

Local Dimensionality Reduction and Supervised Learning Within Natural Clusters for Biomedical Data Analysis

Mykola Pechenizkiy, Alexey Tsymbal, and Seppo Puuronen

Abstract—Inductive learning systems were successfully applied in a number of medical domains. Nevertheless, the effective use of these systems often requires data preprocessing before applying a learning algorithm. This is especially important for multidimensional heterogeneous data presented by a large number of features of different types. Dimensionality reduction (DR) is one commonly applied approach. The goal of this paper is to study the impact of natural clustering—clustering according to expert domain knowledge—on DR for supervised learning (SL) in the area of antibiotic resistance. We compare several data-mining strategies that apply DR by means of feature extraction or feature selection with subsequent SL on microbiological data. The results of our study show that local DR within natural clusters may result in better representation for SL in comparison with the global DR on the whole data.

Index Terms—Classification, dimensionality reduction (DR), local learning, supervised learning (SL).

I. INTRODUCTION

INDUCTIVE (supervised) learning systems were successfully applied in a number of medical domains, for example, in localization of a primary tumor, prognostics of recurrence of breast cancer, diagnosis of thyroid diseases, and rheumatology [1].

However, both researchers and practitioners realize that the effective use of these learning systems requires data preprocessing prior to model induction. This is especially important for multidimensional heterogeneous data presented by a large number of features of different types. The so-called “curse of dimensionality” [2] pertinent to many learning algorithms denotes the drastic raise of computational complexity and classification error with data having large number of features. Hence, the dimensionality of feature space is often reduced before supervised learning (SL) is undertaken.

There are a number of dimensionality reduction (DR) techniques, and according to the adopted reduction strategy, they are usually divided into feature selection (FS) and feature extraction (FE) approaches [3]. The key difference between FS and

FE is that, in the former, a subset of original features is selected while the latter is based on generation of a completely new feature space through a functional mapping, retaining in fewer dimensions as much information about the data as possible [3].

For some problem domains a feature subset may be useful in a subspace of the instance space, and at the same time, it may be useless or even misleading in another subspace. In such a situation, it might be difficult or even impossible to remove irrelevant and/or redundant features and leave only the useful ones using global FS. However, if it is possible to find local homogeneous subspaces of the instance space, then there might be better chances to apply FS successfully (individually within each subspace). For FE, the decision whether to proceed globally over the entire instance space or locally in different subspaces is also one of the key issues. It can be demonstrated that despite being globally high dimensional and sparse, data distributions in some subareas of the instance space are often low dimensional and dense, for example, in physical movement systems [4].

One possible approach for local FS or local FE is to identify clusters (groups) of instances forming homogeneous subspaces of the instance space. Generally, different clustering techniques can be used for this purpose, for example, the k -means or EM techniques [5]. However, in this paper, we emphasize the possibility of applying *natural clustering* that uses contextual features (constructed with the help of domain experts) to split the whole heterogeneous instance space into more homogeneous clusters.

In our experimental study, we apply k -nearest neighbor (k NN) classification to real clinical data to build antibiotic resistance (AR) prediction models. We apply the principle of natural clustering, grouping the instances into clusters related to certain pathogen types. We apply three different wrapper-based sequential FS techniques and three eigenvector-based FE techniques globally and locally and analyze their impact on the performance of k NN.

The results of our study show that natural clustering is an effective approach that facilitates local DR and that the proper selection of a DR technique within each natural cluster enables to get higher predictive accuracy (on average) in comparison with the global SL with or without DR.

The paper is organized as follows. In Section II, we consider the problem of nosocomial infections and AR, and describe data used in our experiments. In Section III, we briefly review related work on local DR and present our approach to the problem. In Section IV, DR techniques used in the experimental study are presented. In Section V, we present the results of experiments with the FE and FS techniques applied globally for

Manuscript received September 13, 2005; revised January 5, 2006. This work was supported in part by the COMAS Graduate School of the University of Jyväskylä, in part by the Academy of Finland, and in part by the Science Foundation Ireland under Grant S.F.I.-02IN.11111.

M. Pechenizkiy and S. Puuronen are with the Department of Computer Science and Information Systems, University of Jyväskylä, 40351 Jyväskylä, Finland (e-mail: mpechen@cs.jyu.fi; sepi@cs.jyu.fi).

A. Tsymbal is with the Department of Computer Science, Trinity College Dublin 2, Ireland (e-mail: Alexey.Tsymbal@cs.tcd.ie).

Digital Object Identifier 10.1109/TITB.2006.875654

the whole data set and locally in clusters for further SL. Finally, in Section VI, we briefly conclude with a summary and present directions for further research.

