Data Mining with Combined Use of Optimization Techniques and Self-Organizing Maps for Improving Risk Grouping Rules: Application to Prostate Cancer Patients

**LEONID CHURILOV, ADYL BAGIROV, DANIEL SCHWARTZ,
KATE SMITH, AND MICHAEL DALLY**

LEONID CHURILOV received a B.Sc. (Hons1) in Mathematics and Statistics and a Ph.D. in Operations Research from the Department of Mathematics and Statistics, University of Melbourne, Australia. He is currently a Senior Lecturer in the School of Business Systems at Monash University, Melbourne. His research interests include health-care systems, integrated process modeling techniques, and the interfaces between process modeling and decision modeling for complex systems.

ADIL BAGIROV received his M.Sc., Baku State University, Azerbaijan, and a Candidate of Sciences, Institute of Cybernetics, National Academy of Sciences, Azerbaijan. He later received a Ph.D. in Optimization from the University of Ballarat, Australia. He is currently a Postdoctoral Research Fellow, School of Information Technology and Mathematical Sciences, University of Ballarat, Australia. His research interests include computational optimization and data mining.

DANIEL SCHWARTZ was a visiting research student at the School of Business Systems, Monash University, Australia.

KATE SMITH is an Associate Professor and Deputy Head in the School of Business Systems at Monash University, Australia. She holds a B.Sc. (Hons) in Mathematics and a Ph.D. in Electrical Engineering, both from the University of Melbourne, Australia. Dr. Smith has published over 130 refereed journal and international conference papers in the areas of neural networks, combinatorial optimization, and data mining, and regularly acts as a consultant to industry in the area of data mining.

MICHAEL DALLY gained his college fellowship in 1993 after training in both Newcastle, NSW, and Wellington, New Zealand, after graduating from Otago Medical School, New Zealand. He has worked as a Radiation Oncologist at the William Buckland Radiotherapy Centre, Alfred Hospital, Victoria, for the past ten years and has been supervisor of registrar training for the past six years. He is also responsible for the safe provision of stereotactic radiosurgery services to the director of Radiation Oncology. This commitment is mirrored in his clinical interests centered around prostate cancer and brain tumors. He is the current chair of the brain tumor group of VCOG under the auspices of the Cancer Council of Victoria.

ABSTRACT: Data mining techniques provide a popular and powerful tool set to generate various data-driven classification systems. In this paper, we investigate the combined use of self-organizing maps (SOM) and nonsmooth nonconvex optimization techniques in order to produce a working case of a data-driven risk classification system. The optimization approach strengthens the validity of SOM results, and the improved classification system increases both the quality of prediction and the homogeneity within the risk groups. Accurate classification of prostate cancer patients into risk groups is important to assist in the identification of appropriate treatment paths. We start with the existing rules and aim to improve classification accuracy by identifying inconsistencies utilizing self-organizing maps as a data visualization tool. Then, we progress to the study of assigning prostate cancer patients into homogenous groups with the aim to support future clinical treatment decisions. Using the case of prostate cancer patients grouping, we demonstrate strong potential of data-driven risk classification schemes for addressing the risk grouping issues in more general organizational settings.

KEY WORDS AND PHRASES: classification, data mining, optimization, prostate cancer, risk grouping.

DATA MINING TECHNIQUES HAVE SHOWN great promise in the area of clinical decision support, and in particular, cancer diagnosis and prognostics. For example, Mangasarian et al. [15] discuss the applications of linear programming to the problem of clinical classification of patients with breast cancer. Bellazi et al. [5] present the use of data mining tools to derive a prognostic model of the outcome of resectable hepatocellular carcinoma. Land et al. [12] discuss a new neural network technology developed to improve the diagnosis of breast cancer using mammogram findings, whereas Walter and Mohan [17] describe *ClaDia*, a learning classifier system applied to the well-studied Wisconsin breast cancer data set. As far as prostate cancer is concerned, Zupan et al. [19] propose a schema that enables the use of classification methods, including machine learning classifiers, for survival analysis of prostate cancer patients, whereas Zhang and Zhang [18] develop and validate *ProstAsure*—a neural network-derived algorithm that analyzes the profile of multiple serum tumor markers and produces a single-valued diagnostic index for early detection of prostate cancer.

