Universiteit Leiden

Computer Science

RNNs Of Heart Rate Modeling As A Function Of Body Movement

Name:	Yazhou Liu
Date:	August 25, 2017
1st supervisor:	Wessel Kraaij
2nd reader:	Arno Knobbe

MASTER'S THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

.

Acknowledgements

I highly thank all my academic supervisors for all their time, feedback, effort and guidance throughout this work.
And of course my parents for their constant support.

RNNs Of Heart Rate Modeling As A Function Of Body Movement

Yazhou Liu
Leiden University
y.liu.21@umail.leidenuniv.nl

Wessel Kraaij
Leiden University
w.kraaij@liacs.leidenuniv.nl

August 25, 2017

Abstract

Recently, there has been an explosion of interest in time series data mining by using machine learning methods and deep learning methods. In the medical field, enormous volumes of data as diverse as human activity, body temperature, body movement and heart rate are recorded, which highlight the importance of data mining methods. Heart rate modeling has always been an important research topic, in the context of medical care and preventive health. According to our knowledge, heart rate modeling has not been studied using deep learning methods. This paper focuses on heart rate modeling as a function of human body movement by using different regression models, such as ridge regression, K nearest neighbors regression and deep learning methods using RNN regression. Through merging and preprocessing the raw data from the LUMC SwitchBox study, we created a training set to construct regression models. We designed and constructed a RNN regression model, which performs better than the state-of-art regression models. We also looked at the influence of the number of hidden layers and the number of hidden units in each layer for RNN regression models. Finally, we investigated the influence of training data size on the quality of the heart rate models.

Contents

1 Introduction

As a growing number of our activities are carried out on portable devices, such as laptops and smartphones, an increasing amount of what we do is recorded. While the Internet is connecting more and more portable devices, it becomes convenient to collect these records and to record them into a dataset which is suitable for machine learning and deep learning methods[10]. Meanwhile, the growing use of time series data has stimulated research in novel data mining algorithms and applications [7].

This study makes use of two datasets collected in the SwitchBox and Gotov study. The SwitchBox study is a European project, which focuses on health maintenance by better homeostasis in old age [11]. The SwitchBox study collected various physiological time series data - using various sensors - in a study focusing on understanding the factors that determine longevity. The data were collected from the offspring of long living persons and the partners of the offspring who were regarded as a normal control group. The Gotov study is a research project on body movement and sports from LUMC (Leiden University Medical Center).

The heart is one of the most important organs in the human body and the heart rate could supply many useful medical diagnostics, which means it is worthwhile to make an effort for heart rate modeling. The heart rate is a complicated function of physical activity, emotion and many other factors. This paper applied some machine learning methods and deep learning methods on the time series accelerator data to predict the heart rate according to the human body movement as an exploratory analysis, for research purposes to explain part of the heart rate variability.

The developments of human brain research inspired the studies of neural network and deep learning methods in the past several decades. In recent years, the neural network and deep learning methods have been applied in various fields, because of stronger CPUs, larger labeled data sets and the power of the deep learning methods itself [10]. Recurrent Neural Networks are one of the most popular deep learning architectures, which have an excellent performance on series tasks such as handwriting detection and time series data prediction. In this paper, we applied Recurrent Neural Networks on the SwitchBox [11] study data to model the heart rate as a function of accelerometer data. According to to our knowledge heart rate modeling has not been studied by deep learning methods. Physical exercise is one of the factors influencing the heart rate. Ultimately we intended to find the heart rate 'anomalies' patterns that cannot be explained by physical exercise heart rate response alone.

In this paper, we merged and preprocessed the raw data from the SwitchBox study to reproduce the proper data set for building regression models. We built a proper RNN regression model, which performed better than the state-of-art regression models, such as ridge regression, K nearest neighbors and decision tree regression. We also had a look at the influence of the number of hidden layers and the number of hidden units in each layer for RNN regression models. We also investigated the influence of training data size on the quality of the heart rate models.

The research objective of this thesis work was to do an exploratory data analysis of the heart rate and accelerator Time Series data collected during the SwitchBox study [11], in form of a regression problem analysis, to hopefully find a method to improve the accuracy of heart rate modeling, which could explain the heart rate function.

