Modeling Long-Term Human Activeness Using Recurrent Neural Networks for Biometric Data

Abstract

This paper explores the feasibility of modeling a person's "activeness" using biometric data retrieved from a fitness tracker. Currently, the notion of activeness of a user at a given period time is defined to be a tuple of three types of biometric data: *heart rate*, *consumed calories*, and the *number of steps taken.* Four recurrent neural network (RNN) architectures are proposed to investigate the performance on predicting the activeness of the user under various length-related hyper-parameter settings. The dataset used in this study consists of several months of biometric time series data gathered by seven users independently. The experimental results show that forecasting the users' activeness is indeed feasible under suitable lengths of input and output sequences.

1. Introduction

With the advances in technology and ever-busy schedules, people tend to lack physical activity, and have increased level of stress. They are hence at a greater risk of suffering from the so-called "modern diseases" such as cardiovascular disease, diabetes, metabolic disorders, and stroke, (Capon, 2012). Attaining a healthy lifestyle, which incorporates a balanced diet and a plenty of exercise, is considered to be key in preventing such diseases.

Recently, many health-care-related devices and services have emerged to aid users in monitoring and improving their physical wellness (Section 2.1). With wearable devices, such as fitness trackers, it has been possible to continuously observe the biometric data produced by a user, and notify the user when he/she has been physically inactive for a period of time. Many services also provide users with general tips on a healthy lifestyle, and motivation to be physically more active during the day, for example, by letting them know how many steps remain to reach the weekly average, or by offering them virtual "badges" to commemorate their physical achievements which can be boasted over a social media platform.

While some of these approaches have been considered to be effective by their users (Findley, 2015), this paper suggests that their usefulness can be further improved if a *long-term predictive model* of the user's "activeness" is incorporated into the health-care services. For example, an application can project the user's activeness for some period of time in the future, and inform him/her of the remaining days before the weight loss goal is (or not) reached. In addition, it may take a more proactive measure, depending on the user's context, and preemptively recommend possible exercises that he/she could perform when the activeness is predicted to be below a threshold.

Many research efforts have been made to accurately model and predict users' heart rates (Cheng et al., 2008; Austin et al., 2013; Sumida et al., 2013; Lipton et al., 2016) and energy expenditures (Keytel et al., 2005; Pande et al., 2013; Bouarfa et al., 2014), often as a means to recognize their simple activities (e.g. walk, run, lying down, etc.) (Sumida et al., 2013) or to identify any medically significant event such as heart failure (Austin et al., 2013; Zheng et al., 2014; Lipton et al., 2016).

As the task of activity recognition or detection of heart failure often involves classifying a relatively short span of time, most existing works utilize machine learning algorithms such as feed-forward neural networks (FFNNs), support vector machines (SVMs), and random forests (RFs) that are known to be effective in learning short-term temporal dependencies among time series data. Furthermore, these works often employ wearable sensors that are specifically designed for a certain type of biometric data, and focus on building an accurate model for the type of data and the task at hand.

In this paper, we slightly shift the perspective, and aim to investigate the feasibility of modeling a user's *long-term* activeness which could, to some extent, represent his/her lifestyle pattern. Currently, our notion of activeness for a given period of time is tracked as a tuple of heart rate, consumed calories, and the number of steps taken by the user.

Instead of utilizing separate wearable sensors for each type of data, a fitness tracker is used to continuously record the three types of biometric data of the user for several months. Along with the three classification models (FFNN, SVM, and RF), we experiment with recurrent neural network (RNN) architectures which are considered to be well suited for learning long-term dependencies among temporal data. While there are many studies of RNN architectures being applied to various sequential modeling tasks—e.g. the stock market (Yaya et al., 2013), energy consumption (Marvugliaa and Messineo, 2012), genetic expression (Noman et al., 2013), speech (Eyben et al., 2013), and language modeling (Mikolov et al., 2010)—few works exists in the domain of wellness modeling. Therefore, this paper explores how the performance of activeness prediction is varied by changing (1) a set of length-related parameters of the training process, and (2) RNN architectures. Experimental results show that the model can learn and predict the user-specific activeness patterns effectively.

The rest of the paper is organized as follows. Section 2 explores the background for this study, while Section 3 describes how the time series datasets are gathered. We illustrate the proposed approach in Section 4, and the experimental results, in Section 5. Finally, the paper is concluded in Section 6 with some directions for future works.

