Utilizing Fused Features to Mine Unknown Clusters in Training Data

Robert S. Lynch, Jr.
Signal Processing Branch
Naval Undersea Warfare Center
Newport, RI, U.S.A.
lynchrs@npt.nuwc.navy.mil

Peter K. Willett
ECE Department
University of Connecticut
Storrs, CT, U.S.A.
willett@engr.uconn.edu

Abstract - ***In this paper, a previously introduced data mining technique, utilizing the Mean Field Bayesian Data Reduction Algorithm (BDRA), is extended for use in finding unknown data clusters in a fused multidimensional feature space. In the BDRA the modeling assumption is that the discrete symbol probabilities of each class are a priori uniformly Dirichlet distributed, and where the primary metric for selecting and discretizing all relevant features is an analytic formula for the probability of error conditioned on the training data. In extending the BDRA for this application, notice that its built-in dimensionality reduction aspects are exploited for isolating and automatically sorting out and mining all points contained in each unknown data cluster. In previous work, this approach was shown to have comparable performance to the classifier that knows all cluster information when mining a single feature containing multiple unknown clusters. Therefore, the primary contribution of the work presented here is to demonstrate that this approach can be extended to cases where the features are fused and contain more than one dimension. To illustrate performance, results are demonstrated using simulated data containing multiple clusters, and where the fused feature space contains relevant classification information.***

Keywords: Adaptive classification, Level two fusion, Discrete data, Unknown data distribution.

1 INTRODUCTION

In [7, 8], the problem of classifying all points in an unknown data cluster was investigated, where the domain of the observed data, or features, describing each class was obscure and highly overlapped. However, within difficult domains such as these it was also discussed that in some situations the target class of interest (e.g., data that produce a desired yield and are thus categorized as the target class) can contain isolated unknown clusters (i.e., subgroups of data points), where the observations within each cluster have similar statistical properties. Notice that in this problem yield represents the primary variable to separate target and nontarget data points, and where stronger yielding points are more desirable. In general, a variable such as the yield is only known for the training data and not the test data. Thus, an important goal in classifying the test data is to choose data points that produce a strong and consistent average yield.[1] Within these situations it turns out that classification performance (i.e., through a minimum probability of error, or high average yield) can be significantly improved if one develops a classifier to recognize, or mine, observations within the clusters as the target class, and where all other nonclustered observations (i.e., both with and without a desired yield) are considered the alternative class (the nontarget class).[2] As was shown previously for the case of mining a single unknown data cluster, a benefit of such a classifier is that subsets of target data points, producing a consistent desired average yield, can be recognized with a minimum probability of error (see the figures in [7, 8]). This was further shown to be in contrast to what is obtained using a traditional supervised learning classification approach, where this latter case produced a much higher probability of error and a lower average yield.

2 Review of mining unknown clusters in a single dimension

Figure 1 illustrates a straightforward example of the problem of interest with a plot containing one thousand samples of one dimensional domain data (a single feature). The data for this figure was generated, for each dimension of each class (i.e., except those within the cluster), to be uniform, independent, and identically distributed. However, with respect to the features each data cluster was generated as Gaussian distributed, with a randomly generated mean, and con-

[1] As an intuitive example, consider financial data where the variable yield represents the return on investment. Often in such data, and with respect to the known features (i.e., economic indicators used to predict the value of an investment) both high (good) and low (bad) yielding data are indistinguishable. However, in such a case there might be small groups of investment possibilities in the training data that if invested in would always produce good consistent yields (and yet not necessarily the best). Thus, the goal of this work is to effectively mine these common data points.

[2] For more information on various approaches to mining data see [2].

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUL 2006		**00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Utilizing fused features to Mine Unknown Clusters in Training Data	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Undersea Warfare Center,Signal Processing Branch,Newport,RI,02841**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
9th International Conference on Information Fusion, 10-13 July 2006, Florence, Italy. Sponsored by the International Society of Information Fusion (ISIF), Aerospace & Electronic Systems Society (AES), IEEE, ONR, ONR Global, Selex - Sistemi Integrati, Finmeccanica, BAE Systems, TNO, AFOSR's European Office of Aerospace Research and Development, and the NATO Undersea Research Centre.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**7**	