Most of the different approaches to the problem of clustering analysis suggested in the literature are mainly based on statistical, neural network, and machine learning techniques. An excellent survey of existing approaches is provided in Jain et al. [8]. At the same time, there are relatively few attempts to utilize optimization techniques for this purpose [14].

Bagirov et al. [4] propose the global optimization approach to clustering and demonstrate how the supervised data classification problem can be solved via clustering. The objective function in this problem is both nonsmooth and nonconvex and has a large number of local minimizers. Problems of this type are quite challenging for general-purpose global optimization techniques. Due to a large number of variables

and the complexity of the objective function, general-purpose global optimization techniques, as a rule, fail to solve such problems. It is very important, therefore, to develop optimization algorithms that allow the decision-maker to find "deep" local minimizers of the objective function. Such "deep" local minimizers provide a good enough description of the data set under consideration as far as clustering is concerned. The optimization algorithm discussed in this paper belongs to this type and is based on nonsmooth optimization techniques.

For many clinicians, however, it is still difficult to trust the decisions of an automated algorithm, and the importance of rules to explain the decisions is critical to the acceptance of the data mining technology. Certainly, automated rule generation methods such as classification and regression trees (CART) are available to find rules describing different (predefined) subsets of the data. When the data sample size is limited though, such approaches tend to find very accurate rules that apply to only a small number of patients, who are frequently outliers. Clinicians are unlikely to trust rules generated under these conditions, particularly when existing classification rules are based on expert clinical knowledge.

In Schwarz et al. [16], it was demonstrated that data mining techniques can play an important role in *rule refinement*, even if the sample size is limited. Instead of using directed knowledge discovery (supervised learning methods) to generate rules or models to assign prostate cancer patients into risk-homogeneous patient groups, Schwarz et al. [16] propose to use undirected knowledge discovery (unsupervised learning methods) to group the patients, explore the existing rules, and identify possible inconsistencies and refinements. A data visualization approach was required, and the self-organizing maps (SOM) [6, 9, 10] were used to achieve alternative groupings of patients and explore the consistency of the risk classifications.

The *objective* of this paper is to expand the findings presented by Churilov et al. [6] by comparing and complementing them with those presented by Schwarz et al. [16]. As a by-product, we provide a convincing illustration of the complex and integrated nature of a real data mining project by demonstrating that imaginative combined use of different data mining methodologies makes such a project a success.

Data Description and Existing Risk Groupings

THE WILLIAM BUCKLAND RADIOTHERAPY CENTER (WBRC) in metropolitan Melbourne, Australia, had been selecting the prostate cancer patients for various treatment approaches using risk groupings derived from a review of published predictive models as shown in Table 1. Patients' diagnostic statistics used in this case study includes pathology Gleason Score (1 = best, 10 = worst), tumor stage (1a = best, 4 = worst), and the PSA (biochemical prostate-specific antigen) level at the time of diagnosis, as well as patient's age. The database consists of 258 de-identified records of hormone-naïve patients who underwent external beam radiotherapy at some stage between January 1, 1990, and December 31, 1997. As part of clinical and PSA follow-up that happened at 3, 6, 9, 12, 18, and 24 months, and then annually, the biochemical (PSA) failure-free survival (bFFS) was calculated from the beginning of radiotherapy

Table 1. WBRC Risk Group Classification

Risk group	Gleason score	Tumor stage	PSA	Five-year bFFS (percent)
Low	2–6	T1c–2a	0–10	76.2
Intermediate	2–6	T1c–2a	> 10–20	53.4
	2–6	T2b–3a	0–20	
High	7	Up to T3a	0–20	37.8
	2–7	Up to T3a	> 20	
	2–7	T3b–4	Any	
	8–10	Any	Any	

until the biochemical failure. Biochemical (PSA) failure was defined as per ASTRO consensus [1] statement or any event that prompted androgen deprivation therapy. Specifically, PSA failure is defined as three successively increasing PSA measures during the follow-up checks after treatment, and is seen as an early indicator that treatment has been unsuccessful. We use PSA failure here as a proxy measure for survivability, since it can be detected earlier and correlates strongly with the bFFS measure. The concept of risk in this situation is directly linked to the chance of a bFFS over a given period of time (usually, five years)—that is, the higher the chance of biochemical (PSA) failure-free survival for a given patient, the lower is the corresponding clinical risk, and the less aggressive is the clinical treatment strategy that can be implemented.