II. DOMAIN: NOSOCOMIAL INFECTIONS

A. Nosocomial Infections

Nosocomial infections and AR are highly important problems today that impact the morbidity and mortality of hospitalized patients as well as their cost of care. It is known that 3%–40% of patients admitted to hospital acquire an infection during their stay, and that the risk for hospital-acquired infection, or nosocomial infection, has risen steadily in recent decades. Formally, nosocomial infections are defined as infections arising after 48 h of hospital admission. Infections arising earlier are assumed to be arisen prior to admission, although this is not always true [6]. The frequency of nosocomial infections depends mostly on the type of conducted operation being greater for so-called “dirty” operations (10%–40%), and smaller for “pure” operations (3%–7%). For example, such a serious infectious disease as meningitis is often the result of nosocomial infection.

Analysis of microbiological data included in antibiograms collected in different institutions over different periods of time is considered as one of the most important activities to restrain the spreading of AR and to avoid the negative consequences of this phenomenon. Traditional hospital infection control surveillance and localization of hospital infection often relies on the manual review of suspected cases of nosocomial infections and the tabulation of basic summary statistics that requires considerable time and resources, and the produced figures and patterns are often not up-to-date. Advanced computer-based analysis methods might help to discover more potentially useful patterns more efficiently [7].

It has been widely recognized that sophisticated, active, and timely intra-hospital surveillance is needed. Computer-assisted infection control surveillance research has focused on identifying high-risk patients, the use of expert systems to identify possible cases of nosocomial infection, and the detection of deviations in the occurrence of predefined events [7].

Treatment of nosocomial infections starts normally with a microbiological investigation. In this investigation, pathogens are isolated and for each isolated bacterium, an antibiogram is built (it represents bacterium’s resistance to a series of antibiotics). The user of the test system can define the set of antibiotics used to test bacterial resistance. The result of the test is presented as an antibiogram that is a vector of couples (antibiotic/resistance). The information included in this antibiogram is used to prescribe an antibiotic with a desired level of resistance for the isolated pathogen.

The antibiogram is not uniquely identified given a bacterium species, but it can vary significantly for the bacteria of the same species. This is due to the fact that the same bacteria of the same species may have evolved differently and have developed different resistances to antibiotics. However, very often groups of antibiotics have similar resistance when tested on a given bacterium species, despite its strains [8].

AR is an especially difficult problem for nosocomial infections in hospitals because they attack critically ill patients who are more vulnerable to infections than the general population and, therefore, require more antibiotics. Heavy use of antibiotics in these patients hastens the mutations in bacteria that bring about drug resistance [9]. According to the Centers for Disease Control and Prevention (CDC) statistics, more than 70% of bacteria that cause hospital-acquired infections are resistant to at least one of the antibiotics most commonly used to treat infections. Persons infected with drug-resistant organisms are more likely to have longer hospital stays and require treatment with second or third choice drugs that may be less effective, more toxic, and more expensive [9]. In short, antimicrobial resistance drives up health care costs, increases the severity of disease, and increases the death rates of some infections.

B. Source, Nature, and Organization of Data

The data used in our analysis were collected in the Hospital of N.N. Burdenko Institute of Neurosurgery using the “Vitek-60” analyzer (developed by *bioMérieux*) over the years 1997–2004 and information systems “Microbiologist” (developed by the Medical Informatics Laboratory of the institute) and “Microbe” (developed by the Russian company “MedProject-3”).

Each instance of the data used in the analysis corresponds to one microbiological test in the database (*sensitivity test*) related to a single specimen that may be blood, liquor, urine, etc. In this paper, we focus on the analysis of meningitis cases only, and the specimen is liquor. The results of the sensitivity tests (sensitive *S*, resistant *R*, or intermediate *I*) were obtained according to the guidelines of the National Committee for Clinical Laboratory Standards (NCCLS) [10].

The information about sensitivity analysis is related to a *patient*, his/her demographical data (*sex*, *age*), and hospitalization in the Institute (*main department*, whether the test was taken while the patient was in *ICU*, *days spent* in the hospital before, etc.).

For the purposes of this analysis, we picked up all 4430 instances of sensitivity tests related to the meningitis cases of the period of January 2002–July 2004. AR domain experts introduced a grouping of binary features for pathogens and antibiotics so that 17 pathogens and 39 antibiotics were combined into 6 and 15 groups, respectively.

Thus, each instance in our data has 30 features (beside the ID-like attributes for records, patients, antibiotics, and pathogens) that included information corresponding to a single sensitivity test augmented with data concerning the type of the antibiotic used and the isolated pathogen, clinical features of the patient and his/her demographics, and the microbiology test result as the class attribute (Table I).