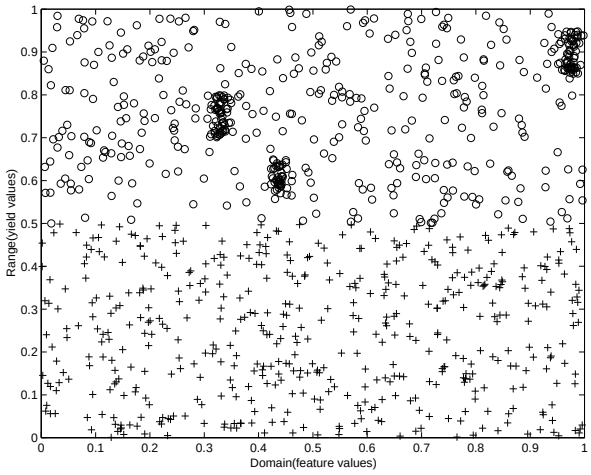

Figure 1: Illustrating the problem of interest with a straightforward example containing one thousand samples of one dimensional domain data (a single feature). In this figure, the ordinate that defines the yield of each data point is plotted versus the domain, where a yield value of 0.5 is used to separate and define the five hundred samples of the target class (i.e., yield > 0.5), and the five hundred samples of the nontarget class (yield < 0.5). It can clearly be seen that the two classes contain many commonly distributed points with respect to the range of the single feature: meaning that they are almost indistinguishable with respect to domain observations. However, notice that three clusters of data points also exist in the target class. Thus, the problem investigated here is to develop a classifier for this data that can essentially mine and recognize all of the points within the positive yielding clusters from all other data points contained in both classes.

strained to be located around the specified "center" yield value. In this case, the ordinate that defines the yield of each data point is plotted versus the domain, where a yield value of 0.5 is used to separate and define the five hundred samples of the target class (i.e., yield > 0.5), and the five hundred samples of the nontarget class (yield < 0.5).

It can clearly be seen in Figure 1 that the two classes contain many commonly distributed points with respect to the range of the single feature. In fact, it was shown in [8] for the single cluster case that traditional supervised classification approaches with this data produce nearly a 0.5 probability of error, and an overall average yield of just slightly more than 0.5. However, notice in Figure 1 that in this case three clusters of data points exist within the target class containing actual respective yields of 0.6, 0.75, and 0.9 (for two cluster results in Tables 1 and 2 below the 0.75 cluster is not used). In this case, each data cluster was randomly placed to be centered somewhere between the yield values of 0.5 and 1, where, as stated previously, the focus of this paper is on extending and refining the classifier developed in [7, 8] to mine multiple clusters in multi-dimensional feature space.

To demonstrate the effectiveness of the algorithm for the problem of Fig. 1, emphasis is placed on illustrating that the new method of mining for each unknown data cluster can obtain an error probability that is comparable to the classifier that knows the total number of clusters, and all data points within each cluster. Further, it will be of interest to show that this performance capability persists given the data contains both relevant and irrelevant features. In this case, the data used to demonstrate all performance results will be simulated based on a multi-dimensional extension of the example shown in Figure 1. This extension is useful because typical real-world problems often involve complicated multi-dimensional feature spaces, where determining the location of any hidden clusters by inspection is nearly impossible. In general, it will be shown that the automatic techniques developed here to find any number of data clusters can easily be applied in higher dimensional feature spaces. However, the drawback with increasing numbers of dimensions is that computational costs also go up.

3 Review of the basic approach to solving the problem

The approach taken in [7, 8] to solve this problem was based on an extension to Mean-Field Bayesian Data Reduction Algorithm (Mean-Field BDRA). The Mean-Field BDRA was developed to mitigate the effects of the curse of dimensionality by eliminating irrelevant feature information in the training data (i.e., lowering M), while simultaneously dealing with the missing feature information problem. The algorithm is based on the BDRA that was first introduced in Ref. [10], and which assigns an assumed uniform Dirichlet (completely noninformative) prior for the symbol probabilities of each class [4]. In other words, the Dirichlet is used to model the situation in which the true probabilistic structure of each class is unknown and has to be inferred from the training data. For more information on the Mean-Field BDRA algorithms see Appendix A of [7, 8], and notice that because of its superior performance with difficult unsupervised training situations the modified version of the Mean-Field BDRA will be used here. In this case, the Mean-Field BDRA is trained with a very fine initial quantization on the feature space (i.e., twenty initial thresholds per feature) to better determine final threshold values for locating each cluster.