Applying the rules shown in Table 1 to the initial data set, it becomes apparent that according to WBRC classification, 8.2 percent of all the patients belong to the “low-risk” group, 51.6 percent belong to the “intermediate-risk” group, and the remaining 38.3 percent of the patients belong to the “high-risk” group. These findings are summarized in Table 2. Note that this classification presents a relatively high percentage of patients in the “intermediate” group, thus complicating the decisions regarding appropriate treatment for many patients.

It is important to note that, as indicated in Table 1, the classification based on three aggregated groups is derived by applying seven rules. If all the seven rules are allowed to generate their corresponding risk groups, Table 3 can be obtained.

Note that separating the data into seven groups according to the rules used by WBRC reveals one very important inconsistency specifically demonstrating that the bFFS rate for the group High2 is, in fact, higher than the bFFS rate for the group Intermediate3. This directly contradicts the way the concept of clinical risk is intuitively defined, and raises the question of the appropriateness of the set of rules used by the WBRC to identify specific risk groups. Note also that this classification, though stipulating better “granularity” of risk groups, is fundamentally based on the set of (sometimes inappropriate) rules reported in the literature rather than “data-driven” rules that are elicited from the practical history observed in WBRC.

Table 2. WBRC Risk Group Classification with Relative Sizes of the Groups

	Low risk	Intermediate risk	High risk
Five-year bFFS (percent)	76.2	53.4	37.8
Percent of records	8.2	51.2	38.3

Table 3. Rule-Based Seven Groups Classification of WBRC Data

Risk groups	Low	Int1	Int2	Int3	High1	High2	High3
Five-year bFFS (percent)	76	68	65	41	31	56	40
Percent of records	8.2	10.9	14.5	25.8	20.3	6.3	11.7

Adopting a data-driven perspective on WBRC classification, we aim to improve the accuracy of the rules as applied to the WBRC patient database. It is important to note that the natural desire of a clinical decision-maker is to ensure that the classification rules ensure the following two conditions are met as closely as possible:

- the PSA failure measure for “high-risk” group should be high, whereas that for the “low-risk” group should be low, if the classification rules are accurate; and
- the size of the “intermediate-risk” group is minimal, thus making sure that the patient is either a “high-risk” or a “low-risk” one in order to make the future treatment decision as tightly linked to the patient’s risk attribute as possible.

These two conditions provide the decision-maker with two objectives that are not necessarily mutually attainable, and can even be conflicting.

Methodology

IT IS IMPORTANT TO NOTE THAT AT THE FIRST STAGE, we are proposing a methodology for *exploring* for identifying inconsistencies in the *existing rules*, rather than *generating* a completely new set of rules. It is our *hypothesis* that, although the existing rules are clinically meaningful, the accuracy of the WBRC patient database could be improved by refining them.

It should be noted that the existing rules make no reference to patient age, but this data is readily available and believed to be important to patient survivability. Clearly, there is room for improvement in the existing rules, by searching for the role of age in the classification structure, and exploring the boundaries of the existing rules.

At the *first stage* of this study, we use an undirected knowledge discovery approach (SOM) to explore the appropriateness of the rules across different segments of the patient database. Then, at the *second stage*, the optimization-based algorithm

for clustering patients into homogenous groupings as far as their survivability is concerned is applied and the corresponding results are validated, compared, and contrasted.

Self-Organizing Maps

SOMs [7, 10, 11, 13] are the best-known unsupervised neural network approach to clustering. Their advantage over traditional clustering techniques, such as the k -means algorithm [8], lies in the improved visualization capabilities resulting from the two-dimensional map of the clusters. Often, patterns in a high-dimensional input space have a very complicated structure, but this structure is made more transparent and simple when they are clustered in a lower-dimensional feature space. Kohonen [10, 11] developed SOMs as a way of automatically detecting strong features in large data sets. SOMs find a mapping from the high-dimensional input space to low-dimensional feature space, so the clusters that form become visible in this reduced dimensionality.

Since traditional clustering techniques find clusters within the input space, which is typically high dimensional and difficult to visualize, the dimension reduction provided by the SOM by projecting the clusters onto a two-dimensional map is seen as one of its main advantages. The software used to generate the SOMs is Viscosity SOMine (www.eudaptics.com), which provides a colorful cluster visualization tool, and the ability to inspect the distribution of different variables across the map [7].

