Temporal Features and Kernel Methods for Predicting Sepsis in Postoperative Patients

JooSeuk Kim*, James M. Blum†, and Clayton D. Scott*
*Department of Electrical Engineering and Computer Science
†Department of Anesthesiology
University of Michigan, Ann Arbor

April 4, 2010

Abstract

Objective: Sepsis represents a major factor in morbidity and mortality in postoperative patients. The systemic inflammatory response syndrome (SIRS) criteria are binary statistics used to identify patients with sepsis, and are based on four physiological variables: body temperature, heart rate, breathing rate, and white blood cell count. However, the SIRS criteria have been criticized for having reduced specificity (high false positive rate), which diminishes their utility in clinical settings. This paper presents new features derived from the same four variables, and a methodology for predicting sepsis in postoperative patients under moderate care.

Methods and material: Data for 1213 sepsis and 26 non-sepsis patients are obtained from post-operative patients on telemetry. We propose new *temporal* features that capture trends and variability in the SIRS variables, and a framework for prediction based on kernel methods. Since the physiological variables of patients in moderate care are sampled irregularly, the temporal features often have missing values. We therefore propose modified kernels that account for these missing values, allowing us to apply existing kernel methods such as the two-class and one-class support vector machines.

Results: We compare the predictive power of the temporal features to those of the SIRS criteria. Performance is evaluated not just when the patients are discharged or sent to intensive care unit (ICU), but also some number of hours in advance. The experimental results show that using temporal features leads to improvements over the SIRS criteria by a statistically significant amount. We also present 6 temporal features that appear to be

most relevant for accurate prediction.
Conclusion: SIRS criteria are based on the extreme values of vital signs in a recent (typically 24-hour) period. The proposed temporal features also take into account the trends and variability of vital signs, and offer substantially improved predictive power in postoperative patients. The implications for clinical practice are potentially reduced time to administration of antibiotics, vasopressors, and IV fluids to patients that may otherwise not conventionally be thought of as suffering from sepsis.

KEYWORDS: sepsis, irregular sampling, missing features, kernel methods

1 Introduction

Sepsis refers to a systemic response arising from infection [1]. In the United States, 0.8 to 2 million patients become septic every year, 30% of which are surgical patients, and hospital mortality for sepsis patients ranges from 18% to 60% [2, 3]. The number of sepsis-related deaths has tripled over the past 20 years due to the increase in the number of sepsis cases, even though the mortality rate has decreased [2]. Because of its high mortality, post-operative surgical patients with possible sepsis are frequently admitted to an intensive care unit (ICU) from moderate care/telemetry unit for monitoring and treatment. Delay in treatment is associated with mortality. Hence, timely prediction of sepsis is critical.

The clinical definition of sepsis is the presence of systemic inflammatory response syndrome (SIRS), together with a known or suspected infection. The phrase systemic inflammatory response syndrome was proposed to describe an inflammatory state affecting the whole body, independent of its cause. SIRS is defined as the presence of two or more of the following "SIRS criteria":

- a body temperature greater than 38° C or less than 36° C
- a heart rate greater than 90 beats per minute
- tachypnea, manifested by a respiratory rate greater than 20 breaths per minute, or hyperventilation, as indicated by a PaCO2 of less than 32 mm Hg
- an alteration in the white blood cell count, such as a count greater than 12,000/cu mm, a count less than 4,000/cu mm, or the presence of more than 10 percent immature neutrophils

within a certain time window, e.g., during the past 24 hours [1].

SIRS criteria are widely used by physicians as a way to identify patients with possible sepsis [4, 5]. However, it has been criticized for having reduced specificity (high false positive rate) at acceptable sensitivities, thus limiting its use in clinical settings [4, 5, 6]. The SIRS criteria depend on the highest and/or lowest values of the four SIRS variables within the window. This motivates us look at more discriminating features within the window, such as the variance and range. We refer to such features as *temporal features* because they reflect how variables change with time.

In addition to temporal features, we also develop a methodology for prediction based on these features. We adopt the general framework of kernel methods, which have proven to be successful in a number of applications [7]. In particular, we employ particular kernel methods known as support vector machines. Based on the maximum-margin principle, SVMs employ kernels to generating nonlinear decision boundaries, and have empirically been shown to generalize well even in the presence of many irrelevant features.

One of the challenges with this approach is that measurements obtained from patients in moderate care are sampled *irregularly*, meaning the time between consecutive samples is not constant. Unlike patients in an intensive care unit (ICU), whose vital signs are monitored constantly, patients in moderate care are monitored based on the severity of their condition and the availability of nursing staff. Irregular sampling leads to situations where there are too few samples available in a given time window to compute some of our proposed temporal features. In other words, the feature vector associated to a patient is prone to have missing values. In our data, about 20% of temporal features are missing due to irregular sampling.

