International Journal of Advanced Science and Technology
Vol. 55, June, 2013

Data Reduction of ICU Data using a Random Selection Approach

Augustine Nsang and Apkar Salatian

School of Information Technology and Communications
American University of Nigeria,
Yola Bypass, PMB 2250, Yola, Nigeria
{augustine.nsang, apkar.salatian}@aun.edu.ng

Abstract

Intensive Care Unit (ICU) monitors generate large volumes of high frequency data from numerous cardiac and respiratory sensors attached to a patient. This presents information overload to medical staff who need to interpret this data to evaluate the physiological status of the patient at any particular point in time. In this paper we present a machine learning technique called random selection to reduce ICU data sets. We will show that this technique derives trends in ICU data sets to enable qualitative reasoning as part of a clinical decision support.

Keywords: *Data reduction; ICU; random selection*

1. Introduction

The Intensive Care Unit (ICU) monitors display and store the continuous recordings of numerous physiologic signals taken from sensors attached to a patient. Typical physiologic signals recorded simultaneously include the heart rate, blood pressure, central venous pressure and the saturation of oxygen in the blood. The ICU monitors display the physiological status of the patient at any particular point in time. The frequency of the data can be higher than 1 value every second which creates information overload for the ICU medical staff. Medical staff, therefore, could benefit from a clinical decision support to assist in the interpretation of the voluminous data.

Reasoning of the ICU data on a point to point basis is computationally expensive. Indeed, [1] argue that time-series analysis could offer a potential for analyzing physiological measurements in the individual patient. Moreover, [2] suggest that clinical decision support systems should provide medical practitioners the ability to perform qualitative reasoning over time – this will relieve the medical staff from reasoning on a point to point basis and facilitate qualitative reasoning. Given continuous ICU data, we wish to identify trends within the data *i.e.*, group sequences of data points which share similar properties.

In this paper we propose a random selection approach as a time-series analysis technique to provide qualitative measurements from the high volume of high frequency data generated by the monitors in the ICU. The random selection approach allows the compression of voluminous monitor data into trends which facilitates clinical decision support.

The structure of this paper is as follows. Section 2 describes a random selection technique for deriving trends from ICU monitor data. Section 3 discusses the results we have obtained by applying our approach to a heart rate trace taken from an ICU monitor. A discussion of applications of our algorithm is given in Section 4. Section 5 discusses related work and a final conclusion is given in Section 6.

2. The Algorithm

In the random selection approach, to reduce a data set D containing n rows of information to one containing k rows of information, a set S_k is formed consisting of k numbers selected at random from the set S given by:

$$S = \{x \in \mathrm{N} \mid 1 \leq x \leq n\}$$

Then, our reduced set, D_R, will be given by:

$$D_R = D(\, S_k, :)$$

That is, D_R is a data set having the same number of columns as D, and the i^{th} row of D_R will be the j^{th} row of D if j is the i^{th} element of S_k.

To demonstrate the random selection approach, consider a data set, *A1***,** which is the heart rate of a patient during a time interval of 15 hours 5 seconds (*i.e.*, 3785 seconds). The data is taken from the monitors of an ICU in the United Kingdom as part of a research project. The data set has 3785 rows, one row for each second in this time interval. The graph representing this information is shown in Figure 1.