In our previous pilot many-sided analysis of this data, we applied a number of different DM techniques trying to build an accurate predictive model, to explore and understand our data and find valuable association rules [11]. Our results showed that applying SL locally could lead to significantly higher accuracy for NB and C4.5 classifiers. However, the best accuracy results

TABLE I
DATA SET CHARACTERISTICS

Name	Type
Patient and hospitalization related	
sex	{Male, Female}
age	[0;72], mean 29.8
recurring stay	{True,False}
days of stay in NSI before test	[0;317], mean 87.5
days of stay in ICU	[0;237], mean 34
days of stay in NSI before specimen was received	[0;169] mean 31.6
bacterium is isolated when patient is in ICU	{True,False}
main department	{0,...,9}
department of stay	{0,...,11}
Pathogen and pathogen groups	
pathogen group1	{True,False}
...	...
pathogen group6	{True,False}
Antibiotic and antibiotic groups	
antibiotic group1	{True,False}
...	...
antibiotic group15	{True,False}
sensitivity	{Sensitive, Intermediate, Resistant}

Major data set characteristics: 4430 instances, 3 classes, 30 features: 4 numerical, 2 categorical and 24 binary. After binarization of categorical features we have got 48 features, of which 4 are numerical and 44 – binary.

with k NN were achieved globally. In this paper, we are interested in applying DR locally in each cluster before SL.

III. LOCAL DR

A. Related Work

Generally, there are several motivating reasons for local DR. For FS, it may be difficult or even impossible to remove irrelevant and/or redundant features from a heterogeneous data set where a feature subset is useful in one subspace of the instance space, and at the same time is useless or misleading in another one. However, if it is possible to identify local homogeneous subspaces, then it might be easier to apply FS successfully. Local FS has been under active study in the ensemble context [12] and in text mining applications [13].

In unsupervised learning, local DR has been actively studied in the context of co-clustering, bi-clustering, projected clustering (CLIQUE and ProClus algorithms), and subspace clustering (see for example [14]).

Also, in FE, a set of local (usually linear) FE models is seen as an alternative to a global (more complex) nonlinear approximation. The major assumption is that while the whole data may require nonlinear FE globally, its individual partitions may be represented with simpler linear models. Then, it is possible to construct a set (an ensemble) of local linear models instead of a global nonlinear one. However, this requires an extra effort in identifying the clusters.

Additional motivating reasons for local FE are the assumptions that a data set may actually consist of separate partitions, which may or may not be connected together as one “piece” and that the intrinsic dimensionality of the data set may vary from one partition to another.

Main well-studied approaches for local principal component analysis (PCA)-based FE include vector quantization local PCA

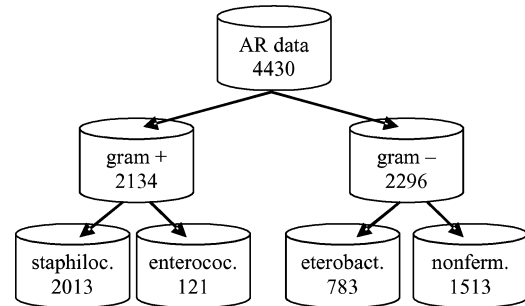


Fig. 1. Natural clustering of data with regard to the pathogens.

(VQPCA) [15], probabilistic PCA mixture model [16], different local Hebbian learning and local linear embedding (LLE) [17] algorithms.

In this paper, we apply (for local DR and SL) a clustering approach that utilizes domain expert knowledge to divide data into clusters where patterns of AR may have different nature. We describe this approach in Section III-B.

B. Natural Clustering Approach

We apply the so-called *natural clustering*, i.e., clustering based on the knowledge of domain experts. The main reason is our belief (supported by the expert opinion and pilot exploratory many-sided data analysis [11]) that the patterns of AR may have different nature and, therefore, can be different for different natural clusters. In particular, we are interested in how the accuracy of classification models varies from one cluster to another and whether it is possible to achieve better accuracy applying DR and a classifier locally in each cluster instead of the use of global DR and classification.

The data for our analysis are relatively high dimensional and heterogeneous; heterogeneity is presented by a number of contextual (environmental) features. Semantically, the *sensitivity* concept is related first of all to the *pathogen* and *antibiotic* concepts. In our study, binary features that describe the hierarchical grouping of pathogens (suggested by domain experts) were used for the hierarchical natural clustering. In our database, the whole data set can be divided into two natural clusters: *gram+* and *gram-* (Fig. 1). Then, the *gram+* cluster consists of the *staphylococcus* and *enterococcus* clusters, and *gram-* cluster consists of the *enterobacteria* and *nonfermentes* clusters.