When dealing with missing data, care must be taken to ensure that the kernel is positive semi-definite which is a technical condition that is required for these methods to work. Previous work on kernels for missing data include [8, 9]. In [8], a polynomial kernel for missing data is proposed. Instead, they introduce a concept of instance margin (the margin of a hyperplane with respect to missing data) and adopt the maximum-margin principle to the setting. The resulting optimization problem is non-convex and solved iteratively. In [9], kernel values for missing data are estimated as an expected value of the kernel conditioned on observed data. As the authors stated, the drawback of this method is that if the dimensionality of the space is large, high dimensional integration technique is needed, which entails high computational cost. We propose a simple but effective method for handling missing data, which combines imputation by zero with a rescaling step inspired by

the notion of instance margin.

The application of machine learning methods such as logistic regression, artificial neural networks, and support vector machines to sepsis-related problems have been explored in different patient monitoring environments [10, 11]. The authors were interested in prediction of death [10] or severe sepsis [11] from sepsis patients rather than predicting sepsis itself from surgical patients. In their settings, since the patients were monitored in an ICU, the patients' vital signs were observed regularly and frequently, and therefore more conventional methods could be applied. We note that the general methodology developed here is applicable to the case of regularly sampled data.

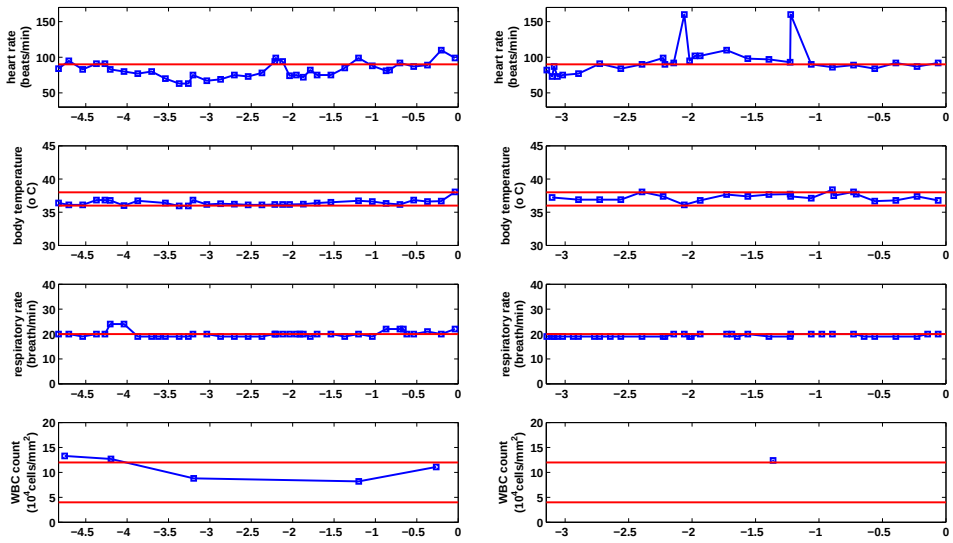
2 Problem Statement

For concreteness, we first describe our motivating application and data before presenting the general methodology.

2.1 Concrete Problem Statement

Institutional review board approved this study at The University of Michigan Hospital, a large, tertiary care facility. The data used for this study are from patients who were admitted for surgery and post-operative care between 7/1/2007 and 10/27/2008. A perioperative electronic medical record (Centricity, General Electric Healthcare, Waukesha, WI) was used to identify patients who were subsequently admitted to a telemetry unit for post-operative care.

Following the SIRS criteria, the variables we used were heart rate, body temperature, respiratory rate, and white blood cell count. For convenience, we refer to the variables as vital signs, even though white blood cell count is technically not considered a vital sign. Hemodynamic and respiratory data were acquired either automatically by a validated electronic interface from the physiological monitors (General Electric Healthcare) or manually by nursing staff. All physiologic data were acquired for each telemetry unit following the admission order protocol and were validated by clinical nursing staff prior to the entry into the medical record. In step-down/telemetry units, patients are monitored only when necessary. Hence, the vital signs obtained are recorded at irregular time intervals. In addition to the four vital signs, our data included demographic information including age, height, weight, and ASA (American Society of Anesthesiologists) classification of each patient, which was obtained from pre-operative history and physical



(a) non-sepsis patient (b) sepsis patient

Figure 1: Examples of vital signs recorded. Horizontal lines correspond to the thresholds defining the SIRS criteria.

examination. The vital signs were recorded up to the point where the patient was admitted to the intensive care unit (ICU) or discharged. Patients admitted to the ICU were reviewed for suspected sepsis. There are 1239 post operative patients, 26 of which become septic. This is consistent with the prevalence of sepsis in postoperative patients which is 1 to 16 percent [2, 12, 13].

We preprocessed the data in order to remove any obvious errors. For example, some body temperature readings had no indication of the temperature scale. If the recorded body temperature was greater than 60, we assumed that it was recorded in $^\circ$ F and converted the value to $^\circ$ C. We excluded patients whose recordings had none of above mentioned vital signs, and any samples in the vital signs that did not make sense were dropped, e.g., heart rate samples equal to zero.

