Predicting Sepsis Severity from Limited Temporal Observations

Xi Hang Cao, Ivan Stojkovic, and Zoran Obradovic

Center for Data Analytic and Biomedical Informatics
Temple University, Philadelphia, PA, USA
{xi.hang.cao,ivan.stojkovic,zoran.obradovic}@temple.edu
http://www.dabi.temple.edu/dabi

Abstract. Sepsis, an acute systemic inflammatory response syndrome caused by severe infection, is one of the leading causes of in-hospital mortality. Our recent work provides evidence that mortality rate in sepsis patients can be significantly reduced by Hemoadsorption (HA) therapy with duration determined by a data-driven approach. The therapy optimization process requires predicting high-mobility group protein B-1 concentration 24 hours in the future. However, measuring sepsis biomarkers is very costly, and also blood volume is limited such that the number of available temporal observations for training a regression model is small. The challenge addressed in this study is how to balance the trade-off of prediction accuracy versus the limited number of temporal observations by selecting a sampling protocol (biomarker selection and frequency of measurements) appropriately for the prediction model and measurement noise level. In particular, to predict HMGB1 concentration 24 hours ahead when limiting the number of blood drawings before therapy to three, we found that the accuracy of observing HMGB1 and three other cytokines (Lsel, TNF-alpha, and IL10) was comparable to observing eight cytokines that are commonly used sepsis biomarkers. We found that blood drawings 1-hour apart are preferred when measurements are noise free, but in presence of noise, blood drawings 3 hours apart are preferred. Comparing to the data-driven approaches, the sampling protocol obtained by using domain knowledge has a similar accuracy with the same cost, but half of the number of blood drawings.

Keywords: health informatics, acute inflammation, therapy optimization, limited temporal data, model predictive control.

1 Introduction and Motivation

Sepsis is a serious condition resulting from uncontrolled systematic inflammatory response to some pathogen infections. This condition is characterized by fast progression, severe symptoms and high mortality rate. In fact this is the number one cause of in hospital death in the USA [1]. Despite the high importance of the problem and substantial amount of researchers effort, not much progress has been achieved in resolving it. The vast heterogeneity of clinical manifestations

S. Džeroski et al. (Eds.): DS 2014, LNAI 8777, pp. 37–48, 2014.

makes identification of sepsis severity challenging. Another difficulty lies in rapid progression of the condition, where a patient goes from mild symptoms of infection to life threatening systematic inflammation condition in just several hours. Treatment consists of administering cocktails of various antibiotics in order to cover the spectrum of possible pathogens (usually bacteria) as much as possible. Even this aggressive treatment often is not enough since mortally in severe sepsis is as high as 30% and up to 70% when septic shock occurs [2].

Two main challenges arise in the problem of reducing the lethality of the sepsis. First it is very important to devise accurate diagnostic techniques that are also able to classify condition as early as possible. After correct diagnosis, it may be even more important that therapy is applied timely and appropriately. Since, in order to be effective, sepsis therapy should be aggressive, treating a person that is healthy is almost as undesirable as not treating an ill patient. The problems of early and accurate diagnostics have been addressed in a number of articles, such as [4], [5] and [8]. Recently, a form of blood purification called Hemoadsorption (HA) was proposed as a complement to antibiotic therapy. It was shown that HA is beneficial when used in animal models of sepsis [6]. It is based on removing certain cytokines from the blood, which are involved in mechanisms of systemic inflammation. Systemic inflammation takes place when these biomarkers enter a positive feedback loop with immune cells resulting in uncontrollable increase in inflammation. This process is known as a cytokine storm and it plays major role in number of conditions including sepsis. By cytokine reduction, HA therapy attempts to regain control over the inflammation process and return it to normal mode.

