A Novel Algorithm to Reduce Machine Learning Efforts in Real-Time Sensor Data Analysis

Majid Janidarmian(✉), Atena Roshan Fekr, Katarzyna Radecka, and Zeljko Zilic

Electrical and Computer Engineering Department, McGill University, Montréal, QC, Canada
{majid.janidarmian,atena.roshanfekr}@mail.mcgill.ca,
{katarzyna.radecka,zeljko.zilic}@mcgill.ca

Abstract. In the fitness and health fields, wearable sensors generate massive amount of information in big data. The machine learning techniques use the data to assess individuals' health in real time and identify trends that may lead to better diagnoses and treatments. Applying efficient algorithms to learn from data can aid physicians to evaluate the state of human actions and diagnose the illnesses. The process of discerning valuable information from wearable sensors is a non-trivial task and is an on-going research area. Many research areas have focused on machine learning-based approaches to sensor data for better understanding and meeting people's needs. However, there are different challenges such as runtime complexity and the number of functions calls associated with these approaches limit us to reach an acceptable accuracy level. To reduce the computational costs of the feature extraction and classification, a novel algorithm is proposed to analyze the variations in the periodic signals. It reduces the learning efforts by detecting any significant changes in the signal. We used the idea of pheromone trail employed in the ant colony optimization algorithm to keep track of the signal updates. The findings of this paper enable the design of a highly effective real-time predictive model for wearable applications.

Keywords: Wearable sensors · Machine learning · Respiratory disorder
Ant colony optimization

1 Background

The wearable sensors, coupled with the advanced data processing and communication technologies have opened the window to a new era of cost-effective remote healthcare services. Recently, much of research has focused on machine learning approaches to sensor data for delivering more intelligence into different health and fitness applications. They enable the remote monitoring of physical activity, vital signals, the early diagnosis of serious conditions, and the remote control of medical treatments [1]. In real-time data analysis, the runtime complexity of the machine learning models and the number of functions calls are important challenges as the whole recognition procedure should quickly handle the online data processing requirements. Although the techniques presented in this paper can be used for different applications, we focus on building an efficient algorithm to analyze accelerometer sensor mounted on the rib cage while

P. Perego et al. (Eds.): MobiHealth 2017, LNICST 247, pp. 83–90, 2018.
https://doi.org/10.1007/978-3-319-98551-0_10

capturing breathing patterns. The wearable motion sensor can be used to detect the small movements of the chest wall that occur during expansion and contraction of the lungs. It has been shown that with proper signal processing, this approach can produce results that closely match the measurements of nasal cannula pressure [2]. For example, the designed system in [3] used accelerometer sensors for diagnosis and treatment of patients with disordered breathing. This method shows a great potential to integrate the use of inertial sensors with machine learning techniques to model a broad range of human respiratory patterns including normal, Bradypnea [4], Tachypnea [5], Kussmaul [6], Cheyn–stokes [7], OSA [8], Biot's breathing [9], Sighing [10], and Apneustic [11] for the goal of cloud-based recognition of different respiratory problems. In addition, accelerometer-derived respiration signal has been proven itself particularly effective in providing an affordable platform for yogic breathing practices [12]. A disordered breathing pattern denotes inefficient oxygen inhalation and carbon dioxide expulsion from the body's tissues. The abnormal respiration is indicative of many diseases such as anemia, asthma, sleep apnea, sudden death syndrome, Chronic Heart Failure (CHF) and Chronic Obstructive Pulmonary Disease (COPD) [7]. For more details, the reader is referred to [1]. Figure 1 shows 30-s samples of eight respiration patterns derived from the accelerometer sensor. In this paper, we introduce an innovative technique that helps distinguish different patterns through analysis of dynamical characteristics of sensors data. It is an effective way to speed up the conventional recognition methods by reducing the number of calls of feature extraction and classification functions. It significantly reduces the usage of classification methods that require computationally expensive algorithms. The proposed algorithm is based on recurrence plot concept, which is the visualization of a square recurrence matrix of distance elements within a cut-off limit [13]. We use the higher-dimensional reconstruction by the method of time delays presented in [14]. From delayed coordinates of a signal $x(t)$, a pseudo-state space can be reconstructed as Eq. 1.