The development of an algorithmic approach, or new training method, for this problem depended on adapting the Mean-Field BDRA to automatically sort and "mine" the data points contained in multiple unknown clusters. In this case, the built-in dimensionality reduction aspects of the Mean-Field BDRA are exploited for isolating each data cluster. Additionally, as the Mean-Field BDRA is a discrete classifier it naturally defines threshold points in the feature space that isolate the relative location of all clusters. Before proceeding with the algorithm developed for multiple clus-

ters, the approach to solving the single cluster case in [7, 8] is discussed next.

3.1 The single cluster case

In general, the automatic cluster mining algorithm developed in [7, 8] to locate a single data cluster strongly relies on the Mean-Field BDRA's training metric, $P(e)$ (as shown in Equation (2) of [7, 8]). The idea is that because the Mean-Field BDRA discretizes all multi-dimensional feature data into quantized cells any data points that are common to a cluster will share the same discrete cell, which also assumes that appropriately defined quantization thresholds have been determined by the Mean-Field BDRA. Therefore, given that all, or most, cluster data points can be quantized to share a common discretized cell they will all also share a common probability of error metric. In other words, locating an unknown cluster, and all of its data points, was based on developing a searching method that looks for data sharing a common $P(e)$. In this case, it is expected that this common error probability value, for all points within a cluster, will be relatively small with respect to that computed for most other data points outside of the cluster. This latter requirement should be satisfied in most situations as data clusters should tend to be distributed differently with respect to data outside of the cluster. As a final step in training, the validity of this cluster can be checked by computing the overall average yield for all points within the cluster (i.e., any grouped data points producing the largest average yield are chosen as appropriately mined data clusters).[3]

3.2 Extension to the multiple cluster case

To extend the idea described above to finding multiple unknown clusters, it was required for the algorithm to have the ability to intelligently sort through and separate data points having common error probabilities. In this case, both the total number of clusters and the number of samples per cluster are assumed unknown to the classifier. Therefore, with multiple data clusters each error probability value was thought of as an indicator to each point within each cluster. Typically, as in the single cluster case, it was expected that with multiple clusters all common error probability values, that is, for all points within each cluster, would be relatively small with respect to that computed for most other data points outside of any cluster. In general, the degree to which this latter requirement was satisfied depended on how differently the clusters tended to be distributed with respect to the non-clustered data.[4] Therefore, a proper data mining algorithm of multiple clusters, and one that is based on the Mean-Field BDRA, will tend to have a higher likelihood of finding leading cluster candidates by focusing on the largest groups of data points that cluster around smaller common error probability values. As the sorting, or mining, continues in this way any data points associated with small error probabilities and that have no other, or a small number of, common data points are rejected as cluster members. The algorithm was designed to automatically stop when all unknown data clusters were found, or when the training error begins to increase. Finally, and as in the single cluster case, the validity of each cluster with respect to the training data was checked by computing the overall average yield for all points within the cluster.

3.2.1 New algorithm training steps

The steps shown below were developed for training a new multiple cluster algorithm using the Mean-Field BDRA, that is, in such a way that all unknown data clusters can be identified with a minimum probability of error. For each of these steps training proceeds in a semi-unsupervised manner in that all target data (yield > 0.5) is utilized without class labels (i.e., no class information at all), and all nontarget data (yield < 0.5) is utilized with class labels (full class information). The motivation for training in this way is to force the Mean-Field BDRA to readily recognize the contrast between target cluster data points and all other data points in both classes that are not like the cluster. Therefore, when adapting class labels for the target class the Mean-field BDRA is more likely to label any cluster data points as target, while grouping most other noncluster "target" data points with the nontarget. The new method of training proceeds with the following steps.

1. Using all available training data (i.e., with all target points unlabeled and all nontarget points labeled), separately train the Mean-Field BDRA by incrementally varying the initial number of discrete levels to be between two and twenty levels.[5]

2. From the separate training runs in the previous step choose the initial number of discrete levels to use for each feature as that producing the least training error (see Equation (2), Appendix A of [7, 8]). Notice that the idea of steps 1 and 2 is to find the best initial number of discrete levels to use for each feature prior to looking for individual clusters. Typically, it is desired to train with as many initial levels as the data will support for best results.