Given their roll in development of sepsis, observing cytokines over time is beneficial in both diagnostic and therapeutic purposes. Recently, they were used in the task of early classification of septic patients [8]. It is shown that applying HA therapy can be guided according to the predicted future values of cytokines in the Model Predictive Control framework [3]. On the other hand, there are also some constraints on cytokines use in the task of predicting sepsis progression. Constraints are mainly posed by limits on various resources. With current technologies fairly large volume of blood is needed to measure a particular (single) cytokine. However, there are at least 150 different cytokines, and many of them are involved in the inflammation process. Instead of measuring all cytokines, in clinical applications just a few of the most informative should be identified. Even when just a few cytokines are measured, measurement needs to be done on a number of different chronological occasions in order to catch the temporal dynamics of their change, which also increases demands on blood that needs to be drawn. In reality the amount of blood that can be drawn from a subject is a limiting factor in temporal observations of cytokines. The total amount of blood drawn over some period of time is limited by cost, data extraction time and even medical protocols. An additional constraint in small animals experiments (e.g mice and rats) is that drawing too much blood can interfere with the state of the subject since volume of the subject's bodily fluids is small.

In this article we are therefore addressing the problem of predicting progression of cytokines from a limited number of temporal observations. Here, we propose an approach for learning from limited temporal observations by utilizing prior knowledge of the interconnections of biomarkers and important internal states of sepsis progression. Using this approach, we discovered a blood drawing and biomarker measuring protocol which balances the constraints, cost, and accuracy.

The rest of the paper is organized as follows. In the second section the dynamical model of sepsis progression along with the process of virtual patient generation is presented. The third section is comprised of several subsections: in section 3.1, a detailed problem formulation of this study is provided; in section 3.2, a domain knowledge based approach is proposed; in section 3.3 and 3.4, alternative data-driven approaches are introduced. In section 4, experiments corresponding to different approaches are described in details, and results are analyzed. Summary and conclusion of this study is provided in section 5.

2 Sepsis Model and Data Generation

2.1 Model of Sepsis Progression

A set of Ordinary Differential Equations (ODE) describing the evolution of severe sepsis in rats is introduced in [7]. The network of interactions included in the model consists of 19 variables and 57 parameters. Out of 19 states 11 are unobservable: CLP protocol (CLP), Bacteria (B), Anti-Inflammatory state (AI), Pro-Inflammatory state (PI), Tissue Damage (D) and five types of Neutrophiles (Nr, Np, Na, Ns, Nt and Nl). Unobservable states are interconnected through equations with each-other and with eight cytokines, which represents variables that can really be measured. These eight observable states are the following plasma cytokines: tumor necrosis factor (TNF), three kinds of interleukins (IL-1b, IL-6 and IL-10), Lselectin (Lsel), high mobility group box1 (HMGB1), creatinine (CRT) and alanine aminotransferase (ALT). Domain knowledge was utilized to relate particular biomarkers that serves as proxies for particular unobservable states. Three cytokines TNF, IL-1b and IL-6 are well known as major Pro-Inflamatory mediators. Similarly IL-10 favors Anti-Inflamaton, while Lsel is related to Neutrophiles. The remaining three cytokines HMGB1, CRT and ALT are indicators of tissue damage and ODEs are devised accordingly. Most of the parameters in the model were fitted from real experimental data, while only a few were adopted from the literature. Experimental data were collected from a set of 23 rats where sepsis was induced by the CLP protocol. Eight longitudinal measurements of eight cytokines were collected at 18, 22, 48, 72, 120, 144, and 168 hours after sepsis induction.

The devised model, although coarse, serves to allow insight into plausible mechanism that drives the progression of sepsis. Moreover it provides a tool for performing experiments on in silico patients, which in turn can lead to new promising hypotheses that could later be evaluated in real experiments.

2.2 Generation of Virtual Patients

For the purpose of conducting experiments on the prediction of sepsis biomarkers, we used the ODE model for the generation of in silico or virtual patients. Every virtual patient behaves according to the mentioned dynamical equation, but each of the patients has a unique set of parameters and therefore unique response to the CLP induction of sepsis. Sets of parameters characterizing each patient were obtained using the following 3-step protocol: First, the valid ranges of parameter values are adopted from [7], and parameters are randomly sampled from those intervals. Next the 19 states model is simulated over time for chosen set of parameters. Finally, the likelihood that the evolution of 8 observable states follows the evolution of the real data from [7] is calculated, and if the likelihood is high enough then the virtual patient has been accepted as valid, or rejected otherwise. In that way, a number of virtual patients is generated for the purpose of training, validating or testing in the conducted experiments, for which setups and results are reported in following section.