Figure 1 shows representative examples of vital signs of patients with and without sepsis. In this figure, each patient's vital signs are time-shifted such that $t = 0$ corresponds to the time when he or she is discharged or admitted to the ICU. Note that the vital signs are sampled at irregular intervals and

Table 1: Notations

# of vital signs	N_d
the set of patterns	$\mathcal{X} = \{x : x = (d_1, \ldots, d_{N_d})\}$
vital signs	$d_i \in \mathbb{R}^{2p_i}, \quad \text{for some} \quad p_i, \quad \forall i = 1, \ldots, N_d$ $d_i = (t_{i1}, v_{i1}, t_{i2}, v_{i2}, \ldots, t_{ip_i}, v_{ip_i})$
class labels	$y \in \{-1, 1\}$

also differ in the number of samples. Our goal is to make accurate, early predictions of sepsis.

2.2 Abstract problem statement

We denote the vital signs of a patient in a structured form $x = (d_1, \ldots, d_{N_d})$ and the set of all x as $\mathcal{X}$. In the cases of sepsis prediction, the number of vital signs is $N_d = 4$. Each d_i corresponds to a vital sign of the patient and is an irregularly sampled time-series, i.e.,

$$d_i = (t_{i1}, v_{i1}, t_{i2}, v_{i2}, \ldots, t_{ip_i}, v_{ip_i})$$

for some $p_i \in \{0, 1, 2, \ldots\}$, where t_{ij} and v_{ij} represent the time and the value of the jth observed sample in d_i. Notice that within one pattern x, each d_i is obtained from irregular sampling, i.e., $(t_{i2} - t_{i1}), (t_{i3} - t_{i2}), \ldots, (t_{ip_i} - t_{i(p_i-1)})$ are typically distinct, and the number of observed samples for each d_i are different, i.e., $p_1, p_2, \ldots, p_{N_d}$ are typically distinct. Furthermore, for any two patterns x and $x' \in \mathcal{X}$, we typically have $p_i \neq p'_i$ for $i = 1, \ldots, N_d$, meaning different variables are recorded different numbers of times. The class label $y \in \{-1, 1\}$ of x is -1 if the patient corresponding to x is septic and 1 otherwise. The training data consists of labeled patients $(x_1, y_1), \ldots, (x_n, y_n)$, where in our application $n = 1213 + 26 = 1239$.

For each *training* patient, $t = 0$ corresponds to the time when he or she is admitted to the ICU or discharged. To assess the performance of early diagnosis, a test patient will be diagnosed not only when he or she is admitted to the ICU or discharged, but also some number of hours in advance. To do this, we will truncate the vital signs of a test patient beyond the time of prediction, and for this patient $t = 0$ corresponds to the time of prediction.

3 Proposed method

Our approach to this problem can be summarized as follows. First, we extract temporal features from vital signs. Because of irregular sampling, some features could be missing. For example, features extracted from temperature are missing in about 3% of the patients and those from white blood cell count are missing in about 50% of the patients. Thus, we define appropriate kernels (or similarity measures) that can handle the missing data. Once we obtain the kernel function, we are in a position where we can apply existing kernel-based machine learning algorithms, e.g., support vector machine. Our primary methodological contributions are feature extraction in irregularly sampled multivariate time series, and adapting kernel methods to handle missing features.

We define our temporal features as follows. Let $\Phi_\Delta : \mathcal{X} \to \mathbb{R}^l$ denote such a feature map, which outputs a vector of length l whose elements consist of temporal features from vital signs, based on a time window of length Δ. The time window is defined as $[-\Delta, 0]$, and Φ_Δ only considers samples that are observed within the window. Samples observed outside the window are ignored. For each vital sign, the temporal features are composed of the mean, standard deviation, range, maximum positive change, maximum negative change, and slope of a line fit using least squares regression. Therefore, $l = 4 \times 6 = 24$. Figure 2 illustrates the procedure of temporal feature extraction. After the extraction, we scale each feature to the range $[-1, 1]$.

Suppose that x and x' represent two patterns. We define a kernel function $k(x, x')$ by applying conventional kernels for Euclidean data to the features $\Phi_\Delta(x)$, $\Phi_\Delta(x')$. For example, for the polynomial kernel (1) or Gaussian kernel (2), we have

$$k(x, x') = \big(\langle \Phi_\Delta(x), \Phi_\Delta(x') \rangle + c\big)^p \tag{1}$$

$$k(x, x') = \exp\big(-\|\Phi_\Delta(x) - \Phi_\Delta(x')\|^2/\sigma^2\big). \tag{2}$$

However, because of irregular sampling there may be missing elements in $\Phi_\Delta(x)$ and/or $\Phi_\Delta(x')$, i.e, there may be vital signs that were never observed during $[-\Delta, 0]$ for x (and/or $[-\Delta, 0]$ for x'), or have insufficient samples to form the statistic. Therefore, we have to modify (1) or (2) to handle missing data.

