Detecting Artifacts in Clinical Data through Projection Retrieval

Abstract

We outline a novel approach to distinguish correct alerts from artifacts in multivariate vital signs data collected at the bedside of critical care patients. The framework selects informative low-dimensional projections of data that allow easy identification and interpretation of artifacts by humans. The results enable designing reliable decision rules that can be used to identify and ignore false alerts on-the-fly. The proposed approach aims at reducing the tedious effort of expert clinicians who must annotate training data to support development of decision support systems. Through our method, the expert intervention is reduced to simply validating the outcome produced by an automated system using a small part of the available data. The bulk of the data can then be labeled automatically. The framework we present makes the decision process transparent and comprehensible to aid the expert validation. The projections jointly form a solution to the learning task. The method works under the assumption that each projection addresses a different subset of the feature space. The purpose is to determine which of the subsets of data correspond to genuine clinical alerts and which are artifacts due to particuliarities of the monitoring devices or data acquisition processes. We show how artifacts can be isolated using a small amount of labeled samples and present our system's utility in identifying patterns in data that are informative to clinicians.

1. Introduction

Clinical monotoring systems are designed to process multiple sources of information about the current health condition of a patient and issue an alert whenever a change of status, typically an onset of some form of instability, requires attention of medical personnel. In practice, a substantial fraction of these alerts are not truly reflective of the important health events, but instead they are triggered by malfunctions or innacuracies of the monitoring equipment. Accidentally detached ECG electrode, transient readings from a dislocated blood oxygenation probe, or many other such problems of minor significance, may and in practice often do yield instability alerts. Frequency of such false detections may cause the ”alert fatigue” syndrome, often observed among medical personnel, particularly in critical care departments. The syndrome may have adverse effects on the quality of care and patient outcomes, if it leads to lowering sensitivity of personnel to alerts and, therefore, to increased risk of missing real crises. In order to maintain and enhance effectiveness of care, it is important to realiably identify and explain the nonconsequential artifacts. In this paper, we outline a novel approach to distinguish correct alerts from artifacts in multivariate vital signs data collected at the bedside of critical care patients. It selects informative low-dimensional projections of data that allow easy identification and interpretation of artifacts by humans. The results enable designing reliable decision rules that can be used to identify and ignore false alerts on-the-fly. They can also reduce data review and annotation efforts by expert clinicians, enhancing their focus on their primary mission of patient care.

The outlined problem can be generalized to any system designed to provide decision support to human users. Typically, this involves automating tasks such as grouping or classification while offering the experts insight into how the learning task was solved and how the model is applied to new data. An ideal scenario for a multitude of practical applications is the following: a domain expert provides the system with preliminary training data for some learning task; the system learns a model for the task (which uses only simple projections); the user provides queries (test points); for a given query point, the system selects the projection that is expected to be the most informative for this point; the system displays the outcome as well as a representation of how the task was performed within the selected projection.

The problem of recovering simple projections for classi-

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

fication has been formalized in (Fiterau & Dubrawski, 2012). The RECIP algorithm proposed there uses point estimators for conditional entropy and recovers a set of low-dimensional projections which classify queries using non-parametric discriminators in an alternate fashion - each query point is classified using one of the projections in the retrieved set. The technique used in this paper is a generalization of RECIP which is applicable to both classification and clustering. The framework, described in Section 3, called Regression for Informative Projection Retrieval (RIPR) can retrieve projections for any task which can be expessed in terms of a consistent loss function. RIPR is designed to work with any type of learner suitable to the particular task. For the application discussed in this paper, we consider linear classifiers (SVM) and nonparametric clustering models (K-means). A classifier or a clustering model is trained for every recovered projection and used for the subset of data assigned to that projection.

The topic of this paper is the application of RIPR to artifact isolation. We illustrate the projections recovered for the task of discriminating artifacts from genuine clinical alerts. Since the types of alerts we focus on are triggered by excessive values of one of the vital signals at a time, we build separate artifact deiscrimination models for alerts on respiratory rate, blood pressure, and oxygen saturation. We evaluate the perfomance of these models at annotating unlabeled data. We also show, through case studies, how the models can help physicians identify outliers and abnormalities in the vital signals. Finally, we outline an active learning procedure meant to reduce the effort of clinicians in adjudicating vital sign data as normal, artifact, or genuine alarm.

2. Related Work

The use of dimensionality reduction techniques is a common preprocessing step in applications where the use of simplified classification models is preferable. Methods that learn linear combinations of features, such as Linear Discriminant Analysis, are not ideal for the task considered here, since we prefer to rely on the dimensions available in the original feature space. Feature selection methods, such as the lasso, are suitable for identifying sets of relevant features, but do not consider interactions between them. Our work fits the areas of class dependent feature selection and context specific classification, highly connected to the concept of Transductive Learning (Gammerman et al., 1998). Other context-sensitive methods are Lazy and Data-Dependent Decision Trees, (Friedman et al., 1996) and (Marchand & Sokolova, 2005) respectively. (Ting et al., 2011) introduce the Feating submodel selection, which performs attribute splits followed by fitting local predictors. (Obozinski et al., 2010) present a subspace selection method in the context of multitask learning.

