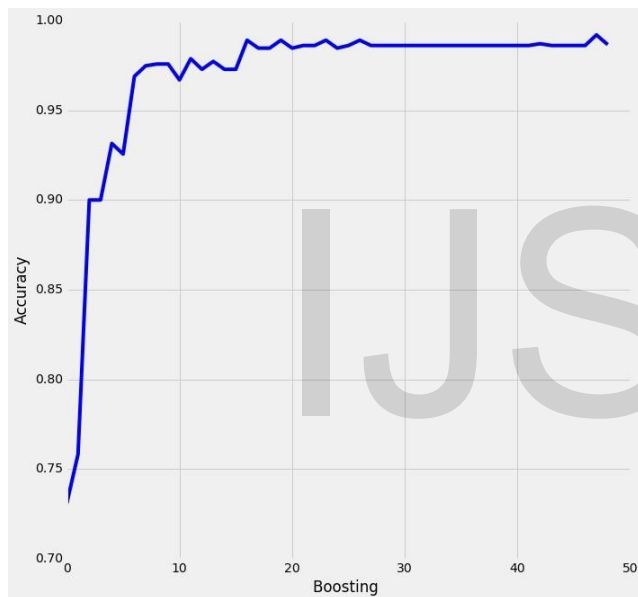
```

m=
n=
a=1
b=1
evaluation_must_Link=0.3333333333333333
evaluation_cannot_Link=0
evaluation_greedy_w=5
evaluation_overall=0.3333333333333333
    
```

a. Evaluation of the MustLink and CannotLink



b. Error Rate Vs. No. of Iterations after applying Adaboost



c. Accuracy Vs. Boosting

In Fig a. we see the Constraint K-Means Evaluation of the clusters. These formed clusters are given to the Adaboost algorithm and from Fig b. and Fig c. we see that the error rate decreases with the no. of iterations. We also achieve accuracy and performance on big data by the proposed double weight semi-supervised ensemble strategy.

VIII. CONCLUSION

We tend to propose the double weight semisupervised ensemble cluster supported with chosen constraint projection to handle the restrictions of ancient semi-supervised cluster ensemble strategies. The projected approach has 3 benefits compared with typical semi-supervised cluster ensemble strategies. (1) It adopts the random topological space technique together with the constraint project procedure to perform high dimensional knowledge cluster. (2) It generates totally different subsets of selected constraints to scale back the result of redundant constraints. (3) Associate in nursing

adaptive ensemble member weight method is meant to emphasize the importance of various ensemble members, and avoid the result of harmful ensemble members. There are many potential directions for future analysis. First, we tend to have an interest in mechanically distinguishing the proper range for the reduced spatiality supporting the background knowledge from providing a prespecified price. Second, we tend to explore different strategies to use supervising in guiding the unattended cluster, e.g., supervised feature cluster.