3 Biomarkers Selection for Prediction of Sepsis Severity from Temporal Observations

3.1 Problem Definition

To determine the proper duration of HA therapy, the severity progression of sepsis is assessed, based on temporal observations of relevant variables, i.e. extending duration if sepsis severity is predicted to increase. In this paper, a cytokine called high-mobility group protein B-1 (HMGB1) is used as the biomarker indicating severity of sepsis. Recently, it has been shown that using HMGB1 in the objective function for model predictive control, the rescue rate was significantly improved [3]. Therefore, our objective is to estimate the value of HMGB1 in the future (typically 24 hours ahead) before applying therapy. Simulations have shown that 18% of septic patients could be rescued with a 4-hour duration HA therapy from the 18th hour since sepsis induction [7]. Therefore, in this problem setup, we would like to predict the value of HMGB1 at the 42nd hour since sepsis induction, while the start of therapy is scheduled at the 18th hour. One may think that this is a typical time series prediction problem, because once we measure HMGB1 for 18 hours, we can deploy any regression model to make predictions. However, in practice we cannot make observations at all historic time points. In our application the minimum time interval between two consecutive blood drawings is 1 hour. Because of this, no more than 18 blood drawings corresponding to hourly observation are possible before the start of therapy. In practice, the number of blood drawings is also limited by blood volume and medical regulations. However, at each blood sample we could measure multiple biomarkers, including HMGB1. That brings up another problem; each individual measurement of a biomarker is very costly, e.g. the cost of measuring 10 biomakers in a blood sample is 10 times as costly as measuring 1 biomarker. Sepsis biomarkers are correlated, and so measurement costs can be reduced as we can predict from

less measurements by utilizing their relationships. In summary, the problem addressed in this study is to balance the number of blood drawings, the number of biomarker measurements, and the prediction accuracy.

In the following 3 subsections we describe 3 methods for determining when to do blood drawings and what biomarkers to measure. A brute-force approach would consist of measuring 8 biomarkers hourly. Following such a protocol, in 18-hours, the number of measurement would be 144 (8×18=144). Considering all 2^{144} combinations of biomarker measurements is infeasible. Therefore, in this paper, we propose using a domain knowledge based approach to select biomarkers. This method is compared to two data-driven approaches based on feature selection and L1 regularization.

3.2 Domain Knowledge Based Sepsis Biomarkers Selection

Numerous sepsis-related studies resulted in understanding of the basic mechanism of sepsis. We propose using existing domain knowledge as clues about biomarkers that are closely related to sepsis progression. In particular, we use domain knowledge to identify biomarkers related to the prediction target HMGB1.

The prediction target HMBG1 is related to tissue damage (D). From the model description in Section 2.1, other cytokines related to D are creatinine (CRT) and alanine aminotrasferase (ALT). Thus, we assume that measuring CRT and ALT could provide information about changes of HMGB1 in the future. However, since CRT, ALT and HMGB1 are proxies for the same internal state, the changes of these three tissue-damage-related biomarkers over time should be similar. Therefore, in the proposed approach we decided not to measure them all, but just select one of them to measure. In order to increase the information gain, we propose that we should select biomarkers that are proxies to different internal states. In other words, for each blood sample, in addition to measuring HMGB1 which is the observable biomarker that estimates tissue damage and is also the prediction target of our interest, we propose measuring L-selectin (Lsel), which is a proxy for peritoneal neutrophil, tumor necrosis factor-α (TNFα), which is a proxy for systemic pro-inflammatory response, and interleukin-10 (IL10), which is a proxy for systemic anti-inflammatory response.

A reduction from 8 cytokines to measuring 4 (Lsel, HMGB1, TNFα, and Il10) in each blood sample is not sufficient, as the problem of when to draw blood remains. Intuitively, measurement at the 18th hour (start of the therapy) gives us the latest status information of the patient. Therefore, we will always draw blood and take measurements at the 18th hour. In the proposed approach we assume that the time between two consecutive blood drawings is the same. If a blood drawing is definite at the 18th hour, the problem becomes finding the most suitable sampling interval of blood drawings. We restrict the number of blood drawings/samples to three, as three is a reasonable number of blood drawings, and it allows us to investigate a variety of different choices of blood sampling intervals in the initial 18-hour period from infection time to the beginning of therapy. We expect that suitable sampling intervals would vary under different situations, such as the level of noise in the measurement data. Thus, we conduct

experiments to see how the preferred sampling interval changes under various conditions.