2. Background

This section briefly introduces some of the commercial devices and services that are proposed to measure and improve a user's "wellness" (Section 2.1), along with some academic research that aim to model biometric data for various tasks (Section 2.2). Moreover, existing works that involve time series modeling using RNN architectures are presented in Section 2.3.

2.1 Devices and services for wellness improvement

According to the Centers for Disease Control and Prevention, USA, 70.7% of American adults over the age of 20 are overweight, and 37.9% of the same group are obese as of 2013-2014 (Centers for Disease Control and Prevention, 2016). As a need for a "fitness revolution" is greater than ever before, fitness devices and services are flooding the marketplace.

Since 2006, the footwear company Nike has introduced the "Nike+ Sports Kit" that records the distance and pace of a walk/run, and transmits the data to the user's smart device. A series of all-around activity trackers have been independently manufactured by Fitbit and Jawbone. These fitness trackers measure the number of steps taken and log the heart rates of the wearer. Based on these measurements (and other biometric information), the consumed calories and the traveled distance are calculated. This study utilizes Jawbone's "UP3" model and Fitbit's "Charge HR" model to continuously record users' heart rate, steps, and calories.

Several fitness centers and health-care providers have devised wellness "scores" or "indices" that aim to quantify the physical fitness of an individual. For example, Life Time Fitness proposes the "myHealthScore" (Life Time Fitness) that is determined by six indicators: blood pressure, triglycerides, total cholesterol to high-density lipoprotein ratio, glucose, body fat, and tobacco use.

Dacadoo introduces the "Health Score" (Dacadoo) which ranges from 1 to 1000, and is calculated from biometric values (gender, age, weight, waist circumference, blood pressure, etc.), emotional values (acquired from self-assessment questionnaires), and lifestyle values (exercise, nutrition, steps, sleep, etc.). Linking with the aforementioned fitness trackers, the Health Score continuously changes throughout the day as the user performs activities such as walking, running, sleeping, etc.

The "Wellness Score" (8 Weeks to Wellness) offered by 8 Weeks to Wellness ranges from 1 to 100, and is calculated using various biomarkers including: body mass index, posture number, core strength and flexibility, body fat percentage, and heart rate.

While these measures claim to represent an individual's state of wellness/health, how the corresponding factors are combined to produce a single value is not known publicly. Furthermore, there is not yet a general consensus even among doctors and medical researchers about what constitutes wellness and how they should be measured. For example, several key dimensions can exist to define wellness—physical, psychological/emotional, social, intellectual, spiritual, occupational, environmental, cultural, economic, and climate—and for each dimension, different researchers may view certain factors more important than other factors, and thus propose different scoring functions (Miller and Foster, 2010).

In addition, the holistic perspective of calculating a single wellness score is not fully grounded on medical examination; after all, the involved factors vary in both characteristics and units.

For these reasons, we specify that this study targets to model a person's physical "activeness", which is kept as a series of tuples of heart rate, consumed calories, and the number of steps, and avoid using the more general term, "wellness".

2.2 Modeling biometric data

While the term "biometric data" in the context of security, generally refers to measurable physical characteristics that help in authenticating an individual (e.g. fingerprint, retina, vein, etc.), this study refers to its more general meaning—the measurable biological quantities of an individual that, unlike the former kind, may change over time. This paper targets three types of biometric data—heart rate, consumed calories, and the number of steps—that reflect how physically active a person is for a given period of time.

The task of modeling human heart rates and energy expenditures (EE) has been widely studied across many disciplines such as sports science, medicine, electrical engineering, and computer science. Keytel et al. (2005) developed a prediction equation for EE from the heart rate by monitoring 115 regularly exercising individuals aged 18 to 45 years old.

The participants performed exercises on a treadmill, and their heart rate and respiratory exchange ratio data were collected. A mixed model analysis identified gender, heart rate, weight, maximal oxygen uptake, and age as important factors in estimating EE.

Cheng et al. (2008) proposed a non-linear state-space control system that modeled the heart rate of a person walking on a treadmill, and later utilized the model to build a computer-controlled treadmill system for regulating the heart rate during exercise.