IX. REFERENCES

- [1].S. Basu, Davidson, and K. L. Wagstaff, *Constrained Clustering*. Boca Raton, FL, USA: CRC Press, 2008.
- [2].O. Chapelle, A. Zien, and B. Scholkopf, *SemiSupervised Learning*. Cambridge, MA, USA: MIT Press, 2006.
- [3].K. Wagstaff, C. Cardie, S. Rogers, and S. Schroedl, "Constrained Kmeans clustering with background knowledge," in *Proc. ICML*, 2001, pp. 577-584.
- [4].D. Klein, S. D. Kamvar, and C. D. Manning, "From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering," in *Proc. ICML*, 2002, pp. 307-314.
- [5].N. Shental, A. Bar-Hillel, T. Hertz, and D. Weinshall, "Computing gaussian mixture models with em using equivalence constraints," in *Proc. NIPS*, 2003.
- [6].S. Basu, M. Bilenko, and R. J. Mooney, "A probabilistic framework for semi-supervised clustering," in *Proc. KDD*, 2004, pp. 59- 68.
- [7].M. Bilenko, S. Basu, and R. J. Mooney, "Integrating constraints and metric learning in semisupervised clustering," in *Proc. ICML*, 2004, pp. 81- 88.
- [8].B. Kulis, S. Basu, I. Dhillon, and R. J. Mooney, "Semisupervisedgraphclustering: A kernel approach," in *Proc. ICML*, 2005, pp. 457-464.
- [9] B. Yan and C. Domeniconi, "An adaptive kernel method for semisupervised clustering," in *Proc. ECML*, 2006, pp. 18-22.
- [10].D. Y. Yeung and H. Chang, "A kernel approach for semi-supervised metric learning," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 141-149, Jul. 2007.
- [11].S. C. H. Hoi, R. Jin, M. R. Lyu, and J. Wu, "Learning nonparametric kernel matrices from pairwise constraints," in *Proc. ICML*, 2007, pp. 361- 368.
- [12].O. Masayuki and Y. Seiji, "Learning similarity matrix from constraints of relational neighbors," *J. Adv. Comput. Intell. Intell. Inform.*, vol. 14, no. 4, pp. 402-407, 2010.
- [13].X. Yin, S. Chen, and E. H. D. Zhang, "Semisupervised clustering with metric learning: An adaptive kernel method," *Pattern Recognit.*, vol. 43, no. 4, pp. 1320-1333, 2010.
- [14].C. Domeniconi, J. Peng, and B. Yan, "Composite kernels for semisupervised clustering," *Knowl. Inform. Syst.*, vol. 28, pp. 99-116, Aug. 2011.
- [15].W. Tang, H. Xiong, S. Zhong, and J. Wu, "Enhancing semi-supervised clustering: A feature projection perspective," in *Proc. KDD*, 2007, pp. 707- 716.

- [16.]D. Zhang, S. Chen, Z. Zhou, and Q. Yang, "Constraint projections for ensemble learning," in Proc. AAAI, 2008, pp. 758-763.
- [17.]F. Wang, "Semisupervised metric learning by maximizing constraint margin," IEEE Trans. Neural Netw., vol. 41, no. 4, pp. 931-939, Aug. 2011.
- [18.]H. Zeng and Y.-M. Cheung, "Semi-supervised maximum margin clustering with pairwise constraints," IEEE Trans. Knowl. Data Eng., vol. 24, no. 5, pp. 926-939, May 2012.
- [19.]A. Mignon and F. Jurie, "Pcca: A new approach for distance learning from sparse pairwise constraints," in Proc. CVPR, 2012, pp. 2666-2672.
- [20.]W. Chen and G. Feng, "Spectral clustering: A semi-supervised approach," Neurocomputing, vol. 77, no. 1, pp. 229-242, Jan. 2012.
- [21.]F. Shang, Y. Liu, and F. Wang, "Learning spectral embedding for semisupervised clustering," in Proc. ICDM, 2011, pp. 597-606.
- [22.]D. Zhang, S. Chen, and Z. Zhou, "Constraint score: A new filter method for feature selection with pairwise constraints," Pattern Recognit., vol. 41, no. 5, pp. 1440-1451, May 2008.
- [23.]E. R. Eaton, "Clustering with Propagated Constraints," Thesis, Univ. Maryland, College Park, MD, USA, 2005.
- [24.]J. Huang and H. Sun, "Lightly-supervised clustering using pairwise constraint propagation," in Proc. 3rd Int. Conf. Intell. Syst. Knowl. Eng., 2008, pp. 765-770.
- [25.]Z. Li, J. Liu, and X. Tang, "Pairwise constraint propagation by semidefinite programming for semisupervised classification," in Proc. ICML, 2008, pp. 576-583.
- [26.]Z. H. Zhou and W. Tang, "Clusterer ensemble," Knowledge-Based Syst., vol. 19, no. 1, pp. 77-83, 2006.
- C. Ding and T. Li, "Adaptive dimension reduction using discriminant analysis and K-means clustering," in Proc. ICML, 2007, pp. 521-528.
- [27.]S. Garcla, A. Fernandez, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," Inform. Sci., vol. 180, no. 10, pp. 2044-2064, May 2010.