We can see from the figure that the clusters *gram-* and *gram+* are approximately of the same size. The further division of the cluster *gram-* results in less balanced clusters. The further division of the cluster *gram+* results in highly unbalanced clusters in terms of the different numbers of instances in them.

In our many-sided exploratory analysis of these data [11], we applied several classifiers to the whole data set and individually to the *gram-* and *gram+* clusters. The results of our experiments demonstrated that, for example, for the Naïve Bayes and C4.5 classifiers, the differences of accuracies between the clusters were quite big and at the same time the average accuracy over the clusters is always higher than the corresponding global accuracy achieved with the whole data set. However, with k NN

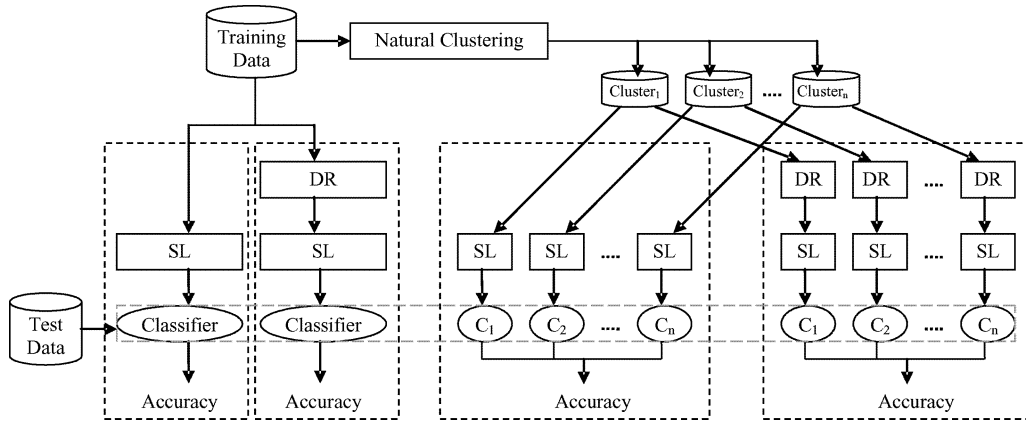


Fig. 2. Experimental approach.

the difference between the average of local accuracies and global accuracy was statistically insignificant for the division into the *gram-* and *gram+* clusters.

We also applied the same classification algorithms for one antibiotic group consisting of three clusters. And the experimental results demonstrated that the average accuracy of classifiers (except C4.5) over clusters (i.e., when they are applied within each cluster) is higher than the accuracies of corresponding global classifiers.

IV. DR TECHNIQUES USED IN THE PAPER

A. FE Techniques

PCA is one of the most commonly used FE techniques. It is based on extracting the axes on which data show the highest variability [18]. Although PCA “spreads out” the data in the new basis (new extracted axes), and can be of great help in unsupervised learning, there is no guarantee that the new axes are consistent with the discriminatory features in a classification problem.

Another approach is to account for class information in FE. One approach is to use some class separability criterion (for example, from Fisher’s linear discriminant analysis), based on a family of functions of scatter matrices: the within-class covariance, the between-class covariance, and the total covariance matrices. Parametric and nonparametric eigenvector-based approaches use the within- and between-class covariance matrices (thus, taking into account class information) and use the simultaneous diagonalization algorithm to optimize the relation between the within- and between-class covariance matrices. The difference between the approaches is in calculation of the between-class covariance matrix. The parametric approach accounts for one mean per class and one total mean, and therefore, may extract at most *number_of_classes-1* features. The nonparametric method tries to increase the number of degrees of freedom in the between-class covariance matrix, measuring the between-class covariances on a local basis. Our previous experiments with parametric and nonparametric FE show that the nonparametric FE is often more robust to different data set characteristics and often results in higher classification accuracy compared to the parametric FE [19].

B. FS Techniques

Greedy hill climbing is one of the simplest search strategies that makes sequential changes to a current feature subset [20]. Often, it is just the addition or deletion of a single feature to/from the subset at a time. We selected the most commonly used sequential FS strategies [21]: forward FS (FFS), backward feature elimination (BFE), and bidirectional search (BS). FFS starts with an empty subset of features and successively adds features one by one. On the contrary, BFE starts with the full subset features and step-wisely deletes features one by one. The BS proceeds in the both forward and backward directions in turn.