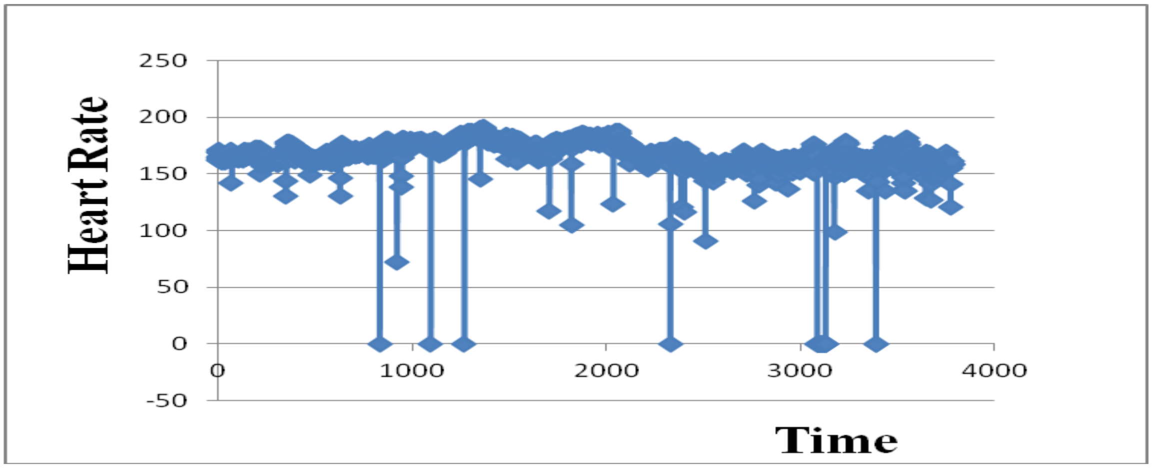

Figure 1. Graph Representing the Original Data Set Containing 3785 Rows

Naturally, it would be very helpful if the data set *A1* is compressed to give us a smaller data set, *A2*, representing the same information. We can do so by using the selection random approach. That is, to reduce the data set to 2000 rows, say, we randomly select 2000 rows from the 3785 rows in *A1* to give the data set *A2*. The MATLAB code which does this compression is given Figure 2.

Our algorithm randomly generates 2000 random numbers between 0 and 1, multiplies each number by 3785 and stores the result (rounded to the nearest integer) in an array *M1*. The array is then sorted to form a sorted array *M2* with 2000 values ranging from 0 to 3785. Then, a new matrix *M3* is generated containing, for each entry, *x*, in *M2*, the x^{th} row of *M*.

```
M  = load('M.txt');
M1 = [];
for i = 1 : 2000
   X = rand(1);
   Y = round(X * 3785);
   M1 = [M1;[Y]];
end
M2 = sort(M1);
M3 = [];
for i = 1 : 2000
   X = M2(i);
   Y = M(X,:);
   M3 = [M3;Y];
end

MX = M3(:,1);
MY = M3(:,3);
M3 = [MX,MY];
```

Figure 2. MATLAB Code for Compression

In other words, if *x* is the i^{th} entry of *M2*, then the x^{th} row of the original data set will become the i^{th} row of *M3*. The matrix *M3* is then stored as the data set *A2*.

The MATLAB code starts by loading the ICU monitor data into the matrix *M*. The first *for*-loop then generates 2000 random numbers between 1 and 3785 and stores them into a one-column matrix *M1*. *M1* is sorted to give *M2*, and the second *for*-loop then selects for each entry, *X*, of *M2*, the X^{th} row of *M* and stores them into a new matrix *M3*. Finally, the first and third columns of *M3* are maintained, and the second discarded.

3. Results

The graph representing the new matrix *M3* is given in Figure 3. As we can see, this graph is almost identical to the graph representing the original data set. Moreover, the noise in the original data has been attenuated. This certainly means that much of the noisy information in the original data set has not been selected by the random selection approach.

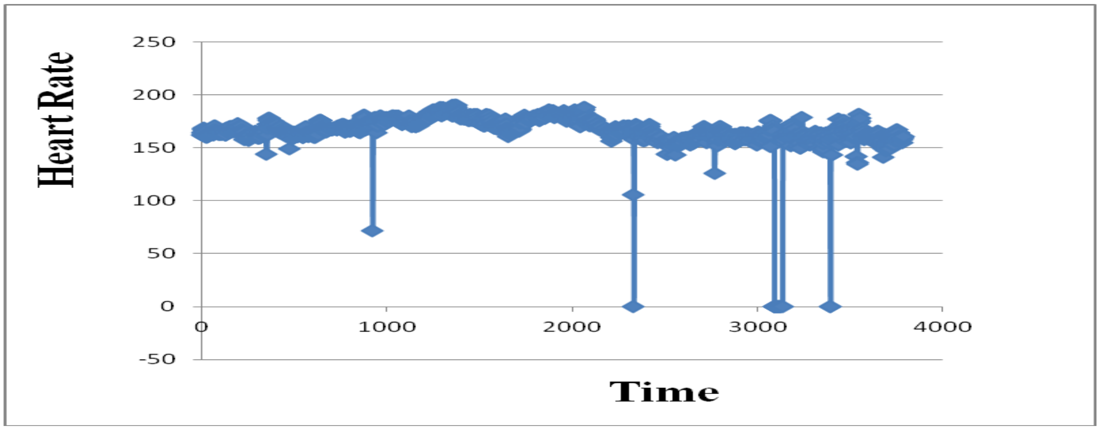

Figure 3. Graph Representing the Data Set Reduced to 2000 Rows

Obviously, we can use the same method to compress the original data set to 1000 rows, and even to 500 rows. *A3* is the data set obtained by reducing the original data set to 1000 rows, and *A4* is the data set obtained by reducing the original data set to 500 rows. The graph representing the data sets *A3* and *A4* are given in Figures 4 and 5. Clearly, these two graphs are almost identical to the graph representing the original data set.