Sumida et al. (2013) introduced an approach that predicted the heart rate of a walking user, utilizing the accelerometer and GPS data obtained from the user's smartphone. The authors used the raw data from the smartphones to calculate the oxygen uptake, which was then fed as a form of input data to an artificial FFNN.

Similarly, Pande et al. (2013) estimated EE for ambulatory activities using accelerometer and barometer sensors in a smartphone. Their model, which also used an artificial FFNN, outperformed calorimetry equations and EE values obtained from fitness trackers.

Bouarfa et al. (2014) targeted a slightly more general setting of estimating EE under "free-living" conditions using a single ear-worn accelerometer. The regression analysis was used to predict EE values while linear discriminant and nearest neighbor classifiers were employed to classify a window of accelerometer values into one of ten activities such as lying down, standing, computer work, vacuuming, etc. The regression model correlated well with the medical golden standard, the doubly labeled water test.

In addition to the general task of modeling heart rate and EE, some works specifically focus on medical problems such as heart failure detection. For example, Austin et al. (2013) compared the classification performance of several machine learning algorithms such as logistic regression, bagging, RF, and SVM, when applied to classifying patients with heart failure (HF) into one of two mutually exclusive subtypes: HF with preserved ejection fraction and HF with reduced ejection fraction. However, in their study, a set of detailed clinical data of patients was used as opposed to the time-series data.

More relevantly, Zheng et al. (2014) proposed a multi-channel deep convolutional neural network (MC-DCNN) for a time-series classification. Their model was applied to a set of electrocardiograph data which had been recorded from 15 patients suffering from severe congestive heart failure. The task was to classify a 2D time-series into one of four types of heartbeats. In their experiments, the proposed MC-DCNN performed better than the nearest neighbor approaches and FFNNs.

Our work is similar to the above studies in that we aim to model a person's heart rate, EE (consumed calories), and number of steps. However, while the above works generally consider time-series data with lengths from a few seconds to a dozen minutes, this work aims to discover a long-term temporal pattern by considering much longer periods of temporal data, ranging from a dozen minutes to days.

Our intuition is that, just as patients suffering from heart disease share a set of distinctive patterns in their electrocardiograph data, a cluster of people, for example, "morning people", may share similar trends in their biometric values when observed for a long period of time. Of course, the specific (short-term) details can depend on individuals and their daily schedules. Surprisingly, our experimental results (Section 5) show that an individually-tailored recurrent neural network model—with no additional information other than the biometric data—can predict such details quite closely.

2.3 Recurrent neural networks

Recurrent neural networks (RNN) represent a class of artificial neural networks where some connections between nodes form a directed cycle (Fig. 1). In essence, RNNs carry out the same task for every element of an input sequence, producing an output that is both dependent on the current input and the results from previous computations. Such recurrent connections enable RNNs to capture information about what has been calculated so far, thus exhibiting a dynamic temporal behavior.

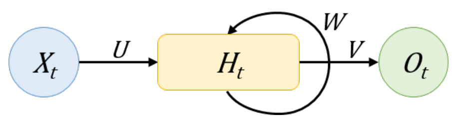


Figure 1: General structure of RNNs.

The recurrent behavior of RNNs has made them an effective solution for various tasks involving sequential data modeling such as for the stock market (Yaya et al., 2013), energy consumption (Marvugliaa and Messineo, 2012), genetic expression (Noman et al., 2013), speech (Eyben et al., 2013), and language modeling (Mikolov et al., 2010).

In the medical domain, RNNs are often used to model physiological signals such as electrocardiograms (Silipo and Marchesi, 1998; Fukuda et al., 2001; Übeyli, 2009). Recently, Lipton et al. (2016) applied long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) RNNs to the task of multilabel classification of multivariate clinical time series data. While the task was to predict the probability distribution of 128 labels (e.g. diabetes, asthma, scoliosis, neoplasm, etc.), the authors improved the performance of the model via auxiliary output training which utilized the remaining 301 diagnostic labels.

To the best of our knowledge, modeling a person's long-term "activeness" using RNN architectures has not been studied previously. This work also utilizes the LSTM cells to model the biometric data effectively, exploring the different layouts of networks and parameter settings.

We briefly mention that LSTM cells are a variation on the architecture of the vanilla RNNs, and designed to overcome the vanishing/exploding gradient problem often faced by the neural network family (Bengio et al., 1994). Greff et al. (2015) conducted a detailed empirical evaluation on the LSTM family.