Search algorithms that implement these strategies may use different selection heuristics, such as considering all possible changes of one feature in the current subset and then selecting the best, or simply choose the first change that improves the merit of the current feature subset. In either case, once a change is accepted, it is never reconsidered (hence, the name “greedy”). There are different stopping rules for greedy FS such as stopping when no feature added to/deleted from the subset improves performance or when any feature added to/deleted from the subset degrades performance. The evaluation of selected feature subsets in our paper is based on the *wrapper* paradigm that assumes interaction between FS and classification [5].

V. EXPERIMENTAL STUDY

A. Experiment Design

In Fig. 2, we present the main idea behind our experimental setup. Our main aim is to analyze whether natural clustering is an efficient approach for the construction of local DR and SL models, i.e., whether FE and FS within natural clusters are able to produce better representation spaces for further SL in comparison with global DR. For this purpose, we collected the accuracies for classification with and without DR on the whole data set, and on each natural cluster.

In our experimental studies, we used an instance-based classifier (*k*NN), and the FFS, BFE, and BS techniques available in the machine learning library “WEKA 3.4.2” [20]. We used the conventional PCA and the class-conditional parametric (*Par*) and nonparametric (*NPar*) FE techniques [13], which we

TABLE II
ACCURACY OF k NN CLASSIFIER FOR EACH APPROACH AND NUMBER OF FEATURES USED BY CORRESPONDING APPROACH FOR CLASSIFICATION ON AVERAGE

Data set	Inst	Feature Selection						noFS	Feature Extraction						noFE		
		FFS		BFE		BS			PCA		Par		NPar				
global	4430	.742	8	.744	8	.738	8	.748	28	.696	34	.682	1	.734	39.1	.719	44
gram- (g-)	2296	.706	6	.709	6	.713	6	.706	24	.662	31	.622	1	.678	31.5	.685	35
gram+ (g+)	2134	.787	5	.784	5	.798	5	.788	24	.745	19	.752	1	.749	32.6	.738	35
enterobacteria	783	.677	4	.677	4	.679	4	.677	23	.635	16	.612	1	.644	28	.643	31
nonfermentors	1513	.716	7	.72	8	.72	9	.716	23	.680	30	.635	1	.700	26.8	.709	31
staphylococcus	2013	.799	5	.757	5	.756	5	.799	23	.766	20	.785	1	.754	33.5	.772	34
enterococcus	121	.736	3	.719	3	.727	4	.736	23	.658	11	.608	1	.631	21.8	.603	28
best	4430	.730	5	.731	5	.733	5	.749	24	.709	19	.696	1	.711	33	.722	34

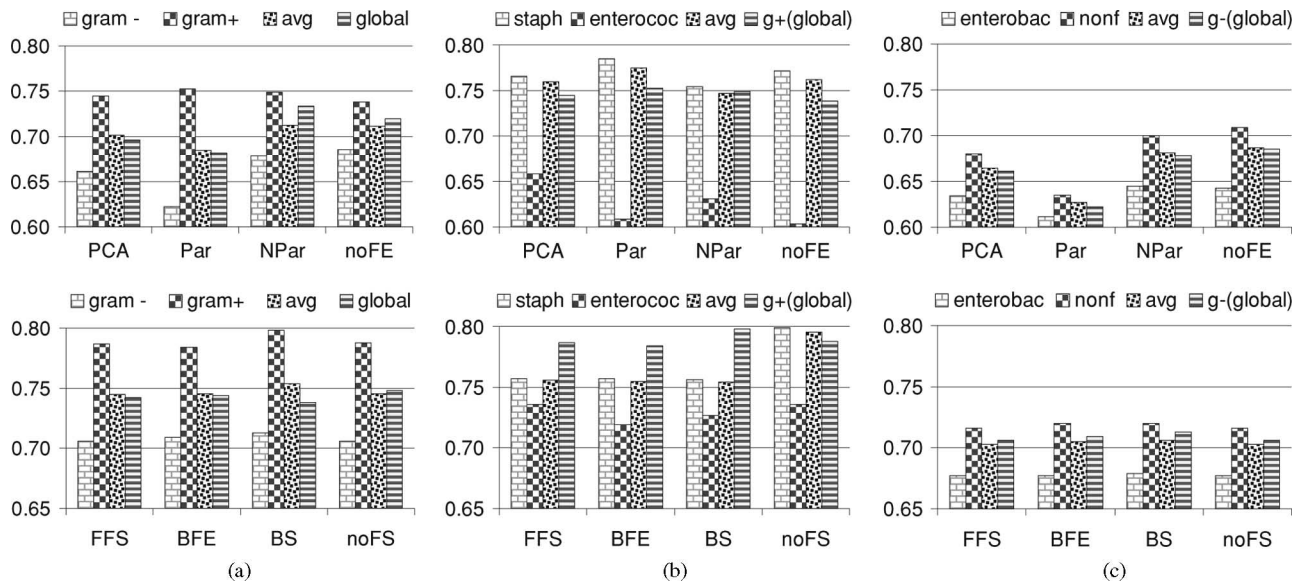


Fig. 3. Accuracy results of k NN classifier with different approaches.