Unlike most of those approaches, RECIP is designed to retrieve subsets of the feature space designed for use in a way that is complementary to the basic task at hand while providing query-specific information.

3. Informative Projection Retrieval

This section describes the formulation of the Informative Projection Retrieval (IPR) problem, then describes an algorithmic framework generalized from the RECIP procedure in (Fiterau & Dubrawski, 2012).

The algorithm solves IPR when the learning task can be expressed in terms of a loss function and there exists a consistent point-estimator for the risk. The derivations in Section 3.1 follow the setup for the RECIP procedure, the main improvement being the formalization of the problem for learning tasks other than classification and the capability to include learners of any given class while RECIP only considered nonparametric classifiers. Section 3.3 shows how divergence estimators are used to customize the framework for classifcation and regression tasks.

3.1. Projection Recovery Formulation

Let us assume we are given a dataset $X = \{x_1 \ldots x_n\} \in \mathcal{X}^n$ where each sample $x_i \in \mathcal{X} \subseteq \mathbb{R}^m$ and a learning task on the space $\mathcal{X}$ with output in a space $\mathcal{Y}$ such as classification or regression. The task solver for the learning task is selected from from a solver class $\mathcal{T} = \{f : \mathcal{X} \to \mathcal{Y}\}$, were the risk for the solver class $\mathcal{T}$ is defined in terms of the loss ℓ as

$$\mathcal{R}(\tau, \mathcal{X}) = \mathbb{E}_{\mathcal{X}} \ell(x, \tau) \quad \forall \tau \in \mathcal{T}.$$

We define the optimal solver for the task as

$$\tau^* \overset{def}{=} \arg\min_{\tau \in \mathcal{T}} \mathcal{R}(\tau, \mathcal{X})$$

We will use the notation $\tau_{\{X\}}$ to indicate the task solver from class $\mathcal{T}$ which is obtained by minimizing the empirical risk over the training set X.

$$\tau_{\{X\}} \overset{def}{=} \arg\min_{\tau \in \mathcal{T}} \hat{\mathcal{R}}(\mathcal{T}, X) = \arg\min_{\tau \in \mathcal{T}} \frac{1}{n} \sum_{i=1}^{n} \ell(x_i, \tau)$$

We formalize the type of model that our IPR framework will construct. Class $\mathcal{M}$ contains models that have a set Π of projections of maximum dimension d, a set $\boldsymbol{\tau}$ of task solvers and a selection function g:

$$\begin{aligned}\mathcal{M} = \{ \qquad\qquad & \Pi = \{\pi; \quad \pi \in \mathbf{\Pi}, dim(\pi) \leq d\}, \\ & \boldsymbol{\tau} = \{\tau; \tau_i \in \mathcal{T}, \tau_i : \pi_i(\mathcal{X}) \rightarrow \mathcal{Y} \quad \forall i = 1 \dots |\Pi|\}, \\ & g \in \{f : \mathcal{X} \rightarrow \{1 \dots |\Pi|\}\} \quad \} \quad .\end{aligned}$$

The set $\mathbf{\Pi}$ contains all the axis-aligned projections. However, the subset $\Pi \subseteq \mathbf{\Pi}$ contained by $\mathcal{M}$ contains only projections that have at most d features. The parameter d is dictated by the application requirements. Values of 2 or 3 are expected since they permit users to view the projections. The selection function g picks the adequate projection π and its corresponding task solver τ to handle a given query x.

Based on this model, we derive a composite solver which combines the benefits of the solvers operating on the low-dimensional projections. The loss of this solver can be expressed in terms of the component losses.

$$\begin{aligned}\tau_{\mathcal{M}}(x) &= \tau_i(\pi_i(x)) \quad \text{where } g(x) = i \\ \ell(x, \tau_{\mathcal{M}}) &= \ell(\pi_{g(x)}(x), \tau_{g(x)})\end{aligned}$$

where $g(x)$ represents the index of the solver for point x is handled and $\pi_i(x)$ is the projection of x onto π_i. Optimizing over the *Informative Projection Model* class $\mathcal{M}$, the IPR problem for learning task $\mathcal{T}$ can be formulated as a minimization of the expected loss:

$$M^* = \arg\min_{\mathcal{M}} \mathbb{E}_{\mathcal{X}} \ell(\pi_{g(x)}(x), \tau_{g(x)}) \qquad (1)$$

Since we are dealing with an unsupervised problem in terms of the selection function, there are limitations on its learnability. One example in which recovery is successful is a dataset containing regulatory features:

$$\forall x \exists x^j \text{ with } x^j \in A \text{ , } \tau^*(x^1 \dots x^m) = \tau^*_A(x^{i_1} \dots x^{i_d})$$

In the example above, for a given point x, j is the regulatory feature. The interpretation is that for all points x whose j^{th} feature is in the set A, the targeted task can be optimally performed by the task solver τ^*_A by considering only features $\{i_1 \dots i_d\}$ of x. The task solver τ^*_A is only trained over samples for which $x^j \in A$.