$$y(t_i) = \left[x(t_i), x(t_i + \tau), x(t_i + 2\tau), \ldots, x(t_i + (D-1)\tau)\right] \tag{1}$$

Where $y(t_i)$ is D-dimensional time-delayed vector of D points that are delayed or offset in time (τ). As shown in Fig. 2, the lag-reconstructed the acceleration vector will provide space-time information for nine breathing patterns performed by a subject. In this figure, each z-axis accelerometer signal is promoted into 3-dimensional space (D = 3) and therefore plotted against itself twice delayed (τ and 2 τ) on a three-axis plot ($\tau = 0.6$ s). As seen, the shape of the trajectory shows the periodic nature and dynamic of breathing patterns. The embedding dimension (D) and delay (τ) between sequential time points in the 1-dimensional signal has to be chosen with a preceding analysis of the data. In this study, the Mutual Information (MI) method [15] and the False Nearest Neighbors (FNN) [16] are employed to estimate the time delay and the embedding dimension, correspondingly. The most common method for choosing a proper time delay is based on finding the first local minimum of the Mutual Information (MI), defined as Eq. 2.

$$M(\tau) = \int_t p(t, t+\tau) log \frac{p(t, t+\tau)}{p(t)p(t+\tau)} dt \tag{2}$$

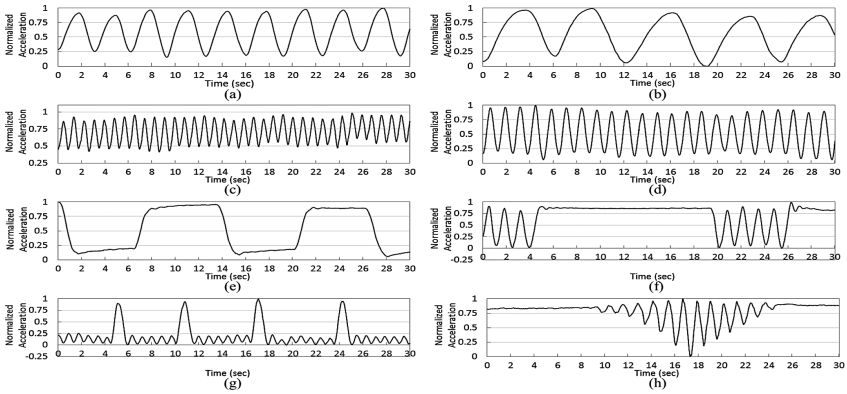

Fig. 1. (a) Normal, (b) Bradypnea, (c) Tachypnea, (d) Kussmaul, (e) Apneustic, (f) Biot's, (g) Sighing and (h) Cheyn-stokes breathing patterns from accelerometer sensor mounted on the subject's rib cage

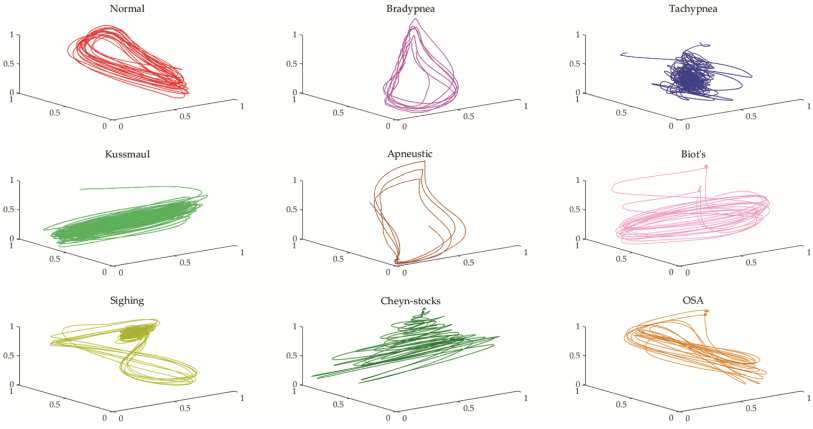

Fig. 2. Delay representation of normalized human breathing signals with embedding dimension 3 and tau 0.6 s.