3. Data Description

Our experiments utilize three types of biometric time series data, namely, heart rate, steps, and consumed calories. Seven graduate students between the ages of 23 to 33 years old participated their biometric data. It is noted that as the participants were all graduate students, the gathered data could be biased towards the group as less active as opposed to a more active group of "athletes" or "outdoor service employees".

However, a simple survey was conducted and revealed that the participants' lifestyle patterns were quite different to one another. For example, three participants described themselves as regularly exercising, while differing in the type and duration of the workouts. Also, the participants' bedtimes and wake-up times were not congruent as well. An extreme case was a participant who operated on a three-day cycle, where he stays up for two days and sleeps for the next entire day. Therefore, we designed experiments to observe if such lifestyles patterns of participants could be effectively learned by our models.

A Jawbone's UP3 fitness tracker was worn by each participant, and used to gather the three types of the biometric data. The device is equipped with a tri-axis accelerometer that detects physical movements, and a bio-impedance sensor that measures heart rate, respiration, and galvanic skin response. While the exact internal logic for the tracker is not known publicly, we believe that the consumed calories are calculated via its own energy expenditure equation that considers the wearer's age, body mass index, number of steps, and heart rate.

Recently, some criticisms have been made on the accuracy of the estimated calories (Feltman, 2014), pointing out that various fitness trackers compute different amount of burned calories when worn by the same user simultaneously. Nevertheless, we believe that the exact value of consumed calories was not so critical to the experiments as the main objective was to observe the long-term trend.

We also note that the tracker computes the heart rate in the form of beats per minute (BPM), but does not yet provide an application programming interface (API) for accessing every value of the recorded time series data (contrary to what was stated in the specification). Thus, we reverse-engineered the web application, and retrieved the individual's heart rates. As the heart rate data in BPM were recorded at an irregular interval (ranging from a few seconds to a few minutes), a linear interpolation was conducted to prepare the data in one-minute-intervals.

As for the consumed calories and number of steps, when an activity of an arbitrary duration was performed, the total sum of each type of data was recorded for that activity. Therefore, for each time stamp in the duration of the activity, we assigned the mean value of each type of data.

After interpolating each type of the time series data, the values underwent a min-max normalization in order to be scaled from 0 to 1. The minimum and maximum values for each type of data were selected by consulting relevant medical documents.

The statistics of the gathered data are presented in Table 1. In order to compare the lifestyle patterns of different users, we prepared the time series data to begin at the same time stamp. The number of samples represents the number of minutes in the recorded duration.

Table 1: Statistics of the Gathered Time Series Data for the Seven Users

User	Start Time	End Time	Duration (Months)	# of Samples
AK	2015-10-12 00:00:00	2016-05-04 14:21:00	6.85	296,062
HJ		2015-11-29 02:32:00	1.60	69,273
JM		2015-11-19 19:05:00	1.29	55,866
KJ		2016-05-02 23:57:00	6.80	293,758
YJ		2016-05-03 18:00:00	6.83	294,841
YS		2016-01-14 10:11:00	3.15	135,972
ZM		2016-04-15 06:53:00	6.21	268,254

4. Method

In this work, we are interested predicting the activeness of a person based solely on his/her previous data. As mentioned in Section 2.3, such task of temporal sequence modeling can be effectively conducted using RNN architectures. This study explores four RNN architectures described in Section 4.2.

4.1 Task settings

Before looking at the architectures, it is necessary to identify the hyper-parameters for training and evaluating an RNN model for activeness prediction, as they directly impact the performance of the model. Figure 2 illustrates the hyper-parameters for (a) training and (b) testing a network (prediction).

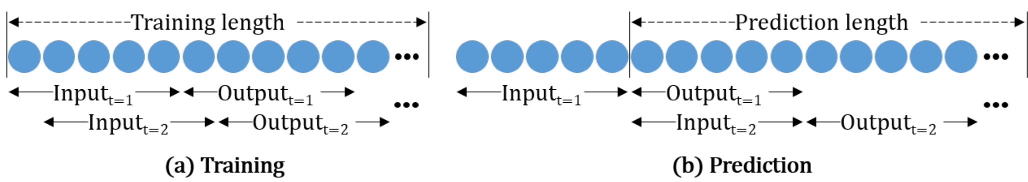


Figure 2: Length-related hyper-parameters for (a) training and (b) testing.