3.2. Projection Recovery Framework (RIPR)

The starting point of the algorithm is writing the empirical version of (1) as a combinatorial problem over multiple projections. The algorithm is designed under the assumption of the existence of low-dimensional embeddings that enable capturing accurate models for the target task. In conformance with this assumption, every sample point x_i can be dealt with by just one projection π_j, in other words $g(x_i) = j$. We model this mapping as a binary matrix B:

$$B_{ij} = I[g(x_i) = j].$$

We write the minimizers of the risk and empirical risk:

$$M^* = \arg\min_{\mathcal{M}} \mathbb{E}_{\mathcal{X}} \sum_{j=1}^{|\Pi|} I[g(x) = j] \ell(\pi_j(x), \tau_j)$$

$$\hat{M}^* = \arg\min_{\mathcal{M}} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{|\Pi|} I[g(x_i) = j] \ell(\pi_j(x_i), \tau_j)$$

Assume now that we can consistently estimate the loss of a tasks learner τ at each available sample, that is

$$\exists \hat{\ell} \text{ s.t.} \quad \forall x \in \mathcal{X} \quad \tau \in \mathcal{T} \quad \hat{\ell}(x, \tau) \overset{n \rightarrow \infty}{\rightarrow} \ell(x, \tau) \qquad (2)$$

Plugging (2) into the minimization yields the final form used to obtain the estimated model:

$$\begin{aligned}\hat{M} &= \arg\min_{\mathcal{M}} \sum_{i=1}^{n} \sum_{j=1}^{|\Pi|} I[g(x_i) = j] \hat{\ell}(\pi_j(x_i), \tau_i) \\ &= \arg\min_{\mathcal{M}, |\Pi| < |\mathbf{\Pi}|} \sum_{i=1}^{n} \sum_{j=1}^{|\mathbf{\Pi}|} B_{ij} L_{ij} \quad , \quad L_{ij} = \hat{\ell}(\pi_j(x_i), \tau_i)\end{aligned}$$

The loss estimators L_{ij} are computed for every point on every subspace of up to the user-specified size d. B is learned through a regularized regression procedure that penalizes the number of projections Π used in the model. This translates to an ℓ_0 penalty on the number of non-zero columns in B, relaxed to ℓ_1. The ℓ_0 penalty is written as $I[B_{\cdot,j} \neq 0]$, while its relaxation is $||B||_{1,1}$.

$$\hat{B} = \arg\min_{B} ||L^* - L \odot B||_2^2 + \lambda \sum_{j=1}^{d^*} I[B_{\cdot,j} \neq 0]$$

where d^* is the number of d-dimensional projections, $L^*_i \overset{def}{=} \min_j L_{ij}$, the operator $\odot$ isojections of size $\leq d$ defined as

$$\odot : \mathbb{R}^{n,d^*} \times \mathbb{R}^{n,d^*} \rightarrow \mathbb{R}^n, \quad (L \odot B)_i = \sum_{j=1}^{d^*} L_{ij} B_{ij}$$

The optimization procedure is described in detail in (Fiterau & Dubrawski, 2012), the key difference to its use here being that we are computing the loss matrix L differently. The technique resembles the adaptive lasso, which gradually reduces the number of non-zero columns in B until a small and stable set of projections is converged upon. As illustrated in 1, the procedure uses the multiplier δ to gradually bias projection selection towards projections that not only perform well but also suit a large number of data points.

Algorithm 1 RIPR Framework

```
  $\delta = [1 \dots 1]$
  repeat
    $B = \arg\min_B ||L^* - L \odot B||_2^2 +$
      $\lambda_1 \sum_{j=1}^{d^*} ||B_{.,j}||_{\ell_1} + \lambda_2 |B\delta|_{\ell_1}$
        subject to
        $||B_{k,.}||_{\ell_1} = 1 \qquad k = 1 \dots n$
    $\delta_k = ||B_{.,j}||_{\ell_1} \qquad j = 1 \dots d^*$ (update muliplier)
    $\delta = (||\delta||_{\ell_1} - \delta)/||\delta||_{\ell_1}$
  until $\delta$ converges
  $\Pi = \{\pi_i; \quad |B_{.,i}|_{\ell_1} > 0 \quad \forall i = 1 \dots d^*\}$
  return $\Pi$
```

3.3. Losses for Classification and Clustering

Next, we show how to formulate IPR for different tasks. The aim is to find projections that are informative for a task given no knowledge of the actual class of solvers that will be used. For instance, we might be given a high dimensional classification problem for which to find a set of low-dimensional projections without considering the classifier - linear, kernel-based or nonparametric - that will ultimately be trained. Therefore, we incline to use nonparametric loss functions. The performance of the method will depend on the estimator's rate of convergence.