Where $p(t, \tau)$ is the joint density function, and $p(t)$ and $p(t + \tau)$ are marginal density functions of $x(t)$ and $x(t + \tau)$, respectively [16]. The first minimum of the MI denotes the time delay, where the signal $x(t + \tau)$ adds maximal information to the knowledge obtained from $x(t)$ [15]. The time delay should be selected in a way that the reconstructed vector serves as independent coordinates while keeping the connection with each other. After finding a proper time delay, the embedding dimension can be calculated. In our algorithm, we used the nearest-neighbor methodology to find the embedding dimension. It steadily increases the embedding dimension and checks whether the neighborhood of all points in the phase space change. The algorithm stops where the amount of false nearest neighbors becomes almost unchanging. In another word, the information of the

system has been completely maximized and no new information can be gained by trying higher dimension [16].

2 The Proposed Method

Algorithm 1 describes how we used embedding dimension (*dims*) and time delay (*taus*) parameters to optimize machine learning calls during real-time classification problems. According to the employed methods explained above, the time delay and embedding dimension for each label are determined in line 2. When a new sensor sample is ready at time $t > (D-1)\tau$, a set of points, $\{points(t-(D-1)\tau), \ldots points(t-2\tau), points(t-\tau), \ldots, points(t)\}$ are defined in line 12. We use binary code to represent positive and negative directions for each point from time $t-1$ to t. For example, if the changes in three points values lie in the $(-, +, +)$ directions, the binary coding for this will be (011) and the resulting decimal number is three (lines 11–17). The obtained decimal number denotes the state number and consequently there is a transition between states once a new sensor sample arrives. We expect to see very similar transitions over time while we are dealing with a periodic signal such as walking or normal breathing. To keep track of the previous recent changes, we use the idea of pheromone trail employed in ant colony optimization algorithm. For example, if the current state is three and the next changes lie in the $(+, -, +)$ directions, binary coding for the new state will be (101) which is five. Therefore, for the transition $3 \rightarrow 5$, a specific amount of pheromone, Δ, will be deposited in line 28. In our experiments, initially there is no pheromone associated to each transition and the pheromone trails in each iteration (every $\frac{1}{f}$ sec) are updated by applying the evaporation technique as follows (line 29):

$$pt_{i,j} \leftarrow (1-\rho)pt_{i,j}, \forall(i,j) \tag{3}$$

Where $pt_{i,j}$ is the existing pheromone trail between state i and j. ρ is the pheromone evaporation coefficient which satisfies $0 < \rho \leq 1$ and is set to $\frac{1}{pt_{i,j}}$ in our study. The chosen value is high enough for a fair adaptation in the underlying problems. However, we believe that it should be experimentally determined under different scenarios. Thus, we read and write pheromones to track the signal behaviors, and more pheromones on each transition increase the probability of that transition being seen. Each new transition is counted, and accordingly, the number of unseen events (when there exists no pheromone associated with the new transition) is updated. However, we need to control the sensitivity of the system to avoid the transient noisy behavior. If the number of detected unseen events is more than the predetermined sensitivity, there exists a major change in the pattern, and the algorithm asks for running machine learning algorithms (*runClassifier*) in lines 22–27. The variable *freshWindow* (line 9 and line 23) is defined to control the number of calls as we have maximum one call for each new window of data according to the frequency and overlap value. If a new pattern keeps occurring, the algorithm will quickly adapt to new state transitions and stop calling the machine learning procedure until it detects a major change in the periodicity of the new pattern.