- **Training length** specifies the total length of the recorded time series data that are used to train the model.
- **Input length** corresponds to the number of time steps that the network takes as an input. When modeling a single type of biometric data, a memory cell receives a one-dimensional vector at a time step t, and updates its cell state using the current input vector x_t and the previous cell state c_{t-1}. Therefore, the input length n determines how many time steps are processed internally by the memory cell before producing an output vector o_t at $t = n$. In a typical case of language modeling, the input length is often set to the average (or maximum) sentence length. However, in our scenario of modeling activeness, it is not so apparent as to how long should the time steps be for each user. Hence, this study explores the variations on this parameter.
- **Output length** refers to the length of the time series data that the model is required to predict for a given input data.
- **Prediction length** represents the total length of the time series data that we want to predict.

We note that other network-based hyper-parameters such as regularization methods, size of hidden layers, and choices in loss and activation functions can also affect the network's performance. This study, however, focuses on exploring the impact of aforementioned length-related parameters on the activeness prediction.

4.2 Network architectures

As there are an infinite number of ways in structuring a neural network model, building an effective network architecture requires much practice and patience. In this study, we

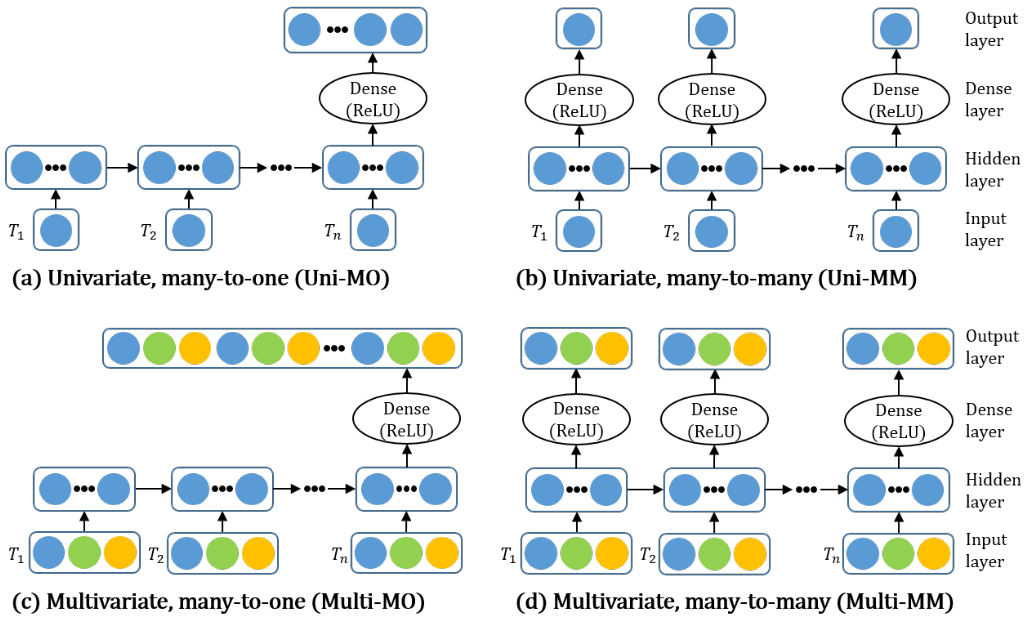


Figure 3: Four proposed RNN architectures.

experiment with the following four RNN architectures—univariate many-to-one (Uni-MO), univariate many-to-many (Uni-MM), multivariate many-to-one (Multi-MO), and multivariate many-to-many (Multi-MM)—depicted in Figure 3.

Architectures (a) and (c) are each formulated in the *many-to-one* fashion where the output is computed only at the last time step of an input data. Notice that these two structures offer more flexibility in choosing the output length than *many-to-many* approach in the sense that the output length can be different to the input length. In *many-to-many* approach, an output vector is computed at each time step, and is in the same dimension as the input vector. A *univariate* architecture models each type of biometric data separately, while a *multivariate* architecture considers the three types of data together in the same model.