The Nonsmooth Optimization Approach to Solving Clustering Problems

The subject of cluster analysis is the unsupervised classification of data and discovery of relationships within the data set without any guidance. The basic principle of identifying these hidden relationships is that if input patterns are similar, they should be grouped together. Two inputs are regarded as similar if the distance between these two inputs (in multidimensional input space) is small.

With this notion in mind, consider set A , which consists of r n -dimensional vectors $a^i = (a_1^i, \dots, a_n^i)$, $i = 1, \dots, r$, where n is the number of data fields/attributes. The aim of clustering is to represent this set as the union of q clusters. Since each cluster can be described by the location of its center, it is instrumental to locate a cluster's center in order to adequately describe the cluster itself. Thus, we address the problem of finding q points that serve as centers of corresponding clusters.

Consider now an arbitrary set X , consisting of q points $x^1, \dots, x^q$. The distance $d(a^i, X)$ from a point a^i that belongs to set A to set X is defined by

$$d(a^i, X) = \min_{s=1, \dots, q} \|x^s - a^i\|, \quad (1)$$

where

$$\begin{aligned}\|x\|_p &= \left(\sum_{l=1}^n |x_l|^p \right)^{1/p}, \quad 1 \leq p < \infty \\ \|x\|_\infty &= \max_{l=1, \dots, n} |x_l|.\end{aligned}\tag{2}$$

The deviation $d(a^i, X)$ from set A to set X can be calculated using the formula

$$\begin{aligned}d(A, X) &= \sum_{i=1}^r d(a^i, X) \\ &= \sum_{i=1}^r \min_{s=1, \dots, q} \|x^s - a^i\|.\end{aligned}\tag{3}$$

Thus, as far as optimization approach is concerned, the cluster analysis problem can be reduced to the following mathematical programming problem

$$\begin{aligned}\min & f(x^1, \dots, x^q) \\ \text{s.t.} & (x^1, \dots, x^q) \in \mathfrak{R}^{n \times q},\end{aligned}\tag{4}$$

where

$$f(x^1, \dots, x^q) = \sum_{i=1}^r \min_{s=1, \dots, q} \|x^s - a^i\|,\tag{5}$$

If $q > 1$, the objective function f in problem (4) is nonconvex and nonsmooth. Note that the number of variables in optimization problem (4) is $q \times n$. If the number q of clusters and the number n of data attributes are large, the decision-maker is facing a large-scale global optimization problem. Moreover, the form of the objective function in this problem is complex enough to not become amenable to the direct application of general-purpose global optimization methods. Therefore, in order to ensure the practicality of the optimization approach to clustering, the proper identification and use of local optimization methods with the special choice of a starting point is very important. Clearly, such an approach does not guarantee the globally optimal solution problem (4). On the other hand, this approach allows one to find a "deep" minimum of the objective function, which, in turn, provides a good enough clustering description of the data set under consideration.

Note also that the meaningful choice of the number of clusters is very important for clustering analysis. It is difficult to define a priori how many clusters represent the set A under consideration. In order to increase the knowledge-generating capacity of the resulting clusters, the optimization-based approach discussed in this paper adopts the following strategy: starting from a small enough number of clusters q , the decision-maker has to gradually increase the number of clusters for the analysis until certain termination criteria motivated by the underlying decision-making situation are satisfied.

As far as optimization is concerned, this means that if the solution of the corresponding optimization problem (4) is not satisfactory, the number of clusters q should be iteratively increased so that problem (4) should be solved with $q + 1$ clusters, and so on, until some termination criterion is met. This implies that one needs to solve repeatedly arising global optimization problem (4) with different values of q —a task even more challenging than solving a single global optimization problem. In order to avoid this difficulty, a step-by-step calculation of clusters is implemented in the optimization algorithm discussed below.

It is also important to note that the form of the objective function in problem (4) allows one to significantly reduce both the number of attributes (feature selection) and the number of records in a data set. Although not essential in the case reported in this paper, this ability proves very important for mining large data sets. The way the proposed algorithm utilizes these features is discussed in detail in Bagirov et al. [4] and Bagirov and Churilov [3].

The following optimization algorithm originally presented in Bagirov and Churilov [3] is used to cluster the prostate cancer patients for the purposes of this study.