3. Sequentially, label each target data point with the correct target label and separately re-train

[3] Typically, data outside of a cluster will have a more random distribution of error probability values that will not necessarily associate with a common yield value.

[4] Intuitively, as data within a cluster becomes distributed more like the data outside of the cluster it obviously becomes less distinguishable. On the other hand, it is reasonable then to conclude that if unknown clusters exist within a data set they will be distinguishable by being distributed differently with respect to all other data points outside of the clusters. Notice that this important assumption about the distribution of the clusters is what has been exploited in developing the methods presented here.

[5] In this case, for the results shown here "all available" training data means 50% of the entire data set.

the Mean-Field BDRA, where all remaining target data points are unlabeled as above.[6] This step produces a set of N_{target} computed training runs equal to the number of target training data points. For each separate run in this step compute a set of "cluster-training" errors by using the training data as a "test" set in which every data point, except for the single correctly labeled target point, is labeled a nontarget.[7]

4. From the set of N_{target} computed cluster-training errors in the previous step sort and group all data points according to those having common error values. The final list of separate cluster-training errors should proceed from the smallest to the largest, and for each value find all data points that share this same error. Observe that this step helps to reveal those data points that are sharing a similar region in quantized feature space.

5. Begin a cluster search and look for the first data cluster using the list obtained in step 4 above. To do this, choose as the best candidate, for *data cluster 1*, as the one having simultaneously the smallest cluster-training error and the largest number of common data points.[8] In this case, call the error associated with all points of this first cluster candidate $P(e|0)$.

6. After selecting the first cluster candidate in the previous step retrain the Mean-Field BDRA with all data points within this cluster labeled as the target class, and where the remaining target data points are unlabeled. Call this new cluster-training error that is computed simultaneously for all data points within *cluster 1* $P(e|1)$. The important point of this step is to determine how statistically similar the selected group of training data points are with each other, or, on the other hand, how different this group is with respect to the nontarget class (which now includes all other "target" data points outside of the cluster).

7. Compare $P(e|1)$ and $P(e|0)$ from steps five and six above. If $P(e|1) \leq P(e|0)$, as it should be in most cases containing data clusters, conclude that *cluster 1* is a valid first data cluster and proceed to step eight. Otherwise, conclude that no substantial data clusters exist, and terminate the algorithm.

8. Proceeding as in step five, proceed to the next most likely candidate on the list and select a second potential cluster (i.e., excluding all points in the first cluster). This new group of points will have simultaneously the next smallest cluster-training error and the largest number of common data points.

9. Retrain the Mean-Field BDRA with all data points within *cluster 2* (and *cluster 1* if selected above) labeled as the target class, and where the remaining target data points are unlabeled. Call this new cluster-training error that is computed simultaneously for all data points within these two clusters $P(e|2)$. Again, if $P(e|2) \leq P(e|1)$, conclude that all clusters in this group are valid data clusters and proceed to the next step. Otherwise, conclude that no more substantial data clusters exist, and terminate the algorithm.

10. Repeat the previous step sequentially for the c^{th} cluster and until all remaining potential clusters have been evaluated from the cluster list. It is important to note that this step always utilizes and trains with all previously determined clusters from the previous steps. As a final step to validate the clusters, compute the average yield for each cluster and, if applicable, select those producing the largest overall yield.

11. Finally, terminate the algorithm when $P(e|c) > P(e|c-1)$, meaning all potential clusters have been evaluated.

4 Review of previous results

The tables appearing in this section illustrate previously obtained (see [7]) performance results with the Mean-Field BDRA using one dimensional data of the type shown in Fig. 1. Before describing these results, the following list describes in more detail the items appearing in the tables below.

Supervised-Unclustered (Sup.-Unclust.) This represents the BDRA classifier that knows the true class labels of each point in the training data, and which is trained in the traditional supervised manner. In this case, and referring to Figure 1, training occurs with all data points above the yield threshold of 0.5 labeled as target, and all points below the threshold of 0.5 labeled as nontarget. In the analogy to financial data, this classifier is utilizing all of the data and is trying to learn how to predict investments that produce a good yield form those that do not.