3.3 Forward Feature Selection Based Biomarkers Identification

We can also treat the problem described in Section 3.1 as a traditional feature selection problem in machine learning. If measurements of 8 biomarkers are available at every hour in the 18-hour history we have 144 features. To reduce measurements we can apply a greedy forward feature selection technique. The forward selection algorithm will try to add features to the candidate set. If the criterion function decreases after adding a feature to the candidate set, that feature will be included to the candidate set (Algorithm 1). In this case, the criterion function is the average root mean squared error (RMSE) in the training set using 5-fold cross validation, while a Linear Regression (LR) model is used as the predictor.

```
input  : X, feature set; f(.), criterion function
output: X_c, candidate set
initialization: S = ∞; gain = true; X_c is empty; x_a is empty;
while gain = true do
    gain = false;
    foreach feature x in X do
        add x to X_c;
        if f(X_c) < S then
            S = f(X_c); gain = true; x_a = x;
        end
        remove x from X_c;
    end
    if gain = true then
        add x_a to X_c; remove x_a from X;
    end
end
```

Algorithm 1. Biomarkers Identification by Forward Selection

3.4 Lasso Regression Based Biomarkers Selection

The Lasso Regression Model is a Linear Regression model that includes an $L_1 - norm$ regulation term to enhance the sparsity of the coefficients (β). The values of the coefficients are found by solving the optimization function (1). Thus, features with non-zero coefficients are relevant to the prediction task. So, Lasso Regression has a built-in functionality of feature selection.

$$\min_{\beta} ||X\beta - y||_2 + \lambda||\beta||_1 \tag{1}$$

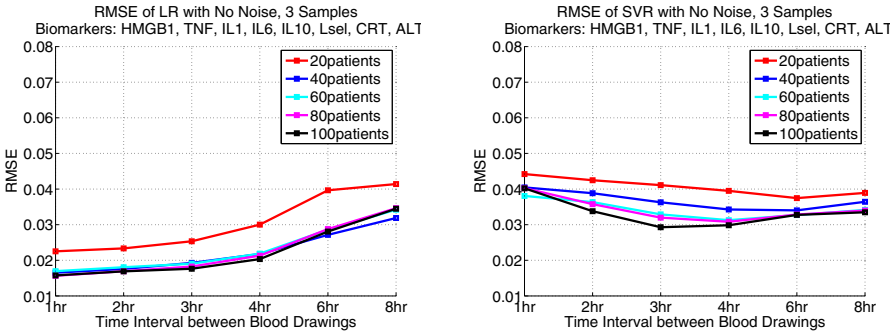
where X is the augmented feature matrix, y is the target vector, β is the coefficient vector, and λ is the regulation coefficient.

4 Experiments and Results

The data used in our experiments were generated by using the system of equations described in Section 2.1. The value of each biomarker measurement is between 0 and 1. In order to simulate real-life conditions, we will add various levels of uniform noise to the generated data.

4.1 Using Domain Knowledge to Select Biomarkers

In the proposed approach based on using prior knowledge, three blood drawings would be made; one of the three blood drawings would be always at the 18th hour. This experiment was designed to answer the following questions: 1. What is the most suitable time interval between two consecutive blood drawings? 2. How do different choices of biomarkers used in the model would affect the prediction accuracy? 3. How does the number of virtual patients in training affect the accuracy? 4. How does the noise level in the data affect the accuracy. For purposes of comparison, a linear model and a nonlinear model were used for prediction. The linear model was Linear Regression (LR), and the nonlinear model was Support Vector Regression (SVR) [10] with radial basis kernel. SVR was implement by using the LIBSVM package [11].



(a) (b)

Fig. 1. RMSE's of LR and SVR model, measuring 8 biomarkers at 3 blood drawings noise free. Results shown from models trained on 1, 2, 3, 4, 6, 8 hours interval between blood drawings and on data from 20 to 100 subjects.