Optimization Algorithm for Clustering Analysis of Prostate Cancer Patients

Step 1—Initialization: Select a tolerance $\varepsilon > 0$. Select an n -dimensional starting point $x^0 = (x^0_1, \dots, x^0_n)$ and solve the following minimization problem:

$$\min f(x) = \sum_{i=1}^r \|x - a^i\| \quad \text{s.t.} \quad x \in \mathbb{R}^n. \quad (6)$$

Let an n -dimensional point x^{1*} be a solution to this problem and f^1 be the corresponding objective function value. Set $k = 1$.

Step 2—Computation of the next cluster: Select an n -dimensional vector x^0 , construct a new $2n$ -dimensional starting point $x^{02} = (x^{1*}, x^0)$ and solve the following optimization problem:

$$\min f^k(x) \quad \text{s.t.} \quad x \in \mathbb{R}^n, \quad (7)$$

where

$$f^k(x) = \sum_{i=1}^m \min \{ \|x^{1*} - a^i\|, \dots, \|x^{k*} - a^i\|, \|x - a^i\| \}.$$

Step 3: Let $x^{k+1,*}$ be a solution to problem (7). Take $x^{k0} = (x^{1*}, \dots, x^{k*}, x^{k+1,*})$ as a new starting point and solve the following minimization problem:

$$\min f^k(x) \quad \text{s.t.} \quad x \in \mathbb{R}^{(k+1) \times n}, \quad (8)$$

where

$$f^k(x) = \sum_{i=1}^m \min_{j=1, \dots, k+1} \|x - a^j\|. \quad (9)$$

Step 4—Termination criterion: Let $x^{k+1,*}$ be a solution to problem (8) and $f^{k+1,*}$ be the corresponding value of the objective function. If

$$\frac{f^{k,*} - f^{k+1,*}}{f^{1,*}} < \varepsilon, \quad (10)$$

then stop, otherwise set $k = k + 1$ and go to Step 2.

Both problems (7) and (8) are nonsmooth optimization problem, and the discrete gradient method developed by Bagirov [2] is used to address these problems.

SOM Implementation and Analysis

AS DESCRIBED IN THE PREVIOUS SECTION, at the first stage, the SOMs were employed to cluster the patients using four inputs (patient age, PSA at time of diagnosis, Gleason score, and tumor stage). Once the clusters have formed, the distribution of other variables, such as PSA failure and WBRC classification group, can be superimposed over the clusters to identify any inconsistent regions.

Figure 1 shows the component planes of the four variables used to generate the clusters (small maps at the bottom), and the superimposition of PSA failure and WBRC risk group number (1 = low, 2 = intermediate, 3 = high). The circled clusters show two groups of patients. The upper cluster contains patients who have been classified as intermediate by the existing rules, and yet the probability of PSA failure is actually quite low. The average age of this group is 68.07 years. The lower cluster contains patients who have been classified as intermediate by the existing rules, and yet the probability of PSA failure is actually quite high. The average age of this group is 60.4 years. Clearly, when patient age is considered, two subsets of the intermediate-risk group can be found with extremely different PSA failure probabilities. In fact, the younger of these groups should be considered high risk, and patients in the older cluster should be considered low risk, according to the superimposed PSA failure statistics. Thus, the SOM has revealed a large group of patients labeled intermediate risk by the existing rules, who should be reclassified. If we introduce a new rule to split patients based on age greater than or less than 65, we find the accuracy improves.

A similar refinement of the existing rules is revealed when the SOM exercise is repeated, but this time using PSA failure as an input along with the other four inputs. The purpose of including PSA failure as an input is not to use the resulting map as a predictive model of risk at the time of diagnosis (since PSA failure is not known until many months or years after treatment), but rather to identify further inconsistencies in the rules and the existing labels.

Figure 1 shows the component planes of the five variables used to generate the clusters and the superimposition of WBRC risk group number. The circled region

Figure 1. SOM Component Planes (Generated Using PSA, Gleason, Age, and Tumor Stage), with Superimposed PSA Failure and WBRC Risk Group

contains a large number of patients currently classified as intermediate risk, and yet, clearly, in the PSA failure negative region of the map. Inspecting the component plane for the variable PSA (at time of diagnosis) reveals that these patients mostly have an initial PSA measure of less than 12. Thus a new rule can be created that classifies this subset of patients as low risk if PSA is less than 12.