3.3.1. Classification

The IPR problem for classification is the topic of the previous work in (Fiterau & Dubrawski, 2012). We state some results obtained in the paper.

Proposition 3.1. *Given a variable $X \in \mathcal{X}$ and a binary variable Y, X sampled from the mixture model*

$$f(x) = p(y=0)f_0(x) + p(y=1)f_1(x) = p_0 f_0(x) + p_1 f_1(x) \ ,$$

$$H(Y|X) = -p_0 \log p_0 - p_1 \log p_1 - D_{KL}(f_0||f) - D_{KL}(f_1||f).$$

The conditional entropy over the points assigned to projection π_j is then shown to be estimated as follows:

$$\hat{H}(Y|\pi(X); \{x|g(x) = j\}) \propto \frac{1}{n}\sum_{i=1}^{n} I[g(x_i) = j]\hat{\ell}(x_i, \tau_{\pi_j}^k)$$

$$\hat{\ell}(x_i, \tau_{\pi_j}^k) = \Big(\frac{(n-1)\nu_k(\pi_j(x_i), \pi_j(X_{y(x_i)} \setminus x_i))}{n\nu_k(\pi_j(x_i), \pi_j(X_{\neg y(x_i)}))}\Big)^{(1-\alpha)|\pi_j|}$$

Above, the notation $\pi(X)$ is used to represent the projection of vector X onto π. Also, we will use X_γ to represent the subset of the sample for which the label is γ. The notation $X \setminus x$ refers to the sample obtained when removing point x from X. The function $\nu_k(x, X)$ represents the k^{th} distance from point x to its k-nearest-neighbor from the sample X. $\tau_{\pi_j}^k$ is the k-nn classifier on projection π_j.

This result is obtained by using the Tsallis α-divergence estimator introduced in (Poczos & Schneider, 2011) and yields an estimator for the loss when the target task is binary classification. α is a constant set to a value close to 1 (such as 0.95) and $|\pi_j|$ is the dimensionality of the subspace π_j.

3.3.2. Clustering

Unlike classification and regression, most types of clustering make it problematic to devise an objective that can be evaluated at every point, mainly because an overview of the data is needed for clustering, rather than local information. Distribution-based as well as centroid-based clustering fit a model on the entire set of points. This is an issue for the IPR problem because it is not known which data should be used for the set of submodels. To bypass this problem, we first learn the projections and the points corresponding to them using density-based clustering, which admits a local loss estimator. We then learn a clustering model (solver) on each projection using only the assigned points.

Density-based clustering uses areas of higher density than the remainder to group points. To achieve IPR for clustering, we consider the negative divergence, in the neighborhood of each sample, between the distribution from which the sample X is drawn and a uniform distribution on $\mathcal{X}$. Let U be a sample of size n drawn uniformly from $\mathcal{X}$. Again, we use the nearest-neighbor estimator converging to the KL divergence. τ_i^{clu} is some clustering technique such as K-means.

$$\hat{\mathcal{R}}_{clu}(\pi_i(x), \tau_i^{clu}) \quad \rightarrow \quad -KL(\pi_i(X)|||\pi_i(U))$$

$$\hat{\ell}_{clu}(\pi_i(x), \tau_i^{clu}) \quad \approx \quad \Big(\frac{d(\pi_i(x), \pi_i(X))}{d(\pi_i(x), U)}\Big)^{|\pi_i|(1-\alpha)}$$

4. Artifact Detection with RIPR

4.1. Vital Sign Monitoring Data

A prospective longitudinal study recruited admissions across 8 weeks to a 24 bed trauma and vascular surgery stepdown unit. Noninvasive vital sign (VS) monitoring consisted of 5-lead electrocardiogram to determine heart rate (HR) and respiratory rate (RR; bioimpedance), noninvasive blood pressure (oscillometric) to determine systolic (SBP) and diastolic (DBP) blood pressure, and peripheral arterial oxygen saturation by finger plethysmography (SpO2). Noninvasive continuous monitoring data were downloaded from bedside monitors and analyzed for vital signs beyond local instability criteria: HR<40 or >140, RR<8 or >36, systolic BP <80 or >200, diastolic BP>110, SpO2<85%. VS time plots of patients whose vital sign parameters crossed the instability thresholds for any reason were visually assessed to judge them as waveform patterns consistent with physiologically plausible instability, or as physiologically implausible and therefore artifactual.