A standard approach to missing data is imputing missing features by zero [14]. When computing an inner product, this is equivalent to ignoring the missing features. Suppose $\mathbf{w} \in \mathbb{R}^l$ is a vector with no missing elements. With the zero imputation, an inner product between $\mathbf{w}$ and $\Phi_\Delta(x)$ is defined

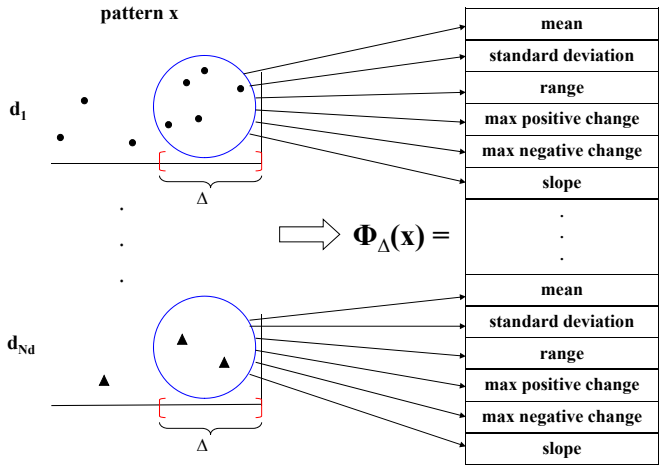


Figure 2: The illustration of temporal feature extraction process

as

$$\langle \mathbf{w}, \Phi_\Delta(x) \rangle_F = \sum_{i \in I_x} \mathbf{w}_i \cdot \Phi_\Delta(x)_i$$

where I_x represents the indices of non-missing elements in $\Phi_\Delta(x)$. As stated in [8], when combined with SVMs, zero imputation underestimates the margin of a hyperplane with a normal vector $\mathbf{w}$ with respect to $\Phi_\Delta(x)$ in a valid subspace. Thus, they introduce a concept of instance margin defined as

$$\frac{y\langle \mathbf{w}, \Phi_\Delta(x) \rangle_F}{\|\mathbf{w}(x)\|} \tag{3}$$

where $\mathbf{w}(x)$ is a vector formed with $\{\mathbf{w}_i\}_{i \in I_x}$, and adopt the maximum-margin principle to the setting. One can consider the instance margin as a conventional margin normalized by $\|\mathbf{w}\|/\|\mathbf{w}(x)\|$,

$$\frac{y\langle \mathbf{w}, \Phi_\Delta(x) \rangle_F}{\|\mathbf{w}(x)\|} = \frac{y\langle \mathbf{w}, \Phi_\Delta(x) \rangle_F}{\|\mathbf{w}\|} \cdot \frac{\|\mathbf{w}\|}{\|\mathbf{w}(x)\|}.$$

One drawback of the method is that due to the term $\mathbf{w}(x)$, the resulting optimization problem is non-convex and solved iteratively.

Here, we propose to impute the missing features by zero and rescale the

non-missing features by $l/|I_x|$,

$$\Phi'_\Delta(x)_k = \begin{cases} l/|I_x| \cdot \Phi_\Delta(x)_k & \text{if } \Phi_\Delta(x)_k \text{ is not missing,} \\ 0 & \text{if } \Phi_\Delta(x)_k \text{ is missing.} \end{cases}$$

The normalization term $\frac{l}{|I_x|} = \frac{\|\mathbf{w}\|_0}{\|\mathbf{w}(x)\|_0}$ can be thought as an approximation to $\frac{\|\mathbf{w}\|}{\|\mathbf{w}(x)\|}$. Then, we can define a new inner product $\langle \cdot, \cdot \rangle_N$ between $\mathbf{w}$ and $\Phi_\Delta(x)$ as a standard inner product between $\mathbf{w}$ and $\Phi'_\Delta(x)$,

$$\big\langle \mathbf{w}, \Phi_\Delta(x) \big\rangle_N = \langle \mathbf{w}, \Phi'_\Delta(x) \rangle = \sum_{i=1}^{l} \mathbf{w}_i \cdot \Phi'_\Delta(x)_i. \tag{4}$$

Another rationale behind this is that (4) is equivalent to

$$\big\langle \mathbf{w}, \Phi_\Delta(x) \big\rangle_N = \frac{l}{|I_x|} \sum_{i \in I_x} \mathbf{w}_i \cdot \Phi_\Delta(x)_i.$$

Therefore, this method of imputation assumes that if missing element $\Phi_\Delta(x)_k$ was observed, $\mathbf{w}_k \cdot \Phi_\Delta(x)_k$ would have produced a similar value to those of non-missing elements.

The inner product between $\Phi_\Delta(x)$ and $\Phi_\Delta(x')$ is then

$$\begin{aligned} \big\langle \Phi_\Delta(x), \Phi_\Delta(x') \big\rangle_N &= \big\langle \Phi'_\Delta(x), \Phi'_\Delta(x') \big\rangle \\ &= \frac{l}{|I_x|} \cdot \frac{l}{|I_{x'}|} \sum_{i \in I_x \bigcap I_{x'}} \Phi_\Delta(x)_i \cdot \Phi_\Delta(x')_i. \end{aligned} \tag{5}$$

For the distance between $\Phi_\Delta(x)$ and $\Phi_\Delta(x')$, we have

$$\begin{aligned} &\|\Phi_\Delta(x) - \Phi_\Delta(x')\|_N^2 = \|\Phi'_\Delta(x) - \Phi'_\Delta(x')\|^2 \\ &= \big\langle \Phi_\Delta(x), \Phi_\Delta(x) \big\rangle_N - 2\big\langle \Phi_\Delta(x), \Phi_\Delta(x') \big\rangle_N + \big\langle \Phi_\Delta(x'), \Phi_\Delta(x') \big\rangle_N. \end{aligned} \tag{6}$$

By replacing the inner product in (1) or the distance in (2) with (5) or (6), respectively, we have well-defined kernel functions. With these kernel functions, we can apply existing kernel-based machine learning methods. In experiments, we applied both zero imputation method and our proposed method with SVMs. The performance with zero imputation method was slightly worse than the proposed method, but not by a statistically significant amount, and the comparison results are not included.