implemented within the same library. We used $k = 7$ and the inverse distance for k NN since these parameters were found to be the best combination for our data (when applied globally) in our pilot studies [11]. The threshold for variance covered was set to 95% for each FE technique and, for $NPar$, we used parameter settings $k = 7$ and $\alpha = 1$.

B. Experimental Results

The main results of experiments are given in Table II. Each row of the table contains the name of the data set (cluster), number of instances in it, accuracy of k NN, and the number of features used for each FS strategy (FFS, BFE, and BS) and FE technique (PCA, Par , $NPar$) averaged over 30 test runs of Monte-Carlo cross-validation. Column $noFS$ is related to the case when no FS was used and column $noFE$ to the case when no FE was used but all the categorical features were binarized, i.e., each categorical feature was replaced with a redundant set of binary features, each corresponding to one value of the original feature.

In Fig. 3, comparison of local and global results of k NN for seven different clusters (including the whole data set) are shown.

The results show similar behavior of FS and FE across the seven different clusters. From the histograms, one can see that DR for our data results in the best accuracy when applied locally to *staphylococcus*, *enterococcus*, and *gram-* clusters.

Applying k NN with DR locally in the *gram+* and *gram-* clusters [see Fig. 3(a)] does not outperform global accuracy. However, the accuracy results for cluster *gram+* are much higher than for the *gram-* cluster. The FE methods were almost equally good for the *gram+* cluster. For the *gram-* cluster, Par was the worst and $NPar$ was the best.

But still, k NN without any FE performed slightly better for *gram-*. FS had no effect on k NN accuracy for the *gram-* cluster. And BS was the only FS technique that increased k NN accuracy for *gram+*.

In Fig. 3(b), one can see that applying k NN and DR individually to the *staphylococcus* and *enterococcus* clusters significantly (according to the Student t -test with 0.95 level of significance) increases overall accuracy. Local Par outperforms local k NN without FE by 1.3% (avg Par versus avg $noFE$), $NPar$ decreases the accuracy of k NN without FE by 2%, and PCA has no effect. Local FS decreases the performance of k NN without

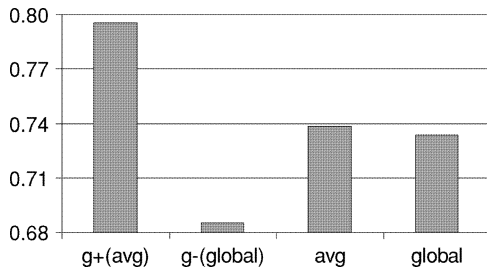


Fig. 4. Classification accuracy with the selection of the best DR technique within each corresponding cluster.

FS by 3.5%–4.5%. Relatively low accuracy of the classifier for the *enterococcus* cluster does not decrease much the average accuracy since this cluster is rather small and contains only 5.7% of instances from *gram+* while *staphylococcus* contains 94.3%.

However, the analysis of how good or bad the performance of certain local model is helps to understand which subsets of instances are harder to classify or which subsets of instances are very noisy.

In Fig. 3(c) one can see that dividing the cluster *gram-* further into *enterobacteria* and *nonfermentes* does not increase the accuracy of *k*NN both with and without local FE or FS.

Now, if we compare the FE and FS horizontal triples of histograms, we can see that for our data the sequential strategies for FS have no success. One exception is the success of BS when applied individually to the *gram+* and *gram-* clusters. The FE techniques have more diverse behavior. In particular, PCA is the best for the *enterococcus* cluster while *Par* is the best for *staphylococcus* [Fig. 3(b) top]. *NPar* demonstrated the best accuracy for global FE on the whole data set. This leads to an idea of adaptive selection of FE method for each cluster—the use of PCA in one cluster, and *Par* or *NPar* in some other cluster may result in significantly higher overall accuracy. Fig. 4 shows the classification accuracies for our data with regard to the selection of the best DR techniques within corresponding clusters—locally in *gram+* and globally in *gram-*.

We compare the impact of FS and FE on classification either globally or locally with the most appropriate selection of clusters: *staphylococcus*, *enterococcus* (joined into averaged results of *gram+*), and *gram-*.