Each alert is associated with a category indicating the type of the chronologically first vital signal that exceeds its control limits. As a result, an alert with labeled as 'respiratory rate'may also include other vitals outside of the bounds that have escalated shortly after the exceedence or respiratory rate is recognized. We extracted a number of features to characterize each of the 813 alert events found in our data. The features are computed for each vital signal independently during the duration of each alert and a short window (of 4 minutes) preceding its onset. The list of features includes common statistics of each vital signal such as mean, standard deviation, minimum, maximum and range of values. It also includes features that are thought to be relevant (by domain experts) in discriminating between artifacts and true alerts. There are a total of 147 features derived from all vital signs as follows:

- The data density, which is the normalized count of signal readings during the alert period, a low value indicates the temporal sparseness of the data, a value of zero simply means there was no data captured in that period;
- The minimum and maximum of the first order difference of vital signal value during alert window. Extreme values indicate a sharp increase/decrease of the signal value;
- The difference of means of vital sign values for the 4 minute window before and after the alert;
- The value of the slope as the result of fitting linear regression to the vital values versus the time index;

4.2. Artifact Classification Models

We now show the classification models obtained to distinguish between artifacts and alerts corresponding to different vitals. We considered alerts associated with different vitals as separate classification tasks. Out of the 813 alert samples, 181 have been identified by expert clinicians as artifacts. Aside from the 813 labeled samples, there is a large amount of data that remains unlabeled. The goal now is to train a separate model for each alert type such that other potential artifacts can be detected in the unlabeled data. Since the classification results will be reviewed by domain experts, we rely on the RIPR framework to extract simple and intuitive projections which will make it easy for clinicians to validate the results.

4.2.1. Respiratory Rate Alerts

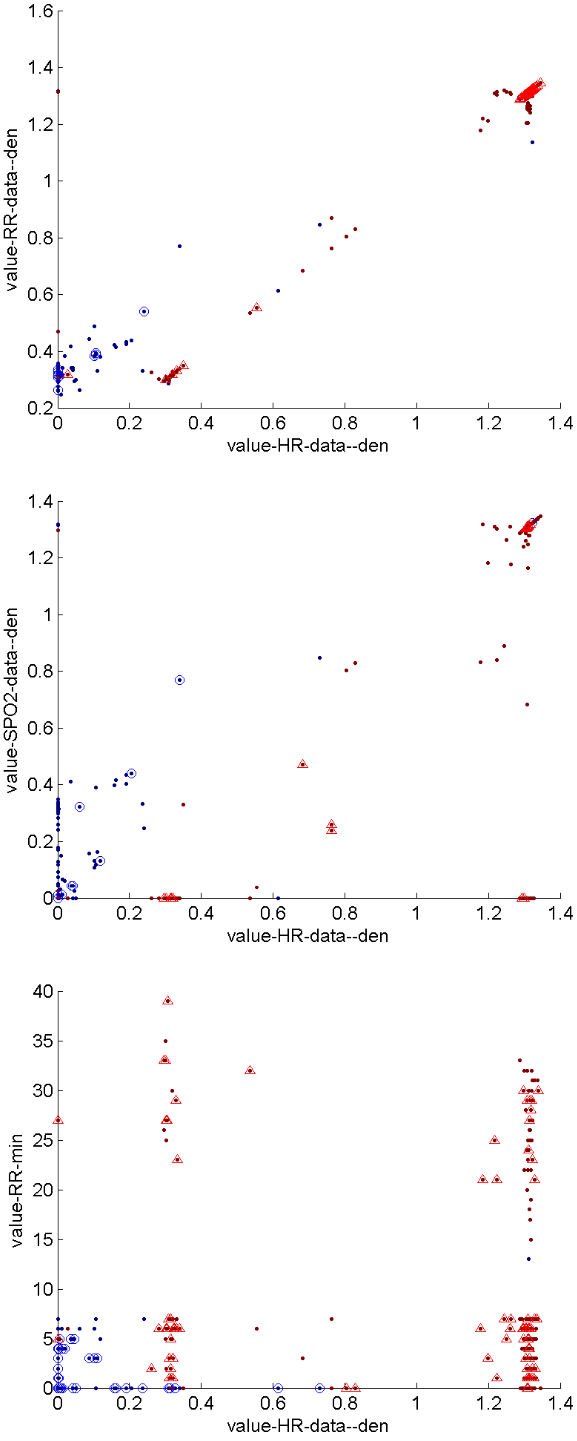
The majority of alerts in our data are associated with the respiratory rate (RR). There are 362 such cases and a significant proportion of these (132 samples) are actually artifacts. Figure 1 shows the set of 2-dimensional projections retrieved by RIPR for the true alarm vs. artifact classification task. All the data points are represented in the plot as dots - the true alerts are shown in blue while the artifacts are shown in red. Recall that each point is only classified using one projection. To illustrate this, we plotted the data assigned for each projection with red circles (for artifacts) and blue triangles (for true alerts). The plots show a good separation between artifacts and true alerts, which was one of our objectives. Also, the projections retrieved use data density features for the RR, SPO2 and HR signals as well as the minimum value for the respiratory rate. The use of these features is consistent with human intuition about what may constitute a respiratory rate artifact. For instance, a lot of missing data often signifies that the probe was removed from the patient for a period of time. The same can be said about minimum values for a vital - the measuring device could have been disconnected or misplaced.