Optimization-Based Grouping: Implementation and Analysis

APPLYING THE OPTIMIZATION ALGORITHM discussed in the previous section to the WBRC prostate cancer data set with $\varepsilon = 0.01$ using the patient's age, Gleason score, tumor stage, and PSA level at diagnosis as input parameters, 10 clusters are produced, as summarized in Table 4. Note that the "tumor stage" field was preprocessed by converting it to the scale of 1–8, where score 1 represents tumor stage 1a and score 8 represents tumor stage 4. The values reported represent the arithmetic mean (S), minimum and maximum value of input parameters [m , M], and the coefficient of variation K percent for all input parameters for every cluster obtained.

Note that clusters 2 and 3, which contain 2 and 6 records, respectively, are formed due to unusually large readings of the PSA and represent statistical outliers of some kind; but due to the nature of the problem domain, these clusters can hardly be excluded or ignored, as extreme values of some parameters (such as very high PSA readings in relatively young patients) may suggest very intensive treatment options for some patients.

Table 4. Ten Clusters Produced by the Optimization Algorithm

Cluster	Size	Tumor stage			Gleason score			PSA			Age		
		S	mM	K percent	S	mM	K percent	S	mM	K percent	S	mM	K percent
1	48	7.2	68	6	7.0	59	12	9.3	2,415	38	68	5,181	10
2	2	7.0	77	0	7.0	77	0	186.0	149,243	28	70	6,673	7
3	6	6.7	48	20	6.7	48	22	77.0	6,199	20	69	5,881	11
4	19	6.2	38	25	6.9	69	12	43.0	3,554	14	69	6,076	6
5	41	3.4	15	29	5.9	210	23	4.8	0.69	49	65	4,478	11
6	29	4.1	25	19	6.4	49	20	25.0	2,036	16	69	5,279	8
7	60	4.6	45	10	6.8	59	14	12.0	519	25	70	5,783	9
8	18	7.3	78	6	6.7	48	14	24.0	1,734	21	68	5,675	7
9	9	6.2	48	24	2.9	24	29	11.0	719	34	69	5,578	10
10	26	3.2	14	21	5.7	37	17	15.0	1,020	18	67	5,476	8

Table 5. Ten Clusters Produced by the Optimization Algorithm with Corresponding bFFS Values

Cluster	10	5	9	8	2	7	1	6	3	4
Percent of records	10	16	3	7	1	23	19	11	2	7
Five-year bFFS (percent)	73	71	67	61	50	47	44	38	17	16

At the next stage of the study, the corresponding bFFS rates were calculated for every cluster, as presented in Table 5 (note that the clusters are sorted in decreasing order of bFFS readings).

In order to comply with the arrangements on risk assessment adopted in WBRC, the percentage levels of five-year bFFS considered as boundaries between risk groups were kept the same. According to these arrangements, clusters 10 and 5 represent the low-risk group, clusters 6, 3, 4, and 1 clearly include high-risk patients, and the remaining clusters include the patients with intermediate risk. Note that the bFFS levels separation between clusters is much more pronounced than in the case of Table 3. This allows, for example, to closely investigate the patients from cluster 7, as they are much more likely to be at the riskier end of the intermediate group than other patients.

Discussion

THIS STUDY DEMONSTRATES THAT data mining techniques can play an important role in rule refinement, even if the sample size is limited. Instead of using directed knowledge discovery (supervised learning methods) to generate rules or models to assign prostate cancer patients into risk-homogeneous patient groups, we propose to use un-directed knowledge discovery (unsupervised learning methods) to group the patients, explore the existing rules, and identify possible inconsistencies and refinements.

The refined rules, as depicted in Figure 2, have improved both the accuracy of the high- and low-risk classifications, and reduced the number of patients in the intermediate-risk group. These two objectives were at first thought to be conflicting. Table 6 shows how the accuracy of the rules has been improved through the methodology of initial inspection, identifying age of 65 as a significant boundary using SOM, and identifying PSA of 12 as a significant boundary using another SOM. The first observation is that the percentage of patients classified in the intermediate-risk group has been substantially reduced from 51.2 percent to 34.8 percent without significantly affecting the accuracy of this class. The number of patients classified as low risk has more than doubled (from 8.2 percent to 20.4 percent), whereas the accuracy of the classification has improved (a reduction in the percentage of patients in the class who have PSA failure). Clinically, this is a positive outcome, since these patients can avoid unnecessary treatment. Similarly, the number of patients classified as high risk

Figure 2. Tree Representation of Refined Rules, After Applying SOM to Reveal Inconsistencies

Table 6. Accuracy and Coverage of Existing Rules and Improved Rules (in percent)

	Existing WBRC rules		After SOM		After optimization	
	PSA failure	Records	PSA failure	Records	PSA failure	Records
Low	23.8	8.2	21.2	20.4	28.4	26.2
Intermediate	46.6	51.2	44.9	34.8	46.9	34.4
High	62.2	38.3	67.0	44.9	64.3	39.4
Coverage		97.7		100.0		100.0

has increased (from 38.3 percent to 44.9 percent), whereas the accuracy of their classification as high risk (as evidenced by PSA failure) has simultaneously improved (from 62.2 percent to 67 percent having PSA failure).