Due to space limitations, we do not present a separate figure but list the main conclusions of the comparison: 1) For the analyzed data set the natural clustering approach was helpful in construction of local FE models, but was not useful for local FS; 2) FE was useful (it improved generalization accuracy) when applied both globally and locally, while FS increased the accuracy of *k*NN only when applied locally to the *gram+* and *gram-* clusters; 3) FS applied locally on these data resulted in higher accuracy produced by *k*NN comparing to local FE. However, we need to point out that this was due to the binarization of categorical features (that is required for FE) that leads to increase in redundant binary features. *k*NN produced almost 3% higher accuracy results for data presented by original categorical (not binarized) features. Perhaps, by analyzing possible reasons of why the accuracy of *k*NN on these data decreases

after binarization, we can improve the overall situation in the FS–FE competition; 4) *Par* produced very poor results when applied globally, but performed surprisingly well in some of the clusters. *NPar* was quite stable across different clusters, and it was the best FE technique for the situation when DR was applied globally.

VI. CONCLUSION AND FUTURE DIRECTIONS

DR is an effective approach to data reduction aimed at focusing on relevant features and improving the quality of data representation. We experimentally compared and showed the benefits of local and global DR (by means of FS and FE) for subsequent SL. In this paper, we applied the *natural clustering* approach using contextual features (constructed with the help of domain experts) to split a real-world clinical data set into more homogeneous clusters to use DR and SL locally.

The results of our experiments demonstrate that the proper selection of a local DR technique can lead to the increase of predictive accuracy in comparison with the global classification with or without DR. The amount of features extracted or selected is always smaller for local than global DR. This shows the usefulness of natural clustering in coping with data heterogeneity and high dimensionality. An interesting direction for further study is to analyze whether the feature subsets extracted are very different from one cluster to another and whether the local predictive models could be easier to examine and interpret for domain experts.

Our future research efforts are going to be directed toward the comparison of a mixture of FE models for SL built on natural clusters and on clusters produced by traditional clustering techniques. We analyzed spatial contextual features related to the categorization of different pathogens. We believe that clustering according to features that contain implicit or explicit information about the timestamp of certain instance may give interesting results too.

Applying PCA in one cluster, and *Par* or *NPar* in some other cluster may result in higher overall accuracy. Therefore, another challenging goal is the adaptive selection of FE method for each cluster according to certain characteristics of the cluster.

ACKNOWLEDGMENT

The authors would like to thank Dr. M. Shifrin and Dr. I. Alexandrova from N. N. Burdenko Institute of Neurosurgery, Russian Academy of Medical Sciences, Moscow, Russia, for the provided data set and helpful discussions. They also thank WEKA developers for the source code used in their work. Last but not least, the authors would like to thank the anonymous reviewers for their valuable comments, which helped improve the quality of the paper.

REFERENCES

- [1] I. Kononenko, "Inductive and Bayesian learning in medical diagnosis," *Appl. Artif. Intell.*, vol. 7, no. 4, pp. 317–337, 1993.
- [2] R. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton, NJ: Princeton Univ. Press, 1961.
- [3] H. Liu and H. Motoda, Eds., *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Norwell, MA: Kluwer, 1998.