Table 2: Methods

method	features	classification algorithm
temp-SVM	temporal features	SVM (linear)
temp-OCSVM	temporal features	OC-SVM (Gaussian)
SI-SVM	SIRS indicators	SVM (linear)
SS	SIRS indicators	sum and thresholded

4 Experiments

4.1 Experimental setting

Recall that there are 1239 post operative patients, 26 of which become septic. To show how the idea of introducing temporal feature works compared to SIRS criteria, we compare 3 feature sets, (a) the temporal features proposed in this paper, (b) SIRS indicators, and (c) SIRS score. By following a clinical convention, all the features are based on vital sign observations during the last $\Delta = 24$ hour window. SIRS indicators are 4 binary variables, each of which indicates whether the corresponding condition in SIRS criteria is met. When there is no observation for a certain vital sign, the corresponding indicator is assumed to be 0. SIRS score is defined as the sum of the SIRS indicators, taking values $0 - 4$. Note that a diagnosis of SIRS is equivalent to SIRS score ≥ 2.

Among machine learning algorithms, we first apply the SVM with the linear kernel (equation (1) with $c = 0$ and $p = 1$) to temporal features and SIRS indicators. Since SIRS score is just one variable, it is directly thresholded without any learning procedure. We also include experimental results when the OC-SVM (one class support vector machine) with Gaussian kernel (equation (2) with $\sigma = 1$) is applied to the temporal features. The OC-SVM uses only the non-sepsis patients as training data, and is motivated by our findings in 4.2 below. These 4 method are summarized in Table 2.

4.2 Exploratory results

A kernel can be considered as a measure of similarity between patterns expressed as an inner product in some feature space [7]. In this section, we plot the histograms of kernels values between patients (with the linear kernel) in Figure 3 to show whether the kernel (or similarity measure) we proposed

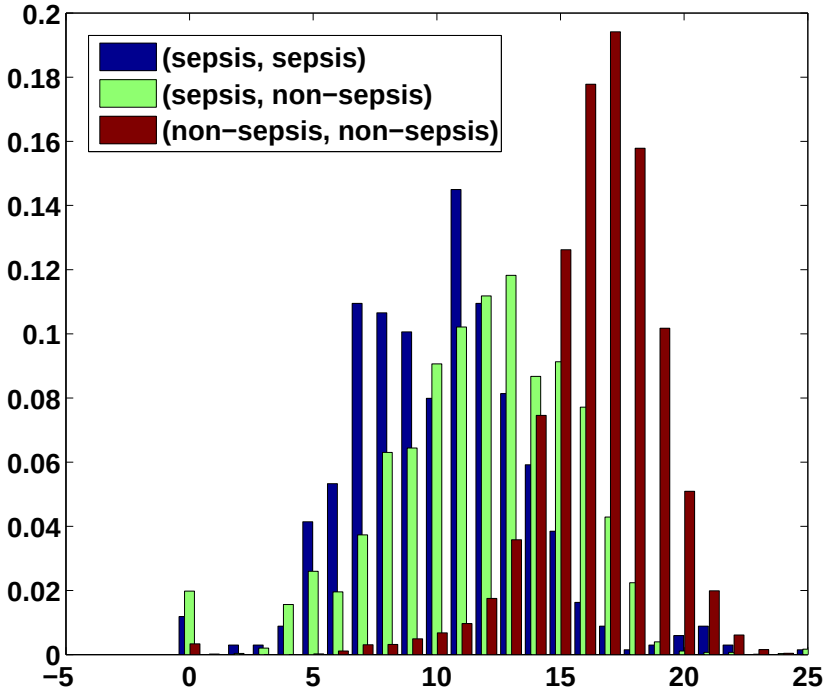


Figure 3: Histogram of similarity $k(x, x')$

previously captures the actual similarity between patients. Similarities between non-sepsis patients are distributed around 18 whereas those between sepsis patients and non-sepsis patients are distributed around 13. One interesting thing to note is the similarities between sepsis patients. We expected that similarities between sepsis patients would have larger values, but they don't. This observation can be explained as *healthy patients are alike but unhealthy patients are unhealthy in their own ways*. This motivated us to try the OC-SVM, since the sepsis patients did not seem to form a homogeneous class.