Figure 4 also shows that the data set *A3* contains even less noise than the data set *A2*.

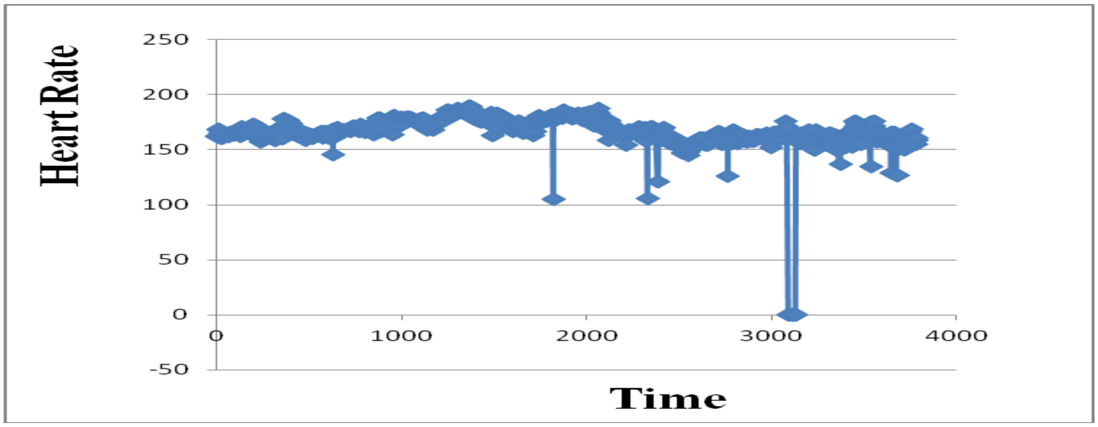

Figure 4. Graph Representing the Data Set Reduced to 1000 Rows

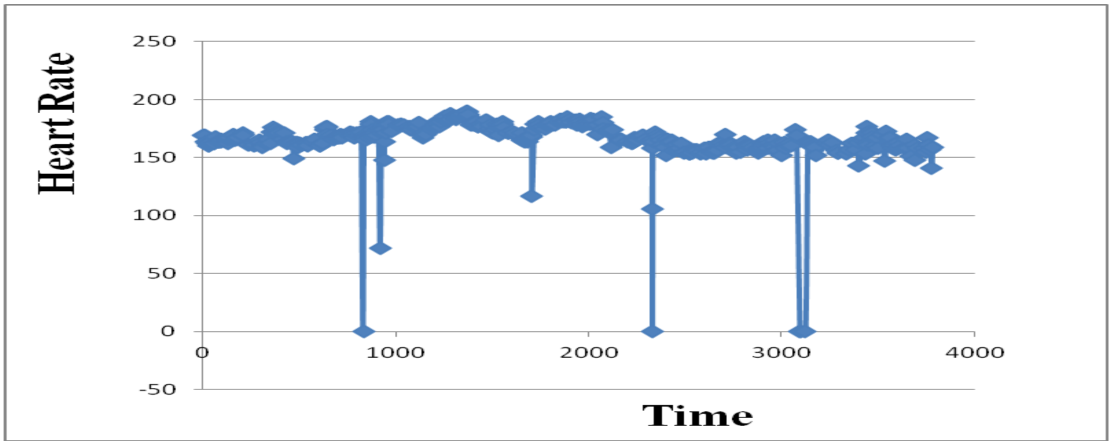

Figure 5. Graph Representing the Data Set Reduced to 500 Rows

Figure 5 shows the dataset *A4* contains more noise than the dataset *A3*. In other words, the noise in the data set has been re-introduced. This is not surprising considering the original dataset has been compressed to about one eighth of the original size.

4. Discussion

The generation of trends facilitates *trend-based (qualitative) reasoning* – this can enhance ICU clinical decision support systems.

Trends allow the identification and removal of clinically insignificant events, identification of clinical conditions and the outcome of therapies for clinical decision support and clinically insignificant events for removal [3] – these are achieved by performing associational reasoning about the temporal relationships between trends based on their endpoints. Clinically insignificant events which could not be removed by standard filtering can be identified and removed by associational reasoning of meeting trends of a single signal or overlapping meeting trends of multiple parameters. Clinical conditions can be identified by identifying particular patterns in overlapping (concurrent) trends of particular parameters**.** The outcomes of therapies are determined by comparing future trends to the trend when the therapy was administered to see if an expectation was met (or not).