As far as optimization-based clustering is concerned, in order to use the same baseline for comparison, 10 clusters produced by the optimization algorithm should now be aggregated into three groups, as summarized in Table 6.

Observe that, using the optimization-based clustering approach, both conditions 1 and 2 for an intuitively good classification discussed in the second section are satisfied. While the percentages of biochemical-failure-free survival (and, consequently,

the measurements for PSA failure) are similar for the initial WBRC approach, SOM clustering, and optimization-based clustering approaches, the resolution capacity of both clustering approaches is definitely much higher. In particular, under SOM classification, 20 percent of patients are classified as low risk; for optimization-based clustering this number is 26 percent, as opposed to 8 percent in the initial WBRC classification—these findings are of extreme importance for clinical judgment, as for a reasonably large number of elderly people they may mean lower dosage of radiation therapy or the absence of other kinds of therapies with significant side effects. Note also that the improvement in the low-risk group is obtained without any reduction effect on the high-risk group and only by the means of reducing the intermediate-risk group.

Summary and Conclusions

THIS PAPER HAS SUCCESSFULLY DEMONSTRATED that both SOMs and optimization-based clustering algorithms can be used to explore existing classification rules, developed by experts, and identify inconsistencies with a patient database. As the proposed optimization algorithm calculates clusters step-by-step and the form of the objective function allows the user to significantly reduce the number of instances in a data set, it can be effectively utilized for clustering in large-scale data sets. Conducting a similar study on a much larger prostate cancer database would therefore allow us to overcome some of the limitations of this study that are due to the modest size of the WBRC database.

A rule-based classification system is important for the clinician to feel comfortable with the decision. Certainly, a decision tree can be used to generate data-driven rules, but as we have shown, for small sample sizes these rules tend to describe outliers that do not necessarily generalize to larger data sets. The benefits of a more exploratory approach to refining the existing rules, which clinicians are already comfortable with, is clear. We have demonstrated that this approach has been able to improve the accuracy of the risk classifications, and reduce the number of patients in the intermediate-risk group, assisting clinicians by providing clearer decisions regarding appropriate treatment for more patients.

Overall, it is important to note that the human effect of this research is difficult to overestimate. As the correct risk classification often means the absence/presence of an extra type of an invasive therapy (often with multiple and strong side effects) a given patient would have to undergo, even minor improvement in risk grouping has direct and unambiguous effect on this patient's quality of life.

It is demonstrated that the proposed approach can support decision-making by improving the accuracy of risk assessment, and it can, therefore, be seen as an evidence-based predictive tool with high-knowledge-generation capabilities. The proposed methodology enables the rules to be refined and improved, without radically altering them beyond recognition. Note also that the suggested risk classification approach is not in any way specific to the domain of clinical medicine. Classification problems of a similar type frequently arise in broader organizational contexts, as the tasks of risk

grouping customers, employees, goods, services, duties, and so on, become the part of the everyday routine for all levels of organizational risk management. Decision support functionality based on the suggested combination of computational tools has a significant potential to enhance organizational information systems.