As a baseline, we compare predictions with measurements of all eight biomarkers in the blood drawings. The number of virtual patients in training varied from 20 to 100, with increments of 20, and the time intervals between blood drawings were 1 hour, 2 hours, 3 hours, 4 hours, 6 hours, and 8 hours. The prediction error is measured by root mean squared error (RMSE) on 2,000 virtual patients. Figure 1a and Figure 1b show the RMSE's using different numbers of virtual patients in training, and using different time intervals between blood drawings

when number of blood drawings was 3, and with no noise present in measurements. The RMSE's of LR are lower than SVR when the time interval is small. From the figures, we learn that if measurements are noise-free, a short time interval between measuremetns (1 hour) will provide lower errors. We also learn that the model trained on observations from 40 virtual patients performs much better than the model trained on 20 virtual patients. However, training with more than 40 virtual patients has not further reduce prediction error. With uniform noise in range of [-0.02, 0.02] present in the measurements, on Figure 2a and Figure 2b, we learn that the RMSE's of LR and SVR here are very similar. In presence of noise, drawing blood with short time intervals was less accurate, and the effect of noise on larger time intervals was less significant. Larger time intervals between blood drawings are more robust to additive noise. In the case of LR, large time intervals between blood drawings result in lower error.

The obtained results provide evidence that including more virtual patients in training would not reduce errors. Therefore, in the following experiments, the number of virtual patients in training is fixed to 100.

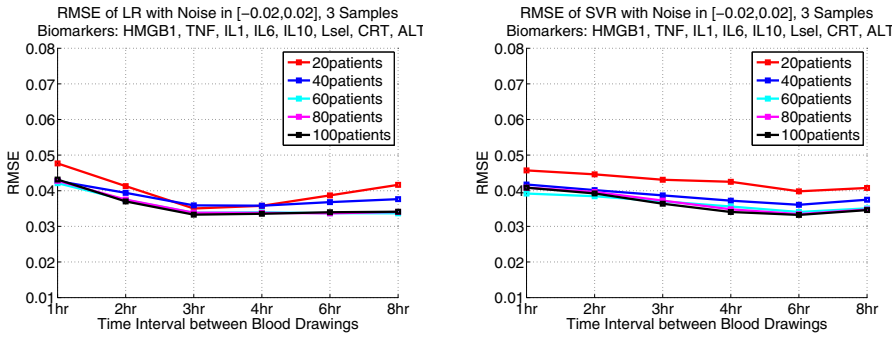


Fig. 2. RMSE's of LR and SVR model, measuring 8 biomarkers at 3 blood drawings with uniform measurement noise in [-0.02, 0.02] range. Results shown are from models trained by using 1, 2, 3, 4, 6, 8 hours interval between blood drawings and on data from 20 to 100 subjects.

Measuring all 8 biomarkers in each blood drawing is not desirable, as the total number of measurements is 24 when taking 3 blood drawings. We would like to obtain similar accuracy by measuring fewer biomarkers. The domain knowledge based approach described in Section 3.2 enabled us to do so. Figure 3a shows the RMSE's of the LR model trained by 3 blood drawings with noise-free, and noisy measurements of HMGB1, TNFα, IL10, and Lsel biomarkers which are related to different internal states that reflect severity of sepsis. In the obtained results, RMSE's in noise-free condition are smaller than the ones in noisy conditions; as the noise level increases, the errors increase. For uniform noise in the [-0.02, 0.02] range errors using these 4 biomarkers are similar to the ones based on all 8

biomarkers (black line in Figure 2a). So, since the number of blood drawings is the same, we could use half the number of measurements to achieve a very similar error. The prediction error when using HMGB1, CRT, and ALT biomarkers is shown in Figure 3b. For noise-free measurements, using these three biomarkers can achieve low error with 1-hour time intervals between blood drawings. For additive uniform noise in [-0.02, 0.02] the errors increases significantly, especially when the time interval between blood drawings is 1 hour. When noise is present, the overall errors using these three biomarkers are significantly higher than the ones when predicting based on 4 biomarkers.