5. Related Work

Data reduction techniques also include dimensionality reduction techniques. Given a collection of n data points (vectors) in high-dimensional Euclidean space, it is often helpful if it can be projected into a lower dimensional Euclidean space without suffering great distortion [4]. In other words, it is helpful if we can embed a set of n points in d-dimensional space, where d is sufficiently large, into k-dimensional space, where k << n (*i.e.* k is much smaller than n), without suffering great distortion. This operation is known as dimensionality reduction.

Traditional dimensionality reduction techniques include *Random Projection, Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Kernel Principal Component Analysis, Discrete Cosine Transform* and *Latent Semantic Analysis.*

In Random Projection [5], the original *d*-dimensional data is projected into a *k*-dimensional (k << d) subspace through the origin, using a random *d x k* matrix *R* whose columns have unit lengths. If X_{nxd} is the original set of *n d*-dimensional observations, then

$$X_{nxk} = X_{nxd}R_{dxk}$$

is the projection of the data onto a lower *k*-dimensional space.

Given *n* data points in R^P as an *n x p* matrix *X,* suppose we want to find the best *q*-dimensional approximation (q << p). SVD based dimensionality reduction consists of two steps:

1. Find the SVD of X. In other words, find *U*, *V* and *D* such that $X = UDV^T$, where *U* is an *n x n* orthogonal matrix, *V* is an *p x p* orthogonal matrix, and *D* is an n *x p* orthogonal matrix
2. The transformed matrix is given by:

$$X^{SVD} = X^T U_q$$

We note that SVD only works if X is a square matrix.

PCA is closely related to the SVD method (described above). The same definitions and explanations hold here, except that, in place of U_q, we are going to define V_q to be the matrix whose columns are the unit vectors corresponding to the q largest right singular values of X. We note that Vq, in this case, is a $p \times q$ matrix. In this case, thus, the transformed matrix is given by:

$$X^{PCA} = XV_q$$

We note that unlike the SVD method, PCA works for all matrices. That is, PCA works even if X were an $n \times p$ matrix with $n \neq p$.

In all of the methods listed above, each dimension in the reduced space is actually a linear combination of the dimensions in the original space. In this paper, we take the position that true dimensionality reduction is obtained when the final set of dimensions is a proper subset of the original set of dimensions. The following dimensionality reduction approaches reduce a dataset to a subset of the original attribute set: the variance approach (Var), the combined approach (CA), the direct approach (DA), the weighted attribute frequency approach (WAF), the best clustering performance (BCF) approach, the new top down approach and the new bottom up approach[6, 7].

With the *Var* approach, the variance within attribute values is used as a criterion for attribute selection. More precisely, let I denote the (sub)set of dimensions considered at a given moment. We start by assuming I is an empty set. Then, dimensions of D are added to I in decreasing order of their variances. That means that a subset of r dimensions will contain the dimensions of top r variances. Intuitively, it is easy to justify why dimensions of low variance are left out as they would fail to discriminate between data. Thus, $I_r = \{i_1,.., i_r\} \subset \{1, . . . , n\}$ is the collection of dimensions corresponding to the top r variances. The reduced data set, D_R, is given by:

$$D_R = D(:, I_r)$$

The i^{th} column of D_R is thus the column of the original data set with the i^{th} largest variance.

For a complete description of all the other six approaches in this category as well as all the other approaches in the first category the reader is advised to see [8].

Apart from data reduction techniques, there seems to be three main approaches to deriving temporal intervals from a set of historical data points: merging existing intervals into larger intervals, classifying data streams through a set of constraints, and wavelets.

Merging algorithms typically involve concatenating existing small intervals into longer intervals until they cannot be merged any more. These are based on domain-specific rules. Examples of merging algorithms are given by [9, 10] and [11]. Merging algorithms depend on domain knowledge to produce trends. The random selection approach does not require domain specific rules and derive trends from the properties of the data.

Series of data can be classified into intervals using a set of constraints. Examples of the use of such algorithms are TrenDx [12] which uses *trend templates* to define and detect trends in a series of time-stamped data, DIA-MON-1 [13] which uses fuzzy classification and [14] which derives trends by dividing the numerical range of a variable into regions of interest. The problem with algorithms that classify data streams through a set of constraints is having knowledge about the *fuzzy* boundaries of where the end points lie. The random selection deduce where the trends begin and end from the properties of the data.