4.2.2. Blood Pressure Alerts

We apply the same procedure for alerts related to blood pressure (BP) signals. There are 96 labeled examples of such alerts out of which 24 are artifacts. The 2-D projections are displayed in Figure 2. This time, though the features used are known to be informative, the class separation is not very clear. This is visible especially in the top right corner of the first plot, where we can observe a substantial overlap between artifacts and true alerts.

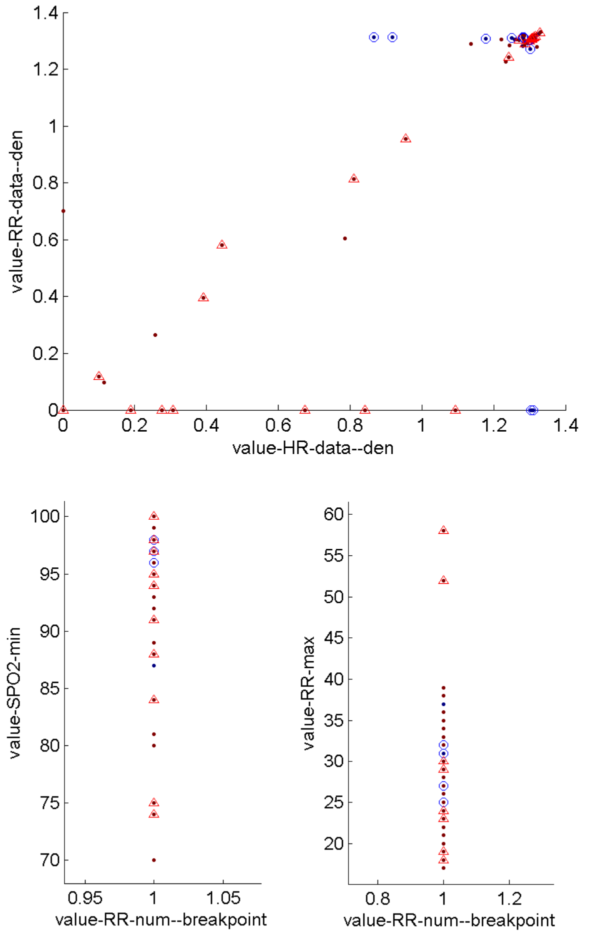
Since in this case using two-dimensional projections appears insufficient to provide a convincing model, we also identified informative 3-dimensional projections. Figure 3 shows the model resulting from this procedure. Only the alerts assigned to the specific projection were shown, in order to avoid overloading the figure. It is noticeable that the addition of the third dimension greatly improves the class separation. Again, the sparsity of data readings is an important feature, though this time the data density of three different

Figure 1. 2-D projections for RR alerts.The artifacts are in blue (circles) and the true alerts are in red (triangles).



vitals needs to be considered for the subset of data presented in the first projection of Figure 3. The second 3-D projection uses the maximum and minimum values of HR and RR to classify artifacts and there exists a hyperplane separating the two classes.

Figure 2. 2-D projections for BP alerts. The artifacts are in blue (circles) and the true alerts are in red (triangles).



4.2.3. Oxygen Saturation Alerts

The alerts based on blood oxygen saturation (SPO2) are more difficult to classify. The training set for these consists of 259 samples out of which only 24 are labeled as artifacts. Figure 4 shows the 2-dimensional projections recovered for this problem. As there is substantial class overlap, we also trained 3-D models, shown in Figure 5. Both 3-D projections of the model use data sparsity features to isolate artifacts, though we must note that the separation is still somewhat noisy.

The remaining alarms are associated with the heart rate. Only one of these is actually an artifact. Predictive accuracy of the presented RIPR models is summarized in Table 1. The results are obtained through leave-one-out cross-validation.

Figure 3. 3-D projections for BP alerts. The artifacts are in blue (circles) and the true alerts are in red (triangles).