4.3 Performance comparison

We assess performance with the AUC (area under curve) of the ROC (receiver operating characteristic). We generate ROCs using different thresholds for the outputs of the decision function. AUCs and their confidence in-

tervals are estimated using 100 bootstrap samples [15]. Let $D = \{(x_i, y_i)\}_{i=1}^n$ denote the original sample and B_k, $k = 1, \ldots, 100$ denote the kth bootstrap sample of size n, which is obtained by sampling with replacement from D. Also let $AUC(D_{tr}, D_{test})$ denote AUC estimate with D_{tr} as training data and D_{test} as test data. Since $AUC(D, D)$ is likely to be higher than the true AUC, a bias reduction is required. This is done with bootstrap samples [15]. Let

$$AUC_k = AUC(D, D) - (AUC(B_k, B_k) - AUC(B_k, D)), \quad k = 1, \ldots, 100$$

and let $\widehat{\sigma}^2$ be the sample variance of $\{AUC_k\}_{k=1}^{100}$. Then, AUC and a $1 - \alpha$ confidence interval are estimated as

$$\widehat{AUC} = \frac{1}{100} \sum_{k=1}^{100} AUC_k$$
$$\widehat{CI} = [\widehat{AUC} - \mathcal{Z}_{\alpha/2}\widehat{\sigma}, \widehat{AUC} + \mathcal{Z}_{\alpha/2}\widehat{\sigma}]$$

where $\mathcal{Z}_\alpha$ is the upper αth quantile for the standard normal distribution.

Now we present comparison results of the 4 methods. We also show the performances of the methods for early prediction. To do this, we truncate a certain amount of time (3, 6, and 12 hours) from vital signs of test patients. The AUC plots are shown in Figure 4. (Parameters are set as $C = 1.0$ for SVM and $\nu = 0.5$ for OC-SVM.) We can easily see that temp-SVM is the best, and temp-OCSVM is the second best. The performances of SI-SVM and SS are similar and worse than the two other method by a significant amount. This suggests that temporal features have more predictive power than SIRS indicators or SIRS score.

More detailed results including AUC and 90% confidence interval estimates along with other choices of parameters are shown in Table 3. For the temp-SVM and temp-OCSVM, the best performance is obtained when either $C = 10.0$ or $\nu = 0.5$, but the other choices also lead to significant improvements over SI-SVM and SS.

4.4 Feature selection

We also investigated which features are most relevant for prediction. In this analysis, the feature set includes demographic data in addition to the temporal features. Previously, we used $\Delta = 24$ hour window for all the vital signs by following a clinical convention. In this section, however, since white blood cell count is less frequently observed, we use two different window

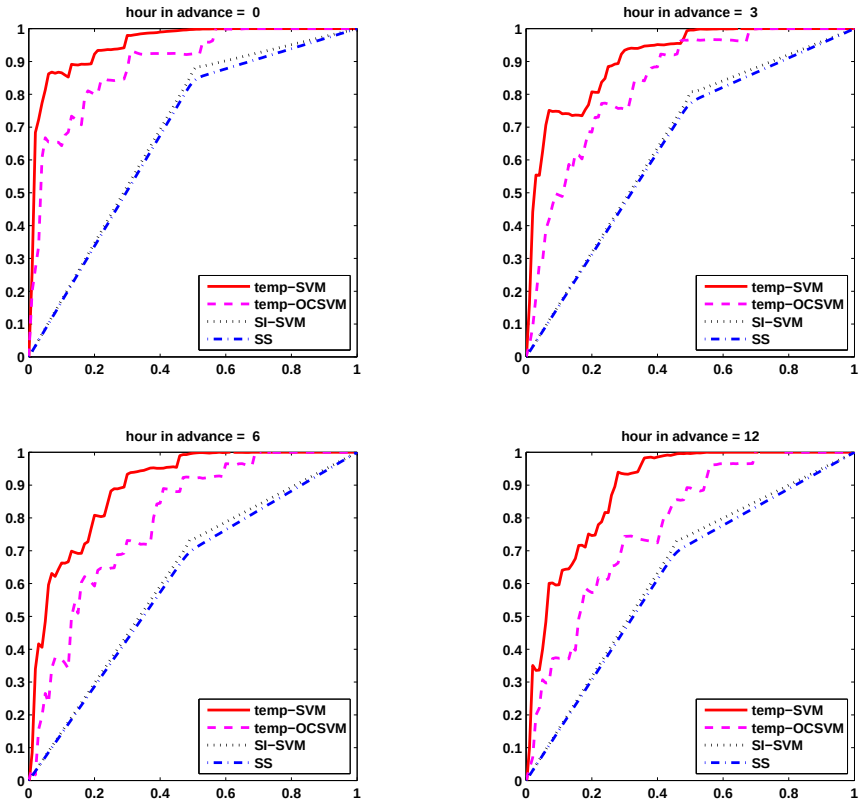


Figure 4: ROC curves of the 4 methods. Parameters are set as $C = 1.0$ for SVM and $\nu = 0.5$ for OC-SVM.

lengths: $\Delta_1 = 24$ for heart rate, respiratory rate, temperature, and $\Delta_2 = 72$ for white blood cell count . We rank the features based on mutual information between the features and labels [16]. To compute the mutual information, we first quantize each feature with 10 bins uniformly located from -1 to 1. Missing features are not included in the computation. The ranking according to mutual information is summarized in Table 4.