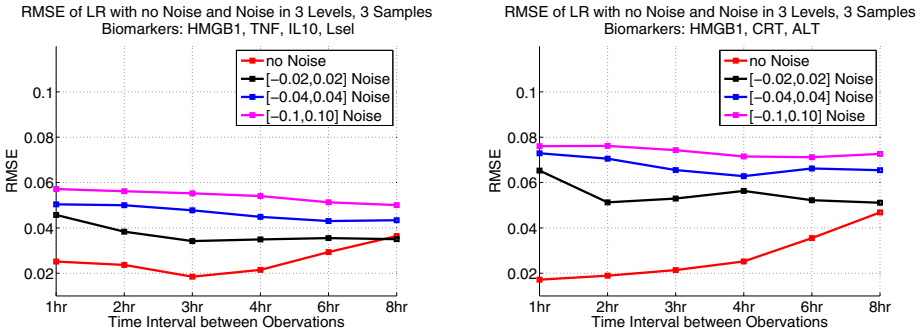


Fig. 3. RMSE's in LR model using 4 and 3 biomarkers measured at 3 blood drawings on (a) and (b) respectively. Results are based on models trained on 100 subjects under noise-free, and different noisy conditions, with sampling interval of 1, 2, 3, 4, 6, and 8 hours between blood drawings.

4.2 Using Forward Selection for Biomarkers Identification

The training set consisted of 100 virtual patients and measurements had [-0.02, 0.02] additive uniform noise. The criterion function of the selection procedure was the average RMSE using 5-fold cross validation on the training set. The selection procedure was repeated 20 times. Selected biomarkers are shown in Figure 4. In all the trails, number of biomarker measurements ranges from 7 to 14, number of required blood drawings ranges from 5 to 8. After testing the model on 2000 virtual patients in each trial, the range of RMSE is from 0.0356 to 0.0623. The minimum RMSE is achieved by 12 biomarker measurements from 7 blood drawings. The minimum RMSE is similar to the one achieved by the domain knowledge based approach, but the number of blood drawings is more than twice large (7 v.s 3). We found that about 21% of the biomarkers were selected from the 18th hour; this result consistent with our intuition that recent measurements are very informative for prediction. Other than the 18th hour, selected biomarkers uniformly span the whole 18-hour period.

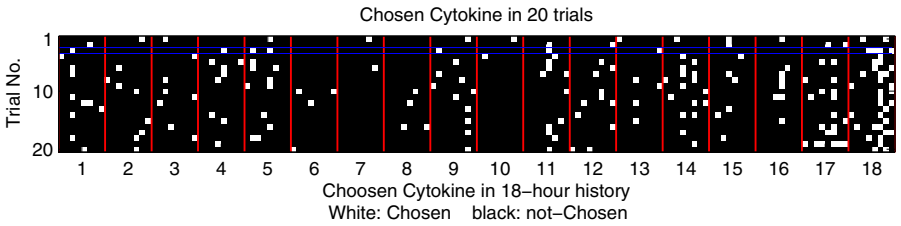


Fig. 4. Biomarker selection using sequential forward feature selection method. A matrix shows the biomarkers selected in 20 trials. The matrix dimension is 20 by 144, where 20 indicates 20 trials and 144 indicates 8 biomakers in 18 hour period (8×18=144).

4.3 Using Lasso Regression for Biomarker Selection and Sepsis Severity Prediction

100 virtual patients with [-0.02,0.02] uniform noise were used for training. 100 different values of the regularization coefficient λ were used to generate models with different numbers of non-zero coefficients. We tested 100 trained linear models (with different non-zero coefficients) on 2,000 virtual patients, and obtained the RMSE of each model. We found that the RMSE's remain low (about 0.035) when models with 12 or more non-zero coefficients (see Figure 5).

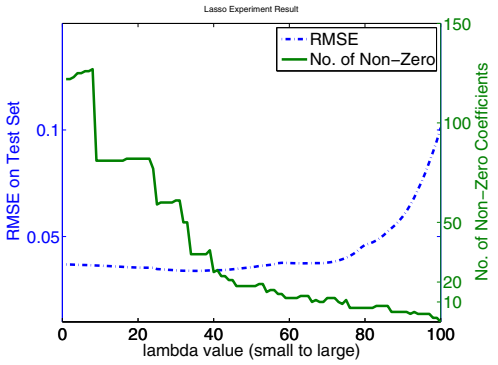


Fig. 5. Lasso regulation for feature selection

Models with 12 non-zero coefficients have RMSE's in range from 0.0347 to 0.0359. In the model achieves minimum RMSE, number of required blood drawings is 6, and number of required biomarker measurements is 12. Although, the minimum RMSE is similar to the one in the domain knowledge based approach, number of blood drawings is twice larger (6 v.s 3).