Algorithm 1: withSmartCalls

Input: *streamData*, *windowSize, overlap, freq, cModel,* $\Delta, \rho,$ *sensitivity, trainingData* (*labels*, :), *maxTau*, *minDim, maxDim*; **output:** *improvement*

1. newTransitions ← 0*; smartCalls* ← 0*; preState* ← 1*; freshWindow* ← *false; numberOfSamples* ← 0*;*

2. [*taus*, *dims*] ← **Find_Taus_Dimensions** (*trainingData* (*labels*, :), *maxTau*, *minDim*, *maxDim*);

3. samplesToWait = **floor** (*windowSize* × *freq* × (1 – *overlap*))*;*

4. **while** (~stop) { % It continues till sensor data is coming

5. **while** (~ **ready** (*streamData.newInstance*));
% Wait till there is a new sample of data (after resampling and filtering)

6. *numberOfSamples = numberOfSamples* + 1;

7. **if mod** (*numberOfSamples, samplesToWait*) == 0
% If we have enough data to run the classifier again

8. *endIndex = numberOfSamples;*

9. *freshWindow* ← *true*;

10. *conventionalCalls* ← *conventionalCalls* + 1;}

11. **for** *i = 1:dim* {

12. *points* (*i*) ← (*end* – (*i* – 1) × *tau*);
% The points values are increased by 1. "end" is the index of the last sensor instance.

13. **if** *streamData* (*points* (*i*)) <= (*streamData* (*points* (*i*) - 1)

14. *bin* (*i*) ← 0;

15. **else**

16. *bin* (*i*) ← 1; }}

17. *nextState* ← **bin2Dec** (*bin*); % Convert binary to decimal

18. **if** *transPheromones* (*nextState, preState*) == 0 {

19. *newTransitions* ← **min** (*newTransitions* + 1, 10);

20. **else**

21. *newTransitions* ← **max** (*newTransitions* – 1, 0); }

22. **if** *newTransitions* > 10 – *sensitivity*) && *freshWindow {*

23. *freshWindow* ← *false;*

24. *featuresVector* ← **featuresExtraction** (*streamData* (*endIndex* – *windowSize* × *freq* + 1: *endIndex*));

25. *label* ← **runClassifier** (*cModel*, *featuresVector*);

26. *tau* ← *taus* (*label*); *dim* ← *dims* (*label*);

27. *smartCalls* ← *smartCalls* + 1; }

28. *transPheromones* (*nextState*, *preState*) ← Δ;
% Δ is the amount of pheromone deposited for the most recent state transition

29. *transPheromones* (:, :) ← **max** ((1 – *ρ*) × *transPheromones* (:, :), 0) ;

30. *preState* ← *nextState*; }

31. *improvement* ← ((*conventionalCalls* – *smartCalls*) / *conventionalCalls*) × 100

32. **return** *improvement;*

Algorithm 2: Find_Taus_Dimensions

Input: *training Data* (*labels*, :), *maxTau*, *minDim, maxDim*; **output:** *taus, dims*
1. **for** *i* = *labels* {
2. *MI* ← **mutualInformation** (*trainingData* (*i*, :), *maxTau*);
3. *valleysLocations* ← **valleysFinder** (*MI*);
4. *taus* (*i*) ← *valleysLocations* (1) ; % First local minimum of the mutual information (MI)
5. *dims* (*i*) ← **falseNearestNeighbors** (*trainingData* (*i*, :)*, minDim, maxDim*); }
6. **return** *taus*, *dims*;