After feature ranking, we evaluated the performance of the best k features, for $k = 1, \ldots, 28$. The results are shown in Figure 5. AUC estimate does not change much until 6 features remain. However, when one more feature is removed from this feature set, the performance drops substantially. The 6 features are the maximum positive change of white blood cell count, the mean/range/standard deviation/maximum positive change of respira-

Table 3: AUC and confidence interval estimates

method	parameter	hours in advance			
		0	3	6	12
temp-SVM	$C = 0.1$	0.94 [0.91, 0.97]	0.88 [0.82, 0.93]	0.86 [0.81, 0.91]	0.84 [0.79, 0.89]
	$C = 1.0$	0.95 [0.93, 0.97]	0.91 [0.87, 0.95]	0.89 [0.85, 0.93]	0.88 [0.84, 0.92]
	$C = 10.0$	0.95 [0.92, 0.98]	0.92 [0.87, 0.96]	0.90 [0.85, 0.95]	0.90 [0.85, 0.95]
temp-OCSVM	$\nu = 0.1$	0.80 [0.70, 0.89]	0.70 [0.60, 0.80]	0.64 [0.53, 0.75]	0.63 [0.51, 0.74]
	$\nu = 0.3$	0.87 [0.80, 0.94]	0.81 [0.73, 0.88]	0.75 [0.67, 0.84]	0.71 [0.62, 0.80]
	$\nu = 0.5$	0.89 [0.83, 0.95]	0.83 [0.77, 0.89]	0.79 [0.73, 0.86]	0.75 [0.67, 0.82]
SI-SVM	$C = 0.1$	0.28 [0.22, 0.33]	0.34 [0.27, 0.40]	0.37 [0.30, 0.44]	0.35 [0.28, 0.43]
	$C = 1.0$	0.69 [0.63, 0.74]	0.65 [0.59, 0.71]	0.62 [0.55, 0.69]	0.65 [0.57, 0.73]
	$C = 10.0$	0.69 [0.63, 0.74]	0.65 [0.59, 0.71]	0.62 [0.55, 0.69]	0.65 [0.57, 0.73]
SS	N/A	0.67 [0.61, 0.73]	0.64 [0.57, 0.71]	0.61 [0.53, 0.68]	0.65 [0.57, 0.73]

tory rate, and the mean of heart rate.

5 Conclusion and Implications

In this paper, we propose a method for predicting sepsis in postoperative patients. Our methodology is based on the extraction of temporal features from the same physiological variables that define SIRS. Unlike the SIRS criteria, which only reflect the extreme values of vital signs in a given window, these temporal features also capture the trends and variability of vital signs. We also developed a framework for applying kernel methods with missing values, which allows our methodology to be applied to patients in moderate care, whose vital signs tend to be irregularly sampled.

Table 4: Feature ranking according to mutual information.

	Feature	Ranking		Feature	Ranking
	age	19		BMI	22
	ASA	25		gender	24
heart rate	mean	6	temperature	mean	18
	standard dev.	13		standard dev.	16
	range	8		range	15
	max pos. change	10		max pos. change	11
	max neg. change	23		max neg. change	14
	slope	27		slope	28
respiratory rate	mean	1	WBC count	mean	17
	standard dev.	4		standard dev.	9
	range	3		range	12
	max pos. change	5		max pos. change	2
	max neg. change	7		max neg. change	21
	slope	20		slope	26

The combination of temporal features in kernel methods leads to significant improvements in predictive power compared to the more conventional SIRS score. We evaluated these methods based on their ability to predict the presence of sepsis several hours in advance of when the patients were actually transitioned to an ICU or discharged. For example, when making predictions for six hours in advance, and assuming a specificity (false positive rate) of 10 percent, our method achieves a true positive rate of 66 percent, while the basic SIRS criteria lead to a true positive rate of around 15 percent. We also identified six specific temporal features that appeared to be most relevant for prediction of sepsis in postoperative patients.

The proposed temporal features were chosen a priori, and are intended to be simple yet general. Hand tuning of these features for particular variables may lead to further improvements.

The implications for clinical practice are potentially reduced time to effective administration of appropriate sepsis treatment. It has been well documented that delay in appropriate antibiotic administration in a septic patient has a profoundly negative impact on survival [17]. Furthermore, "goal directed" sepsis therapy is now recommended for early intervention in septic patients. Rivers, et al showed a reduction in mortality from 46.5