Supervised-Clustered (Sup.-Clust.) This represents the BDRA classifier utilizing supervised training and that knows which data points are contained in clusters. In this case, only data points contained in clusters above the threshold of 0.5 in Figure 1 are labeled as target. Notice, that all data points above the 0.5 threshold that are not in a cluster, and every data point below this threshold, are labeled as nontarget. In

[6] In this important step of training the multi-dimensional fused features are reduced for each target data point, which determines the best subset of fused quantized features for that configuration of data.

[7] This error is computed based on counting the number of wrong decisions made under each hypothesis.

[8] Typically, the first error value on the list has both the absolute smallest error and the largest number of common points. However, because the algorithm is suboptimal this does not have to always be the case.

the financial data analogy, this classifier is trying to learn how to predict good consistent investments (i.e., those having similar statistical properties and cluster in feature space) from all other investments (i.e., both good and poor yielding investments that are indistinguishable in feature space).

Unsupervised-BDRA (Unsup.-BDRA) This represents the semi-unsupervised Mean-Field BDRA classifier that utilizes the eleven training steps listed above, and where all data points above the 0.5 yield threshold of Figure 1 are not assigned any class labels. Further, recall that all training data points below the 0.5 threshold are labeled as nontarget (i.e., poor yielding data). Returning to the financial data analogy, as in the previous case this classifier is trying to learn how to predict good consistent investments (i.e., those that cluster in feature space) from all other investments (i.e., both good and poor yielding investments that are indistinguishable). However, the difference in this case is that the Mean-Field BDRA has no prior knowledge about any clusters in the data. The goal is to adaptively mine each cluster and its associated location, that is, with respect to the feature space and yield values.

As a final note before describing results, observe that Figure 1 contains more than one cluster indicating that possibly more than one class exists in the data. In general, and in all results presented here, it is assumed that only two classes exist. That is, a target class represented by data above threshold and a nontarget class represented by data below threshold. However, a future extension of the methods developed in this paper will be to determine if performance can further be imported by assuming that individual data clusters are separate classes.

Table 1 (see the caption above) illustrates interesting aspects with regard to classifying data that contains isolated clusters. Observe in this table that average classification results are poor when all of the training data are labeled correctly, and training proceeds in a supervised manner (see the unclustered results column), given the classifier has no knowledge about any data clusters. However, it can also be seen (see the clustered results column) that performance improves dramatically when the classifier is given precise knowledge about the location of all points within the data clusters.

The error probabilities in Table 1 indicate that there is only a slight difference in the results if the data contains respectively either two or three clusters. For example, in the unclustered results column the three cluster case is slightly better as more clusters are providing information to help discriminate the classes (as a comparison to this, [7, 8] single cluster results using supervised training produced an error probability of near 0.5). On the other hand, in the clustered column the two cluster case appears to perform slightly better. In this situation, with three clusters an increasing number of isolated quantized cells also causes more false positive classifications to occur in the regions containing all clusters.

Table 1: Classification performance results for the Mean-Field BDRA (i.e., w/o a cluster mining algorithm applied) with supervised training (i.e., data with yields greater than 0.5 are called target and those with yields less than 0.5 are called nontarget), for two and three cluster data of the type shown in Fig. 1. Appearing in this table is the average probability of error computed on an independent test set (50% training/50% test), for the respective number of unknown clusters shown. In this case, supervised training results appear for both unclustered the classifier has no knowledge about the data clusters (i.e., see Sup.-Unclust. definition above), and the clustered classifier which knows all data points in each cluster that are labeled as target (i.e., see Sup.-Clust. definition above). In producing these results the Mean-Field BDRA trains with twenty initial discrete levels of quantization.

# of clusters	Sup.-Unclust.	Sup.-Clust.
2	0.400	0.104
3	0.388	0.126

As a final observation in Table 1, notice that the initial number of discrete levels per feature was chosen to be twenty by the Mean-Field BDRA. For the supervised training case shown in this table the initial number of discrete levels used for each feature was chosen to be consistent with that used below in obtaining the modified results of Table 2. In all cases, when obtaining these results the actual number of initial discrete levels per feature was incrementally varied between two and twenty by the Mean-Field BDRA. The final value of ten shown was determined by the Mean-Field BDRA to be that producing the smallest cluster-training error with the clustering algorithm applied.