5. Experiments

5.1 Choosing an RNN architecture

While separately conducting all experiments on each of the four RNN architectures would be ideal, due to time and resource constraints, we select the best performing one to be the specimen for the subsequent experiments. We individually train each RNN architecture using the first *one month* of the time series data, and generate the time series data for *the next week*.

The effectiveness of each architecture is evaluated by grid searching the network-based hyper-parameters such as the number of hidden units, dropout rate, and choice of activation function. We found that LSTM cells with 64 hidden units were best for the univariate architectures in general, while the memory cells with 192 hidden units were effective for the multivariate ones. Each architecture was trained to minimize the mean squared error (MSE) using a recently proposed optimization method called *Adam* (Kingma and Ba, 2014). We also note that a *dropout rate* of 0.2 was used in every layer; and *rectified linear unit* was chosen to be the non-linear activation function for the fully-connected (dense) layers.

The top-5 results were a *multivariate many-to-one* (Multi-MO) architecture with 192 hidden units ($MSE = 0.00140$), followed by three *univariate many-to-one* (Uni-MO) architectures with 64 (0.00152), 128 (0.00156), and 256 (0.00177) hidden units, and a *univariate many-to-many* (Uni-MM) architecture with 128 hidden units (0.00202).

However, as we did not exhaustively search the network-based hyper-parameter settings, we cannot yet to claim the superiority of one architecture over another. Nevertheless, following the results, we selected the Multi-MO model with 192 hidden units to be the specimen model for all subsequent experiments.

5.2 Effect of varying length parameters

Four experiments were conducted to evaluate the effect of the four length-related parameters on the performance of activeness prediction. We specify that when one parameter was varied, the other three parameters were fixed as follows: training length = 1 month, input length = 15 mins, output length = 15 mins, and prediction length = 7 days.

Table 2 shows the average MSEs for each experiment where Min, 1Q, Median, 3Q, and Max represent the minimum, first quartile, median, third quartile, and maximum MSE values respectively. We can observe that the prediction error is generally reduced when *more training data* are used to predict a *shorter period of time* using *shorter input* and *output* time steps. Interestingly, we note that the minimum MSE for the experiment of one-day prediction is particularly high; and this was due to one person who had been ill during the day, i.e., the day that was selected to evaluate the model was quite dissimilar from the person's "typical" day.

Table 2: Average MSEs when Varying the Length Parameters

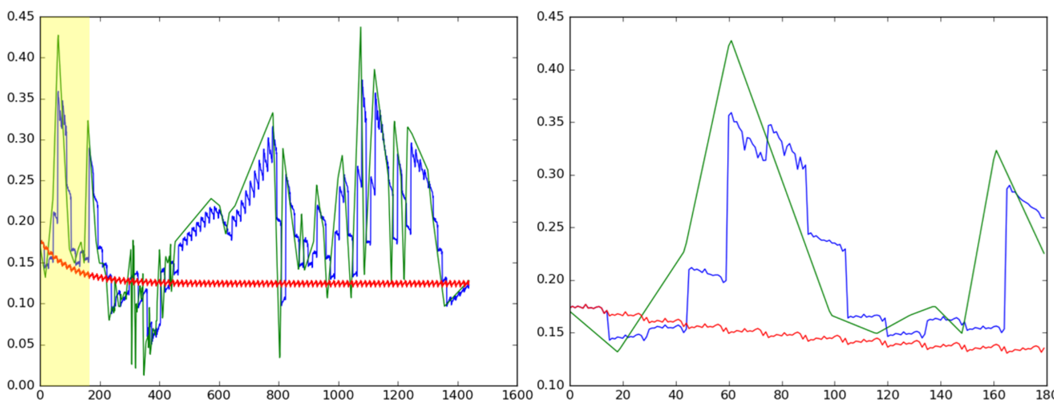
Experiment	**Variation**	**Min**	**1Q**	**Median**	**3Q**	**Max**
Training length (# of users: 7)	1 Day	0.00774	0.01242	0.01365	0.01410	0.01638
	1 Week	0.00160	0.00176	0.00252	0.00303	0.00606
	1 Month	0.00086	0.00111	0.00140	0.00158	0.00203
Prediction length (# of users: 5)	1 Day	0.00043	0.00081	0.00106	0.00176	0.00524
	2 Weeks	0.00026	0.00084	0.00134	0.00224	0.04365
	1 Month	0.00045	0.00106	0.00152	0.00224	0.05310
Input length & output length (# of users: 7)	30 Mins	0.00044	0.00078	0.00131	0.00226	0.00523
	1 Hour	0.00043	0.00151	0.00220	0.00300	0.01734
	6 Hour	0.00086	0.00199	0.00297	0.00420	0.01728
Input length (# of users: 7)	20 Mins	0.00030	0.00057	0.00088	0.00139	0.00223
	30 Mins	0.00027	0.00085	0.00134	0.00200	0.01145
	1 Hour	0.00038	0.00106	0.00161	0.00293	0.01199
Output length (# of users: 7)	20 Mins	0.00024	0.00058	0.00110	0.00136	0.00460
	30 Mins	0.00034	0.00104	0.00168	0.00235	0.00758
	1 Hour	0.00038	0.00148	0.00220	0.00341	0.01455