3 Experimental Results

The evaluation was performed on data from 10 healthy volunteers, five males and five females aged 27 to 48 with (Mean ± SD) 34.80 ± 6.89. The tests lasted for about 35 min per subject. The ethical approval was received from McGill University Ethics Committee. All participants were informed about the experimental procedures before starting the trial sessions. The subjects were asked to perform nine introduced breathing patterns, each for 1 min in sitting position (torso at about 90° angle to the floor). For simulating apnea in Cheyn-stokes, Biot's and OSA breathing exercises, the subjects paused their breathing for at least 10 s. We asked the participants to prolong their inspiration and expiration during Apneustic maneuver for at least 5 s. Finally, for the Sighing pattern, they performed normal breathing, which is followed by deep periodic of inspiration every 3–7 s. OSA breathing pattern is similar to Biot's breathing pattern; however, it has a different phase shift between chest and abdomen compared to Biot's breathing. The SPR-BTA spirometer [17] is also used in all tests to make sure that the subjects were not over emphasizing the breathing movements. The LIS3DH 3-axis accelerometer with 12-bit resolution is used and secured by a soft and elastic strap which is easy to attach and comfortable to wear. The sensor is mounted on the subject's chest in the middle of sternum region. In our tests, the sensor is sampling with 50 Hz. The proposed algorithm is validated on breathing disorder classification in which we deal with 1D motion signal (z-axis). Figure 3 plots a raw breathing signal for three different patterns. In this example, the time delay $\tau = 2$ s is selected for the phase space reconstruction. Given the time delay, we take the embedding dimension as 4 for the windowed breathing signal. Therefore, we have $2^4 \times 2^4$ states. This figure also shows the amounts of pheromones ($\Delta = 12$) in each transition at nine different moments. The proposed technique can detect any major changes or motion artifacts in the signal. Figure 4 shows the improvement in number of machine learning functions calls in each breathing pattern performed by different subjects. In average, the number of functions calls reduced by in average 52.75%, 47.37%, 42.71%, 48.35%, 19.28%, 10.69%, 20.21%, 9.51%, 20.43% for Normal, Bradypnea, Tachypnea, Kussmaul, Apneustic, Biot's, Sighing, Cheyn-stocks and OSA breathing patterns, respectively. The results indicated an average improvement

of more than 31% on all different breathing maneuvers with no reduction in classification accuracy.

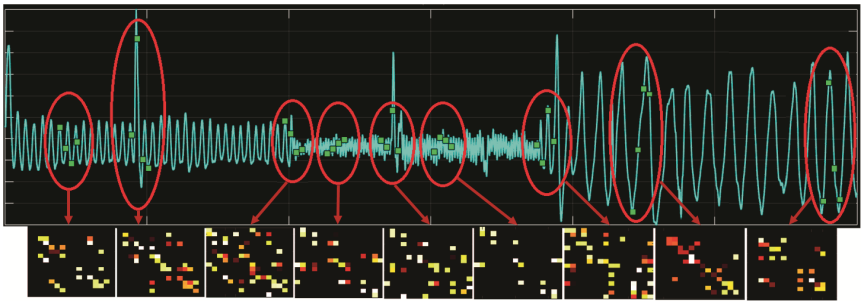

Fig. 3. Raw breathing signal for three different patterns and the pheromone trail updates during the procedure. We show the *transPhermone* updates at nine different moments. As the embedding dimension is four in this test, the size of *transPhermone* is 16×16.

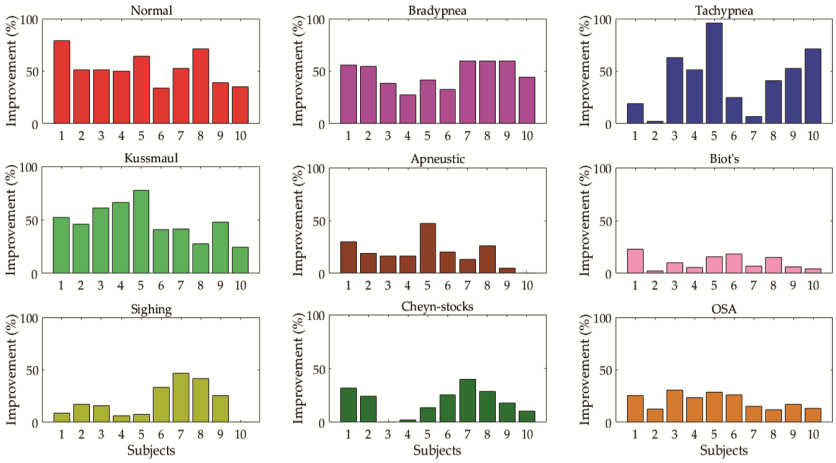

Fig. 4. The improvement in number of machine learning functions calls in each breathing pattern performed by each subject

4 Conclusion

We proposed an innovative approach to speed up the conventional recognition methods by reducing the number of calls of feature extraction and classification functions. It is a very fast algorithm to analyze the dynamical characteristics of sensor data at each sample to detect any significant change in the signal. When working with breathing data derived from the accelerometer sensor, an average improvement of more than 31% was obtained. This finding enables the design of a highly effective real-time predictive model for wearable applications.