[15] use wavelets to derive trends in ICU data. Wavelets are a mathematical tool for the extraction of information from many different kinds of data. They are a wave-like oscillation with amplitude that starts out at zero, increases, and then decreases back to zero. Sets of wavelets are generally needed to analyze data fully. Wavelets have the risk of losing features in the data from the compression process. Certainly, features in the data are also lost by the random selection process. This is clear since only a small proportion of the original data set are selected and the rest discarded. In fact, the random selection method will hardly work if the data does not have a uniform trend from the first row to the last row.

6. Summary and Conclusions

Trend detection using the random selection approach can relieve the medical staff from the complexity of having to reason qualitatively about the relationships between each data item in large data sets. Reasoning qualitatively with trends allows the identification of clinical conditions, clinically insignificant events and the outcome of therapies. Therefore, the random selection approach can facilitate a clinical support system.

Our approach has potential and our results are encouraging. We need to do more tests to derive the optimal number of trends from the random selection approach. Although not fully developed, we believe it to be a step forward in the development of systems for the interpretation of ICU data.

References

[1] M. Imhoff, M. Bauer, U. Gather and D. Lohlein, “Time Series Analysis in Intensive Care Medicine”, Applied Cardiopulmonary Pathophysiology, vol. 6, **(1997)**, pp. 263-281.

[2] R. Fried, U. Gather and M. Imhoff, “Pattern Recognition in Intensive Care Online Monitoring”, Schmitt, Manfred, Teodorescu Horia-Nicolai, Jain Ashlesha, Jain Ajita, Jain Sandhya and Jain Lakhmi. (eds): Studies In Fuzziness and Soft Computing, Computational Intelligence processing in Medical Diagnosis. Springer-Verlag Company, Chapter, Physica-Verlag Heidelberg, New York, vol. 6, **(2002)**, pp. 129-156.

[3] A. Salatian, “Interpreting Historical ICU Data Using Associational and Temporal Reasoning”, 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2003), Sacramento, California, USA, **(2003)**, pp. 442-450.

[4] D. Achlioptas, “Database Friendly Random Projections”, Journal of Computer and System Sciences, **(2003)**, pp. 671-678.

[5] A. S. Nsang and A. Ralescu, “A Review Of Dimensionality Reduction Methods And Their Applications”, Proceedings of the Twentieth Midwest Artificial Intelligence and Cognitive Science Conference, **(2009)**, pp. 118-123.

[6] A. S. Nsang and A. Ralescu, “Approaches to Dimensionality Reduction to a Subset of the Original Dimensions”, Proceedings of the Twenty-First Midwest Artificial Intelligence and Cognitive Science Conference, **(2010)**, pp. 70-77.

[7] A. S. Nsang and A. Ralescu, “More Dimensionality Reduction to a Subset of the Original Attribute Set”, Proceedings of the Twenty-First Midwest Artificial Intelligence and Cognitive Science Conference, **(2010)**, pp 109-116.

[8] A. Nsang, “Novel Approaches to Dimensionality Reduction: An Empirical Study”, Lambert Academic Publishing, Saarbrücken, Germany, **(2011)**.

[9] Y. Shahar, “A Framework for Knowledge-Based Temporal Abstraction”, Artificial Intelligence in Medicine, **(1997)**, pp. 79-133.

[10] A. Salatian and J. R. W. Hunter, “Deriving trends in historical and real-time continuously sampled medical data”, Journal of Intelligent Information Systems, vol. 13, **(1999)**, pp. 47-74.

[11] D. DeCoste, “Dynamic across-time measurement interpretation”, Artificial Intelligence, vol. 51, **(1991)**, pp. 273-341.

[12] I. S. Kohane and I. J. Haimowitz, “Hypothesis-Driven Data Abstraction with Trend Templates”, Symposium on Computer Applications in Medical Care, **(1993)**.

[13] F. Steimann and K. P. Adlassnig, “Two-Stage Interpretation of ICU Data Based on Fuzzy Sets”, Working Notes of AAAI Spring Symposium Series: Interpreting Clinical Data, Stanford University, **(1994)** March 21-23, pp. 152-156.

[14] S. Miksch, W. Horn, C. Popow and F. Paky “Therapy Planning Using Qualitative Trend Descriptions”, Proceeding of the 5th Conference of Artificial Intelligence in Medicine Europe (AIME 95), Pavia, Italy, **(1995)**.

[15] A. Salatian, F. Adepoju and A. Odinma, “TRENDER - A Wavelets Approach to Deriving Trends in Respiratory ICU Monitor Data”, 14th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2010, Volume II, Orlando, Florida, USA, **(2010)**, pp 23-27.