From an application standpoint, it would beneficial to find the parameter setting that would provide (approximately) the longest possible length of prediction with an acceptable mean error rates. This would, of course, be dependent upon individual users, and require efficient hyper-parameter tuning methods (Bergstra and Bengio, 2012; Young et al., 2015).

5.3 Closed-loop predictions

So far, the RNN model predicted an output sequence when the true input sequence—i.e., the sequence of a user's actual activeness data—was provided. This section illustrates how the predicted results are affected when the previously predicted sequences are fed back into the model to conduct a closed-loop prediction.

Figure 4 illustrates the closed-loop predictions of heart rates, which are drawn in red, for the next (a) 24 hours and (b) 3 hours. Note that the blue lines represent the predicted values while the green lines show the actual activeness values of the user. During the closed-loop predictions, we observe that the network exhibits a typical limit cycling behavior of dynamical systems where its oscillating trajectory is converging toward a fixed point.



(a) Closed-loop predictions for 24 hours

(b) Closed-loop predictions for 3 hours

Figure 4: Closed-loop predictions of heart rates for the next (a) 24 hours and (b) 3 hours.

As the output length in this model is 15 minutes, we can see that the model approximates the real data by sequences of 15-minute data. So far, the longest length of an output vector considered in this experiment is 6 hours. It is interesting to note that the MSE rates do not increase linearly with the length of output vector, and more in-depth experiments should be conducted to closely examine the trade-off between the output length and MSE.

6. Conclusion

In this work, we explored the feasibility of modeling a user's activeness using biometric data retrieved from fitness trackers. We proposed four RNN architectures, and later selected one (Multi-MM) to further investigate the performance under various length parameter settings. We are currently developing a health-care application that aims to increase a user's activeness through proactively recommending (and learning) activities that the user likes to perform, and have prominent effects on his/her activeness.

References

8 Weeks to Wellness. Wellness score. URL `http://8ww.com/what-is-a-wellness-score/`.

Peter Austin, Jack Tu, Jennifer Ho, Daniel Levy, and Douglas Lee. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. *Journal of Clinical Epidemiology*, 66(4):398–407, 2013.

Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, 1994.

James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(1):281–305, 2012.

Loubna Bouarfa, Louis Atallah, Richard Kwasnicki, Claire Pettitt, Gary Frost, and Guang-Zhong Yang. Predicting free-living energy expenditure using a miniaturized ear-worn sensor: An evaluation against doubly labeled water. *IEEE Transactions on Biomedical Engineering*, 61:566–575, 2014.

Anthony G. Capon. Diseases of modern civilisation, 2012. URL `http://www.natsoc.org.au/our-projects/biosensitivefutures/part-4-facts-and-principles/human-health-issues/diseases-of-modern-civilisation#references`.

Centers for Disease Control and Prevention. Health, United States, 2015, with special feature on racial ethnic health disparities. Technical report, May 2016.

Teddy Cheng, Andrey Savkin, Branko Celler, Steven Su, and Lu Wang. Nonlinear modeling and control of human heart rate response during exercise with various work load intensities. *IEEE Transactions on Biomedical Engineering*, 55(11):2499–2508, 2008.

Dacadoo. Health score. URL `https://info.dacadoo.com/product/overview/`.

Florian Eyben, Felix Weninger, Stefano Squartini, and Bjorn Schuller. Real-life voice activity detection with lstm recurrent neural networks and an application to hollywood movies. In *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 483–487, Vancouver, Canada, 2013.