References

1. Janidarmian, M., Roshan Fekr, A., Radecka, K., Zilic, Z.: Multi-objective hierarchical classification using wearable sensors in a health application. IEEE Sens. J. **17**(5), 1421–1433 (2017)
2. Bates, A., Ling, M.J., Mann, J., Arvind, D.K.: Respiratory rate and flow waveform estimation from tri-axial accelerometer data. In: International Conference on Body Sensor Networks (BSN), pp. 144–150 (2010)
3. Roshan Fekr, A., Janidarmian, M., Radecka, K., Zilic, Z.: A medical cloud-based platform for respiration rate measurement and hierarchical classification of breath disorders. Sensors **14**, 11204–11224 (2014)
4. Kou, Y.R., Shuei Lin, Y.: Bradypnea. In: Lang, F. (ed.) Encyclopedia of Molecular Mechanisms of Disease, pp. 241–243. Springer, New York (2009). https://doi.org/10.1007/978-3-540-29676-8
5. William, H., Myron, L., Robin, D., Mark, A.: Current Diagnosis and Treatment in Pediatrics, 21st edn, p. 989. McGraw-Hill Professional, New York (2002)
6. Kaufmann, P., Smolle, K.H., Fleck, S., Lueger, A.: Ketoacidotic diabetic metabolic dysregulation: pathophysiology, clinical aspects, diagnosis and therapy. Wien. Klin. Wochenschr. **106**, 119–127 (1994)
7. Brack, T., Thüer, I., Clarenbach, C.F., et al.: Daytime cheyne-stokes respiration in ambulatory patients with severe congestive heart failure is associated with increased mortality. Chest J. **132**(5), 1463–1471 (2007)
8. Hofsoy, D.A., Clauss, J.F., Wolf, B.: Monitoring and therapy of sleep related breathing disorders. In: Proceedings of the 6th International Workshop Wearable Micro and Nano Technologies for Presonalized Health, pp. 41–44, June 2009
9. Farney, R.J., Walker, J.M., Boyle, K.M., Cloward, T.V., Shilling, K.C.: Adaptive servoventilation (ASV) in patients with sleep disordered breathing associated with chronic opioid medications for non-malignant pain. J. Clin. Sleep Med. **4**, 311–319 (2008)
10. Aljadeff, G., Molho, M., Katz, I., Benzaray, S., Yemini, Z., Shiner, R.J.: Pattern of lung volumes in patients with sighing breathing. Thorax J. **48**, 809–811 (1993)
11. Respiratory center, The American Heritage R_Medical Dictionary (2007)
12. Janidarmian, M., Fekr, A.R., Radecka, K., Zilic, Z.: Haptic feedback and human performance in a wearable sensor system. In: 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Las Vegas, NV, pp. 620–624 (2016)
13. Webber, C.L., Zbilut, J.P.: Recurrence quantification analysis of nonlinear dynamical systems. In: Riley, M.A., Van Orden, G.C. (eds.) Tutorials in Contemporary Nonlinear Methods for the Behavioural Sciences, National Science Foundation, Arlington, VA, pp. 26–95 (2005)
14. Takens, F.: Detecting strange attractors in turbulence. In: Rand, D., Young, L.-S. (eds.) Dynamical Systems and Turbulence, Warwick 1980. LNM, vol. 898, pp. 366–381. Springer, Heidelberg (1981). https://doi.org/10.1007/BFb0091924
15. Fraser, A.M., Swinney, H.L.: Independent coordinates for strange attractors from mutual information. Phys. Rev. A **33**, 1134 (1986)
16. Kennel, M.B., Brown, B., Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys. Rev. A **45**, 3403–3411 (1992)
17. Spirometer. http://www.vernier.com/products/sensors/spr-bta/