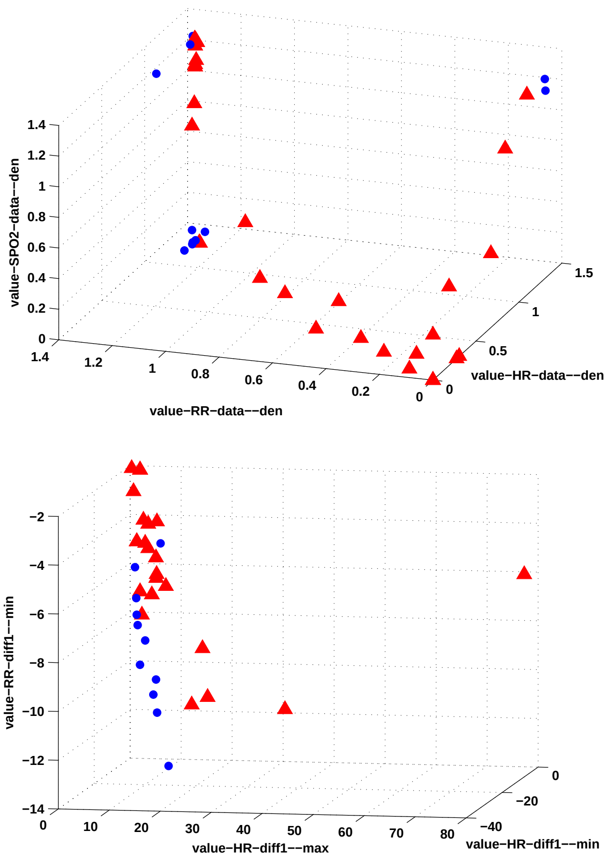
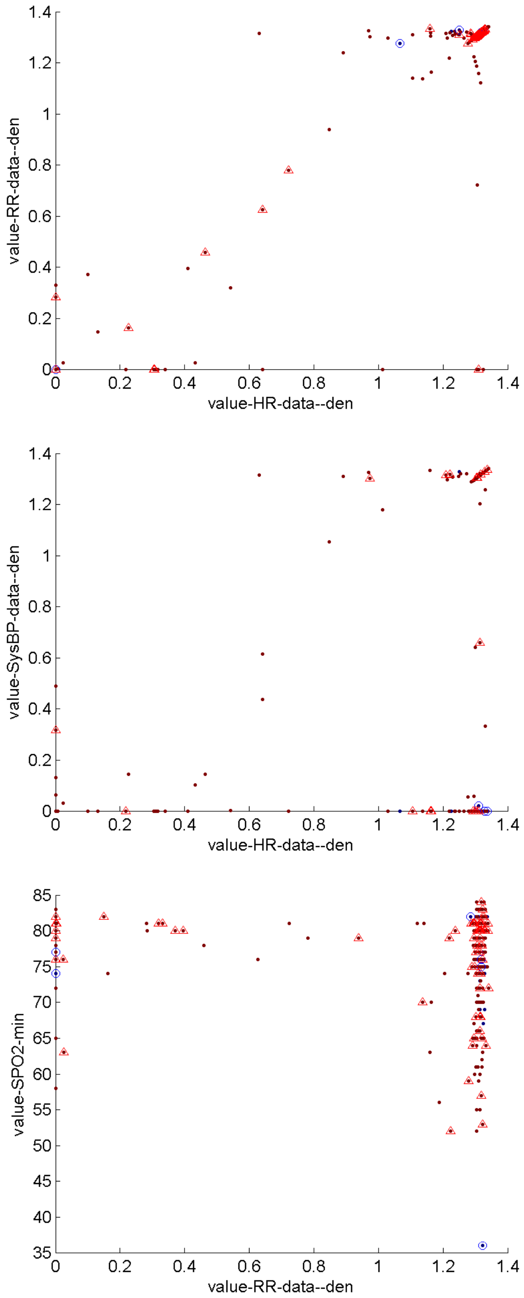


Figure 4. 2-D projections for SPO2 alerts. The artifacts are in blue (circles) and the true alerts are in red.