REFERENCES

1. ASTRO (American Society for Therapeutic Radiology and Oncology Consensus Panel). Consensus statement: Guidelines for PSA following the radiation therapy. *International Journal of Radiation Oncology, Biology, and Physics*, 37, 5 (1997), 1035–1041.
2. Bagirov, A.M. Minimization methods for one class of nonsmooth functions and calculation of semi-equilibrium prices. In A. Eberhard, R. Hill, D. Ralph, and B.M. Glover (eds.), *Progress in Optimization: Contribution from Australasia*. Dordrecht: Kluwer Academic, 1999, pp. 147–175.
3. Bagirov, A.M., and Churilov, L. An optimization-based approach to patient grouping for acute healthcare in Australia. In P. Sloot, D. Abramson, A.V. Bogdanov, J.J. Dongarra, A.Y. Zomaya, and Y.E. Gorbachev (eds.), *Computational Science—ICCS 2003: Lecture Notes in Computer Science*. Berlin: Springer, 2003, pp. 20–29.
4. Bagirov, A.M.; Rubinov, A.M.; and Yearwood, J. A heuristic algorithm for feature selection based on optimization techniques. In R. Sarker, H. Abbas, and C.S. Newton (eds.), *Heuristic and Optimization for Knowledge Discovery*. Hershey, PA: Idea Group, 2002, pp. 13–26.
5. Bellazzi, R.; Azzini, I.; Toffolo, G.; Bacchetti, S.; and Lise, M. Mining data from a knowledge management perspective: An application to outcome prediction in patients with resectable hepatocellular carcinoma. In S. Quaglini, P. Barahona, and S. Andreassen (eds.), *Artificial Intelligence in Medicine: Eighth Conference on Artificial Intelligence in Medicine in Europe, AIME 2001*. Berlin: Springer-Verlag, 2001, pp. 40–49.
6. Churilov, L.; Bagirov, A.M.; Schwartz, D.; Smith, K.A.; and Dally, M. Risk grouping of prostate cancer patients with optimization. In R.H. Sprague Jr. (ed.), *Proceedings of the Thirty-Seventh Annual Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society Press, 2004, pp. 136–144.
7. Deboeck, G., and Kohonen, T. *Visual Explorations in Finance with Self-Organizing Maps*. London: Springer-Verlag, 1998.
8. Jain, A.K.; Murty, M.N.; and Flynn, P.J. Data clustering: a review. *ACM Computing Surveys*, 31, 3 (1999), 264–323.
9. Keeney, R. Common mistakes in making value trade-offs. *Operations Research*, 50, 6 (2002), 935–945.
10. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 1 (1982), 59–69.
11. Kohonen, T. *Self-Organisation and Associative Memory*. New York: Springer-Verlag, 1988.
12. Land, W.H., Jr.; Masters, T.; Lo, J.Y.; McKee, D.W.; and Anderson, F.R. New results in breast cancer classification obtained from an evolutionary computation/adaptive boosting hybrid using mammogram and history data. In M.J. Embrechts, H.F. VanLandingham, and S.J. Ovaska (eds.), *SMCia/01—Proceedings of the 2001 IEEE Mountain Workshop on Soft Computing in Industrial Applications*. Los Alamitos, CA: IEEE Computer Society Press, 2001, pp. 47–52.
13. Lin, C.; Chen, H.; and Nunamaker, J.F. Verifying the proximity and size hypothesis for self-organizing maps. *Journal of Management Information Systems*, 16, 3 (Winter 1999–2000), 57–70.
14. Mangasarian, O.L. Mathematical programming in data mining. *Data Mining and Knowledge Discovery*, 1, 1 (1997), 183–201.
15. Mangasarian, O.L.; Street, W.N.; and Wolberg, W.H. Breast cancer diagnosis and prognosis via linear programming. *Operations Research*, 43, 4 (1995), 570–577.
16. Schwartz, D.; Smith, K.A.; Churilov, L.; Dally, M.; and Weber, R. Improving risk group-

ing rules for prostate cancer patients using self-organizing maps. In A. Abraham, M. Koppen, and K. Franke (eds.), *Design and Application of Hybrid Intelligent Systems*. Amsterdam: IOS Press, 2003, pp. 126–135.

17. Walter, D., and Mohan, C.K. ClaDia: A fuzzy classifier system for disease diagnosis. In *Proceedings of the 2000 Congress on Evolutionary Computation*. Los Alamitos, CA: IEEE Computer Society Press, 2000, pp. 1429–1435.

18. Zhang, Z., and Zhang, H. Development of a neural network derived index for early detection of prostate cancer. In *Proceedings of IJCNN'99—International Joint Conference on Neural Networks*. Los Alamitos, CA: IEEE Computer Society Press, 1999, pp. 3636–3641.

19. Zupan, B.; Demsar, J.; Kattan, M.W.; Beck, J.R.; and Bratko, I. Machine learning for survival analysis: A case study on recurrence of prostate cancer. *Artificial Intelligence in Medicine*, 20, 1 (August 2000), 59–75.

Copyright of *Journal of Management Information Systems* is the property of M.E. Sharpe Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.