Table 2: Classification performance results appear for the Mean-Field BDRA (i.e., with a cluster mining algorithm applied) and unsupervised training (i.e., using the algorithmic steps described above), for two and three cluster data of the type shown in Fig. 1. Appearing in this table is the average probability of error computed on an independent test set (50% training/50% test), for the respective number of unknown clusters shown. Notice, that for comparison the error probabilities are repeated for the supervised clustered case of Table 1.

# of clusters	Unsup.-BDRA	Sup.-Clust.
2	0.110	0.104
3	0.134	0.126

Observe that the utility of the data clustering algorithm developed here can clearly be seen in the results of Table 2. In this table, and observe for both the two and three cluster cases, that the error probability of the cluster mining algorithm is only about one percent higher than it is for the clustered supervised classifier that knows everything. This is significant because the cluster mining algorithm used here has no prior information at all about the clusters.

Table 3: Average yield results for the multiple cluster cases of Tables 1 and 2, and for comparison previously obtained single cluster results are also shown. In each of these cases, the actual average yield for all data clusters is 0.75. Appearing for two and three clusters are computed average yields for the unsupervised Mean-Field BDRA based classifier of Table 2, and the Supervised unclustered classifier of Table 1. For the single cluster case, yield values are based on averaging the one-dimensional results for actual cluster yields of 0.6 and 0.9.

# of clusters	Unsup.-BDRA	Sup.-Unclust.
1	0.666	0.512
2	0.608	0.555
3	0.622	0.588

In Table 3 it can be seen that the cluster mining algorithm developed here is improving the overall average yield for all numbers of clusters over that of the Supervised classifier. This implies that the new algorithm is improving the quality of the decisions in that it is declaring a proportionately larger ratio of high yielding data points as the target. However, notice also that as the number of clusters increases yield performance of the supervised classifier improves with respect to that of the unsupervised Mean-Field BDRA. Intuitively, as more clusters appear in the data classification performance with supervised training should improve as each cluster provides additional information. This implies that in some cases it might be best for an algorithm such as the Unsupervised Mean-Field BDRA to mine for clusters individually, as opposed to collectively as a group.

5 Extension for mining multi-dimensional data

In this section, results with multi-dimensional data are presented, where the one dimensional case shown in Figure 1 is extended in Figure 2 by utilizing two fused features to mine unknown data clusters.

The results in Table 4 indicate that the cluster mining algorithm developed here can be effectively extended to mine data clusters in multi-dimensional feature spaces. For example, it can be seen in this table that with respect to the performance metrics of the average probability of error and average yield, the Mean-Field BDRA outperforms the Supervised Classifier.[9] It is also apparent, and as expected, that relative to the one dimensional case shown above (see Tables 1, 2, and 3), performance has improved for both classifiers when utilizing two fused features in the training data. In other words, for the three cluster case results shown in Table 4 all error probabilities are now lower, and all associated yields higher than that previously shown above. This, of course, is a direct result of the additional information provided about the target class by having more than one relevant feature in the data.

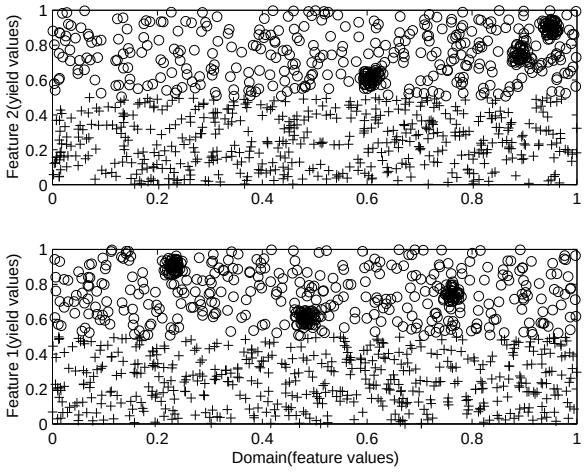

Figure 2: Illustrating the problem of interest by extending the example shown in Figure 1 to mining unknown data clusters in a two dimensional fused feature space. Notice, and as shown in Figure 1, the ordinate defining the yield of each data point is plotted versus the domain, where a yield value of 0.5 is used to separate and define the five hundred samples of the target class (i.e., yield > 0.5), and the five hundred samples of the nontarget class (yield < 0.5). However, in this case observe that separate plots also appear for each feature under both classes. It can clearly be seen that amongst the commonly distributed data three unique clusters exist in the target class for each dimension, where each cluster contains an equal number of points (i.e., they represent the same data points). Therefore, the goal for this problem is to extend previous results and develop a classifier that can mine and recognize all points within the two dimensional positive-yielding data clusters contained in the target class. In other words, the overall objective is to illustrate the impact that feature level fusion has on the performance of the algorithm developed here.