- [4] S. Vijayakumar and S. Schaal, "Local dimensionality reduction for locally weighted learning," in *Proc. IEEE Int. Symp. Comp. Intell. Robot. Autom.*, 1997, pp. 220–225.
- [5] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools With Java Implementations*. San Francisco, CA: Morgan Kaufmann, 2000.
- [6] R. P. Gaynes, "Surveillance of nosocomial infections: A fundamental ingredient for quality," *Inf. Control Hosp. Epidemiol.*, vol. 18, no. 7, pp. 475–478, 1997.
- [7] S. Brossette, A. P. Sprague, W. T. Jones, and S. A. Moser, "A data mining system for infection control surveillance," *Methods Inf. Med.*, vol. 39, no. 4–5, pp. 303–310, 2000.
- [8] E. Lamma, M. Manservigi, P. Mello, A. Nanetti, F. Riguzzi, and S. Storari, "The automatic discovery of alarm rules for the validation of microbiological data," in *Proc. 6th Int. Workshop IDAMAP 01*, 2001, pp. 1–7.
- [9] *The Problem of Antibiotic Resistance, NIAID Fact Sheet*, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, U.S. Department of Health and Human Services, USA [Online]. Available: <http://www.niaid.nih.gov/factsheets/antimicro.htm>
- [10] *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically: Approved Standard: Sixth Edition and Performance Standards for Antimicrobial Susceptibility testing*, NCCLS, Wayne, PA, 2004.
- [11] M. Pechenizkiy, A. Tsymbal, S. Puuronen, M. Shifrin, and I. Alexandrova, "Knowledge discovery from microbiology data: Many-sided analysis of antibiotic resistance in nosocomial infections," in *Professional Knowledge Management: Third Biennial Conference, WM 2005, Kaiserslautern, Germany, Apr. 10–13, 2005, Revised Selected Papers*, K.-D. Althoff, A. Dengel, R. Bergmann, M. Nick, and T. Roth-Berghofer, Eds. Berlin, Germany: Springer-Verlag, 2005, pp. 360–372.
- [12] S. Puuronen and A. Tsymbal, "Local feature selection with dynamic integration of classifiers," *Fundamenta Informaticae*, vol. 47, no. 1–2, pp. 91–117, 2001.
- [13] B. How and W. Kiong, "An examination of feature selection frameworks in text categorization," in *Proc. Inf. Retrieval Technol., AIRS'05*, LNCS, G. Lee *et al.*, Ed. New York: Springer-Verlag, 2005, vol. 3689, pp. 558–564.
- [14] C. C. Aggarwal, J. Han, J. Wang, and P. Yu, "A framework for high dimensional projected clustering of data streams," in *Proc. VLDB Conf.*, 2004, pp. 852–863.
- [15] N. Kambhatla and T. K. Leen, "Dimension reduction by local principal component analysis," *Neural Comput.*, vol. 9, pp. 1493–1516, 1997.
- [16] M. E. Tipping and C. M. Bishop, "Mixtures of principal component analysers," in *Proc. IEEE 5th Int. Conf. Artif. Neural Netw.*, 1997, pp. 13–18.
- [17] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [18] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.
- [19] M. Pechenizkiy, "Impact of the feature extraction on the performance of a classifier: kNN, Naive Bayes and C4.5," in *Proc. 18th CSCSI Conf. AI'05*, LNAI 3501, B. Kegl and G. Lapalme, Eds. New York: Springer-Verlag, 2005, pp. 268–279.
- [20] R. Caruana and D. Freitag, "Greedy attribute selection," in *Proc. ICML94*, San Francisco, CA: Morgan Kaufmann, 1994, pp. 26–28.
- [21] D. W. Aha and R. Bankert, "A comparative evaluation of sequential feature selection algorithms," in *Proc. 5th Int. Workshop Artif. Intell. Stat.*, D. Fisher and H. Lenz, Eds., 1995, pp. 1–7.

Mykola Pechenizkiy received the Ph.D. degree in computer science from the University of Jyväskylä, Jyväskylä, Finland, in 2005.

He was a Software Developer and Researcher at the Niilo Mäki Institute. Since 2002, he has been a Researcher in the Department of Computer Science and Information Systems, University of Jyväskylä. He was also a Visiting Researcher at the Northern Ireland Engineering Laboratory and at the Machine Learning Group, Trinity College, Dublin, Ireland. His research interests include machine learning, data mining and knowledge discovery from databases, and he has about 20 refereed publications in these areas since 2002.

Alexey Tsymbal received the Ph.D. degree in computer science from the University of Jyväskylä, Jyväskylä, Finland, in 2002.

He is currently a Research Fellow in the Department of Computer Science, Trinity College, Dublin, Ireland. His research interests include machine learning and data mining, knowledge-based systems, and applications of AI techniques to biomedical domains. He has more than 40 refereed publications in these areas.

Dr. Tsymbal was a General Co-chair of the 18th IEEE International Symposium on Computer-Based Medical Systems CBMS'2005 (Dublin, Ireland) and is currently the Special Tracks Chair for CBMS'2006 (Salt Lake City, UT).

Seppo Puuronen received the Ph.D. (Economics) degree in computer science from the University of Jyväskylä, Jyväskylä, Finland, in 1988.

He was as Associate Professor and Lecturer of Computer Science at the Lappeenranta University of Technology from 1983 to 1988. During 1988–2000, he was a Lecturer, Senior Researcher, and Acting Full Professor, Assistant Professor, and Associate Professor at the University of Jyväskylä. Since 2000, he has been a Full Professor of Information Systems at the University of Jyväskylä. He was the Head of the Department of Computer Science and Information Systems, University of Jyväskylä, during 1988–1989, 1990–1991, and 1997, and the Head of the COMAS Graduate School during 1999–2001. He has written several international journal and conference papers. His main research interests are data mining, knowledge discovery, and computers in education.