4.3. Case Studies

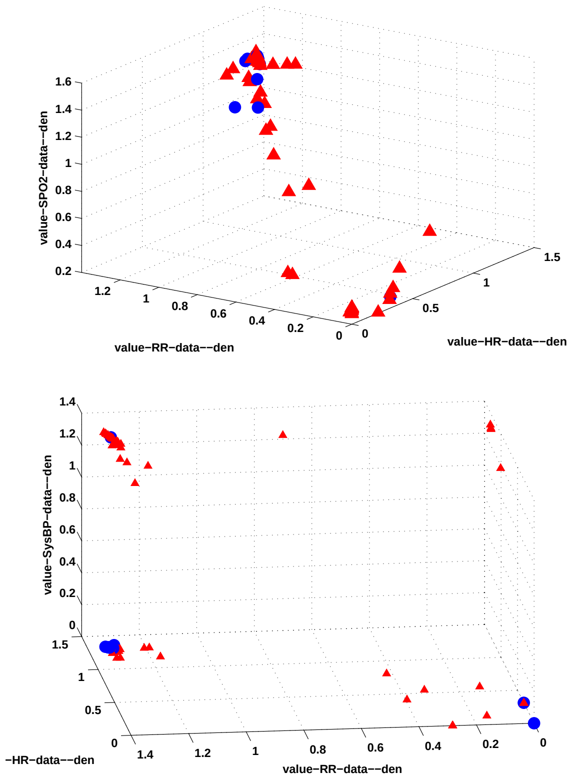
4.3.1. Outlier Detection

A good indication of the invalidity of a RR alert is the lack of HR data. So a simple decision rule - as stated by the clinicians - would be just to see whether there is HR data, if there is HR data, then the RR alert is an artifact, otherwise, it could be real. In classifying RR-based alerts, the algorithm correctly picked HR data density as the most important dimension. The top right of the second graph of Figure 1 contains two blue circles representing samples that would be classified as non artifact according to the projection. Both of them have continuous stream of data, but the RR signals are irregular. This is a different type of artifact. Because there are very few this type of artifact, and the algorithm is designed to retrieve a small set of projections, they end up being misclassified. The vital signs corresponding to these two samples are presented in Figure 6 below. Further investigation showed that

Table 1. Classification Accuracy of RIPR models. Precission and recall in recovering the genuine alerts.

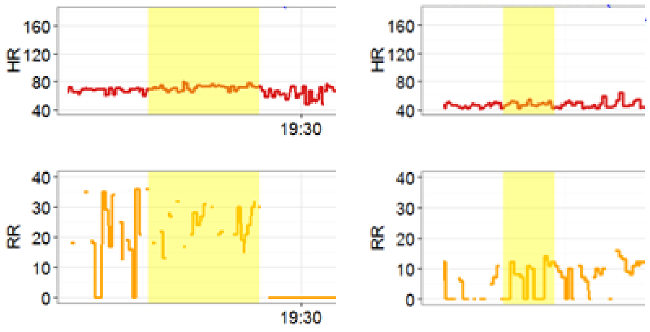
Type	RR 2D	BP 2D	BP 3D	SPO2 2D	SPO2 3D
Acc.	0.978	0.833	0.886	0.911	0.912
Prec.	0.979	0.858	0.896	0.929	0.918
Rec.	0.991	0.93	0.958	0.945 0.996	

Figure 5. 3-D projections for BP alerts. The artifacts are in blue (circles) and the true alerts are in red (triangles).



variance of the signal values provides a reliable way to detect these outliers. Thus, expert attention was focused on this more problematic type of artifact rather than on the type which represents the majority of cases and is relatively easy to handle automatically.

Figure 6. Vital signs of RR artifact outliers



4.3.2. Finding Errors in Data

On the other hand, some samples were classified by the system as artifacts while the domain experts considered them true alerts. On closer inspection, they seemed to exhibit artifact-like features - with little or no recorded values in the HR signal. When we drilled down to look at the data, we found that the samples were actually labeled incorrectly in the training set. Therefore, the RIPR approach can also be useful in detecting inconsistencies due to human error.

5. Conclusions

This paper outlines the use of a machine learning algorithm to support annotation of clinical data. We have shown the models that our RIPR framework produces for automatic data labeling and how the retrieved low-dimensional projections make it possible for domain experts to quickly validate the assigned labels. We also illustrated how RIPR models can be used to find special cases and incomplete or invalid data. Thus, the proposed framework promises to be useful to clinicians by partially automating annotation of medical data in a hman understandable and intuitive manner.

References

Fiterau, M. and Dubrawski, A. Projection retrieval for classification. In *Advances in Neural Information Processing Systems (NIPS 2012)*, volume 24, 2012.

Friedman, Jerome H., Kohavi, Ron, and Yun, Yeogirl. Lazy decision trees, 1996.

Gammerman, A., Vovk, V., and Vapnik, V. Learning by transduction. In *In Uncertainty in Artificial Intelligence*, pp. 148–155. Morgan Kaufmann, 1998.

Marchand, Mario and Sokolova, Marina. Learning with decision lists of data-dependent features. *JOURNAL OF MACHINE LEARNING REASEARCH*, 6, 2005.

Obozinski, Guillaume, Taskar, Ben, and Jordan, Michael I. Joint covariate selection and joint subspace selection for multiple classification problems. *Statistics and Computing*, 20(2):231–252, April 2010. ISSN 0960-3174. doi: 10.1007/s11222-008-9111-x. URL http://dx.doi.org/10.1007/s11222-008-9111-x.

Poczos, B. and Schneider, J. On the estimation of alpha-divergences. *AISTATS*, 2011.

Ting, Kai, Wells, Jonathan, Tan, Swee, Teng, Shyh, and Webb, Geoffrey. Feature-subspace aggregating: ensembles for stable andunstable learners. *Machine Learning*, 82:375–397, 2011. ISSN 0885-6125. URL http://dx.doi.org/10.1007/s10994-010-5224-5. 10.1007/s10994-010-5224-5.