4.4 Overall Comparison of Different Approaches

The sampling protocol design objectives were low prediction error as well as small number of blood drawings and biomarker measurements. The prediction error (RMSE) of different approaches, their required number of blood drawings, and number of biomarkers measurements are shown at Table 1. Uniform noise in [-0.02, 0.02] range was added to the signal to simulate reality. The smallest RMSE was achieved by measuring all eight observable biomarkers. However, the error was just slightly larger when using only half of measurements selected based on knowledge of sepsis mechanism.

Table 1. Comparison of Different Approaches under Uniform Noise in [-0.02,0.02]

Approach	Best RMSE in test	No. of blood drawings	No. of biomarker measurements
Data-Driven: Forward Selection	0.0356	7	12
Data-Driven:Lasso Regression	0.0347	6	12
Domain Knowledge 8 biomarkers	0.0338	3	24
Domain Knowledge: 4 biomarkers	0.0341	3	12
Domain Knowledge: 3 biomarkers	0.0511	3	9

5 Summary and Conclusion

In this study, we used different approaches to characterize options for obtaining temporal observations of biomarkers in an 18-hour period to predict the value of HMGB1 in the future 24th hour. From the data-driven approaches, we learned that with blood drawings at proper times, 12 biomarker measurements were sufficient to make good predictions. Additional biomarker measurements would not improve the prediction accuracy. Inspired by the data-driven results, we came up with an approach that utilized domain knowledge of the interconnections of biomarkers and important internal states of sepsis progression. Using this approach, we discovered a blood drawing and biomarker measuring protocol which balances the constraints, cost, and accuracy.

Acknowledgment. This work was funded, in part, by DARPA grant [DARPAN66001-11-1-4183] negotiated by SSC Pacific grant.

References

1. Liu, V., Escobar, G.J., Greene, J.D., Soule, J., Whippy, A., Angus, D.C., Iwashyna, T.J.: Hospital Deaths in Patients With Sepsis From 2 Independent Cohorts. Journal of American Medicine Association (May 18, 2014)
2. Russel, J.A.: The current management of septic shock. Minerva Med. 99(5), 431–458 (2008)
3. Ghalwash, M., Obradovic, Z.: A Data-Driven Model for Optimizing Therapy Duration for Septic Patients. In: Proc. 14th SIAM Intl. Conf. Data Mining, 3rd Workshop on Data Mining for Medicine and Healthcare, Philadelphia, PA, USA (April 2014)
4. Ghalwash, M., Radosavljevic, V., Obradovic, Z.: Utilizing Temporal Patterns for Estimating Uncertainty in Interpretable Early Decision Making. In: Proc. 20th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining, New York, NY, USA (August 2014)
5. Ghalwash, M., Ramljak, D., Obradovic, Z.: Patient-Specific Early Classification of Multivariate Observations. International Journal of Data Mining and Bioinformatics (in press)
6. Peng, Z.Y., Wang, H.Z., Carter, M., Dileo, M., Bishop, J.V., et al.: Acute removal of common sepsis mediators does not explain the effects of extracorporeal blood purification in experimental sepsis. Kidney Int. 81, 363–369 (2011)
7. Song, S.O.K., Hogg, J., Peng, Z.-Y., Parker, R.S., Kellum, J.A., Clermont, G.: Ensemble models of neutrophil tracking in severe sepsis. PLoS Computational Biology 8(3) (2012)
8. Ghalwash, M.F., Radosavljevic, V., Obradovic, Z.: Early diagnosis and its benefits in sepsis blood purification treatment. In: IEEE International Conference on Healthcare Informatics (ICHI), International Workshop on Data Mining for Healthcare, Philadelphia, PA, USA (September 2013)
9. Hogg, J.S., Clermont, G., Parker, R.S.: Real-time optimization of cytokine-selective hemoadsorption devices for treatment of acute inflammation. Journal of Critical Care 26(2), e14 (2011)
10. Vapnik, V., Golowich, S.E., Smola, A.: Support vector method for function approximation, regression estimation, and signal processing. In: Advances in Neural Information Processing Systems, pp. 281–287. Morgan Kaufmann Publishers (1997)
11. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2(3), 27 (2011)