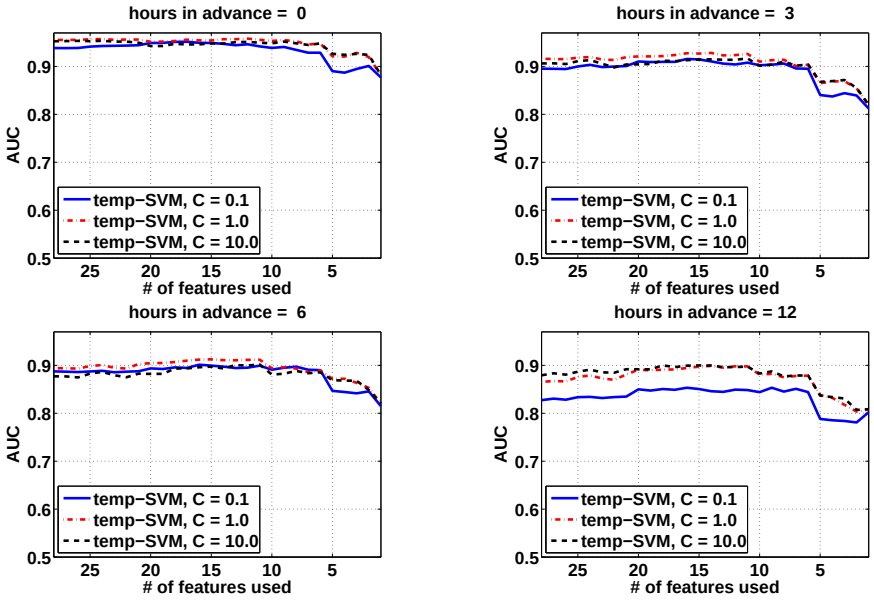


Figure 5: AUC plot as the number of features used decreases.

percent to 30.5 percent when such a protocol was implemented [18]. This protocol has become a key component of the Surviving Sepsis Campaign [19]. This therapy includes the administration of antibiotics, vasopressors, and IV fluids to patients that is thought to be suffering from sepsis. However, something must trigger this response. While profoundly septic patients are occasionally missed, it is not uncommon for a patient to begin expressing signs of sepsis that are not detected by clinicians, and only in retrospect become recognized as the start of a patient's demise. The algorithm we describe above has the potential to enable clinicians to focus on a potential sepsis diagnosis earlier. Further prospective trials will be required to determine if such triggered interventions reduce morbidity and mortality.

Acknowledgments

J. Kim and C. Scott were supported in part by NSF Award No. 0830490.

References

[1] American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference, Definitions for sepsis and organ failure and

guidelines for the use of innovative therapies in sepsis, Critical Care Medicine 20 (1992) 864–874.

[2] T. R. Vogel, V. T. Dombrovskiy, S. F. Lowry, Trends in postoperative sepsis: Are we improving outcomes?, Surgical infections 10 (1) (2009) 71–78.

[3] J. M. O'Brien, et al., Sepsis, The American Journal of Medicine 120 (2007) 1012–1022.

[4] R. P. Dellinger, et al., Surviving sepsis campaign guidelines for management of severe sepsis and septic shock, Critical Care Medicine 36 (1) (2008) 296–327.

[5] G. V. Bochicchio, et al., Persistent systemic inflammatory response syndrome is predictive of nosocomial infection in trauma, The Journal of Trauma 53 (2) (2002) 245–250.

[6] C. L. Sprung, et al., An evaluation of systemic inflammatory response syndrome signs in the Sepsis Occurrence in Acutely ill Patients (SOAP) study, Intensive Care Medicine 32 (3) (2006) 421–427.

[7] B. Schölkopf, A. J. Smola, Learning with Kernels, MIT Press, Cambridge, MA, 2002.

[8] G. Chechik, G. Heitz, G. Elidan, P. Abbeel, D. Koller, Max-margin classification of incomplete data, Journal of Machine Learning Research 9 (2008) 1–21.

[9] A. J. Smola, S. Vishwanathan, T. Hofmann, Kernel methods for missing variables, Proc. of the Tenth Int. Workshop on Artificial Intelligence and Statistics.

[10] F. Jaimes, J. Farbiarz, D. Alvarez, C. Martínez, Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room, Critical Care 9 (2005) R150–R156.

[11] B. Wang, The prediction of severe sepsis using svm model, Ph.D. thesis, National Chung Chen University (2006).

[12] R. Merkow, K. Bilimoria, M. McCarter, D. Bentrem, Effect of body mass index on short-term outcomes after colectomy for cancer, Journal of the American College of Surgeons 208 (1) (2009) 53–61.

[13] S. Behrman, B. Zarzaur, Intra-abdominal sepsis following pancreatic resection: incidence, risk factors, diagnosis, microbiology, management, and outcome, The American Surgeon 74(7) (2008) 572–578.

[14] R. Little, D. Rubin, Statistical analysis with missing data, 2nd Edition, Wiley, 2002.

[15] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall/CRC, 1994.

[16] T. Cover, J. Thomas, Elements of Information Theory, John Wiley and Sons, New York, 1991.

[17] S. F. Costa, Impact of antimicrobial resistance on the treatment and outcome of patients with sepsis, Shock 30 (2008) 23–29.

[18] E. Rivers, et al., Early goal-directed therapy in the treatment of severe sepsis and septic shock, The New England Journal of Medicine 345(19) (2001) 1368–1377.

[19] R. P. Dellinger, et al., Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2008, Critical Care Medicine 36(1) (2008) 296–327.