As a final observation, it should be pointed out that computational time substantially increased when mining clusters in two dimensions, that is, as compared to the one dimensional case. Thus, it can be expected that in higher dimensional spaces the computational costs will increase much more significantly when uti-

[9] Recall, the supervised classifier knows the correct class labels for each fused feature vector.

Table 4: Classification performance results using two dimensional data appear for the Mean-Field BDRA (i.e., with a cluster mining algorithm applied) and unsupervised training (i.e., using the algorithmic steps described above), for the three cluster data of the type shown in Fig. 2. As with previous tables, shown in this table is the average probability of error computed on independent test data (50% training/50% test), for the respective number of unknown clusters shown. Also, in parenthesis the average yield is given with each respective error probability. Notice, that for comparison results are also shown for the supervised unclustered classifier.

# of clusters	Unsup.-BDRA	Sup.-Unclust.
3	0.121(0.643)	0.308(0.612)

lizing the methods developed here. However, to significantly improve these computational costs it is possible to first independently mine and train each dimension separately. This is then followed by jointly training on the reduced joint quantized feature space. As an example, prior work in [9] demonstrated that this technique can not only reduce computational costs but can also improve classification performance. This will be a topic of future research for the methods contained in this paper.

6 Summary

In this paper, a previously developed data mining technique based on the Mean-Field Bayesian Data Reduction Algorithm (BDRA) has been extended to mine multiple unknown clusters in fused multi-dimensional feature spaces. The new method employs an intelligent search through the feature space by sorting and separating out data points having common error probabilities. In other words, the algorithm works by finding commonly grouped cluster data points that are in the same quantized region of the feature space. For the simulated data shown, finding all clusters was typically based on estimating and lowering the false decision rate of the training data, given all candidate points within each cluster are labeled as a target. In all cases, classification results revealed that the new clustering algorithm was able to find all significant clusters within the data. Further, and as expected, the algorithm was able to improve performance over the single dimensional case by utilizing the additional information contained in two relevant fused features.

Acknowledgments

This research was supported by a Grant from Dr. David Drumheller of the Office of Naval Research.

References

[1] R. A. Dempster, A. P. Laird, and S. G. Walker, "Maximum Likelihood from incomplete data via the EM algorithm (with discussion)," *J. R. Statistical Society*, B, 39, 1977, pp. 1-38.

[2] *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, NY, NY, 2001.

[3] S. Haykin, *Neural Networks: A Comprehensive Foundation, Second Edition*, Prentice Hall, Inc., Upper Saddle River, NJ, 1999.

[4] G. F. Hughes, "On the Mean Accuracy of Statistical Pattern Recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, January 1968, pp. 55-63.

[5] A. Jain and D. Zongker, "Feature Selection: Evaluation, Application, and Small Sample Size Performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, February 1997, pp. 153-158.

[6] J. Kittler, "Feature Set Search Algorithms," *Pattern Recognition and Signal Processing*, C. H. Chen, ed., Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands, 1978, pp. 41-60.

[7] R. S. Lynch, Jr. and P. K. Willett,"An Algorithmic Approach to Mining Unknown Clusters in Training Data" *Proceedings of the SPIE Symposium on Defense and Security*, Orlando, FL, April 2006.

[8] R. S. Lynch, Jr. and P. K. Willett, "Utilizing unsupervised learning to cluster data in the Bayesian Data Reduction Algorithm" *Proceedings of the SPIE Symposium on Defense and Security*, Orlando, FL, April 2005.

[9] R. S. Lynch, Jr. and P. K. Willett, "Quantizing features independently in the Bayesian Data Reduction Algorithm," Proceedings of the 2004 IEEE International Symposium on Systems, Man, and Cybernetics, The Hague, Netherlands, October 2004.

[10] R. S. Lynch, Jr. and P. K. Willett, "Bayesian Classification and Feature Reduction Using Uniform Dirichlet Priors," *IEEE Transactions on Systems, Man, and Cybernetics*, **vol. 33**, no. 3, June 2003.