Rachel Feltman. Fitness trackers are no good at counting calories, and other lessons from wearing four at once, March 2014. URL `http://qz.com/184639/fitness-trackers-are-no-good-at-counting-calories-and-other-lessons-learned-from-wearing-four-at-once/`.

Kenyon Findley. Perceived effectiveness of fitness trackers among adults. *Current Issues in Health*, Paper 1, 2015.

Osamu Fukuda, Yoshihiko Nagata, Keiko Homma, and Toshio Tsuji. Evaluation of heart rate variability by using wavelet transform and a recurrent neural network. In *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1769–1772, Istanbul, Turkey, 2001.

Klaus Greff, Rupesh Srivastava, Jan Koutník, Bas Steunebrink, and Jürgen Schmidhuber. Lstm: A search space odyssey. *arXiv preprint*, 1503(04069):1–18, 2015.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

Lara Keytel, Julia Goedecke, Timothy Noakes, H Hiiloskorpi, Raija Laukkanen, Lize van der Merwe, and Estelle Lambert. Prediction of energy expenditure from heart rate monitoring during submaximal exercise. *Journal of Sports Sciences*, 23(3):289–297, 2005.

Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint*, 1412(6980):1–15, 2014.

Life Time Fitness. myhealthscore. URL `https://www.lifetimefitness.com/en/programs-and-events/myHealthScore-landing-page.html`.

Zachary Lipton, David Kale, Charles Elkan, and Randall Wetzel. Learning to diagnose with lstm recurrent neural networks. In *Proceedings of the 4th International Conference on Learning Representations*, San Juan, Puerto Rico, 2016.

Antonino Marvugliaa and Antonio Messineo. Using recurrent artificial neural networks to forecast household electricity consumption. In *Proceedings of the 2nd International Conference on Advances in Energy Engineering*, pages 45–55, Beijing, China, 2012.

Tomas Mikolov, Martin Karafiat, Lukas Burget, Jan Cernocky, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Proceedings of the 2010 Interspeech*, pages 1045–1048, Makuhari, Japan, 2010.

Gord Miller and Leslie Foster. Critical synthesis of wellness literature. Technical report, Department of Geography & Faculty of Human and Social Development, University of Victoria, February 2010.

Nasimul Noman, Leon Palafox, and Hitoshi Iba. *Natural Computing and Beyond: Winter School Hakodate 2011, Hakodate, Japan, March 2011 and 6th International Workshop on Natural Computing, Tokyo, Japan, March 2012, Proceedings*, chapter Reconstruction of Gene Regulatory Networks from Gene Expression Data Using Decoupled Recurrent Neural Network Model. Springer Japan, 2013.

Amit Pande, Yunze Zeng, Aveek Das, Prasant Mohapatra, Sheridan Miyamoto, Edmund Seto, Erik Henricson, and Jay Han. Energy expenditure estimation with smartphone body sensors. In *Proceedings of the 8th International Conference on Body Area Networks*, pages 8–14, Brussels, Belgium, 2013.

Rosaria Silipo and Carlo Marchesi. Artificial neural networks for automatic ECG analysis. *IEEE Transactions on Signal Processing*, 46(5):1417–1425, 1998.

Mayu Sumida, Teruhiro Mizumoto, and Keiichi Yasumoto. Estimating heart rate variation during walking with smartphone. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Zurich, Switzerland, 2013.

Elif Übeyli. Combining recurrent neural networks with eigenvector methods for classification. *Digital Signal Processing*, 19(2):320–329, 2009.

OlaOluwa Yaya, Olanrewaju Shittu, and Mohammed Tumala. Estimates of bull and bear parameters in smooth threshold parameter nonlinear market model: A comparative study between nigerian and foreign stock markets. *European Journal of Business and Management*, 5(7):107–123, 2013.

Steven Young, Derek Rose, Thomas Karnowski, Seung-Hwan Lim, and Robert Patton. Optimizing deep learning hyper-parameters through an evolutionary algorithm. In *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments*, pages 1–5, Austin, Texas, United States, 2015.

Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and Leon Zhao. *Web-Age Information Management: 15th International Conference, WAIM 2014, Macau, China, June 16-18, 2014. Proceedings*, chapter Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks, pages 298–310. Springer International Publishing, 2014.