In this paper, firstly we introduce the related work in Chapter 2, which includes time series prediction and heart rate modeling. Second, we discuss the methods of Kalman Filter, the heart rate and Recurrent Neural Networks in Chapter 3. Third, we introduce the detailed description of the data and the data merging and preprocessing in Chapter 4. Fourth, our approaches are applied to the SwitchBox study in Chapter 5. Finally, a discussion and conclusion are provided in Chapter 6 and Chapter 7 respectively.

2 Related Work

The related work of this paper includes time series data regression and heart rate modeling. In this section, we give a brief overview of various strategies and methodologies, to provide a context for our research/

2.1 Time Series Data Prediction

Usually, time series data are a collection of observations at multiple time periods. The characteristics of time series data are large, numerical and continuous. The growth of time series usage especially in medical and financial domains attracts plenty of research and development on time series data mining [7]. There are three common aspects of time series data mining: representation techniques, distance measures, and indexing methods [5]. Another significant aspect of time series data mining is prediction, which is a key point of this paper.
Time series data prediction always is a major topic of debate and there is a large volume of research and development on this hot topic. We are not in a position to introduce all of them, so we will introduce three excellent, commonly used regression models and regard them as baselines for the Recurrent Neural Networks regression models which this paper focuses on.
One of the baselines is ridge regression models, which are also called regularized linear regression. Regularized linear prediction is based on the ordinary least squares (OLS) approach and optimized by adding a regularized parameter to avoid over-fitting problem, which have an excellent performance on real-world data such as R-R intervals of ECG signals [12]. Theoretically, ridge regression could to be faster to predict and steady while the accuracy is similar to other regression models [23]. A thesis found that local ridge regression consistently perform better than the K nearest neighbors regression models (KNN) and the kernel smoothing method in term of traffic forecasting [20]. In a recent research, ridge regression is applied to genome wide selection in maize [16].
The second baseline that we picked is K nearest neighbors regression method (KNN). The basic idea of K nearest neighbors regression method is to calculate the property value for the object by finding K closest training examples in feature space [1]. Olga $et\ al.$ investigated that the weighted K nearest neighbors regression model (KNNimpute) performs more robust and sensitive for missing value estimation than a singular value decomposition (SVD) based method (SVDimpute) [21]. Recently, one successful study demonstrates that "the proposed technique, such as K nearest neighbors, method can excavate hidden patterns/relationships in EEGs and give greater understanding of brain functions from a system perspective, which will advance current diagnosis and treatment of epilepsy" [2]. In multi-label learning problems, Min-Ling $et\ al.$ designed experiments show that ML-KNN method achieves a better performance than some well-established multi-label learning algorithms [25].
The last baseline which we propose using is decision tree regression models. "Decision tree regression models are a non-parametric supervised learning method whose goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features." [18] [17] [9] In Robert $et\ al.$'s research, GAFDT (genetic algorithms of fuzzy decision tree) model performs better than other regression methods for financial time series data prediction on various stocks in Taiwan Stock Exchange Corporation (TSEC) [13]. In other domains, M.A. $et\ al.$ found that decision tree methods have several advantages for classification of land cover from remotely sensed data [6]. Similarly, R.S.$et\ al.$ also applied decision tree methods to global land cover classifications at 8km spatial resolutions [4].

Basically, ridge regression models, K nearest neighbors models and decision tree models are robust regression models for time series data prediction, which could represent the state-of-art time series data prediction.

2.2 Heart Rate modeling

Heart rate modeling is another major topic of discussion. In basic research, most of heart rate models are based on the exercise level. Cheng $et\ al.$ introduced the heart rate prediction model based on a nonlinear system response during and after treadmill walking exercise [3]. Another heart rate model which responds dynamics to moderate exercise was investigated by Steven $et\ al.$ [19]. However, "Current heart rate models were developed for a specific scenario and evaluated on unique data sets only". [8] Matthias $et\ al.$ investigated a heart rate model based on analytical models and machine learning approaches, which are trained by the data of indoor environments such as treadmills and bicycle ergo-meters, as well as the data of outdoor environments recorded by smart phone [8]. In order to improve the models, the heart rate prediction model is not only based on exercise data, but also from other individual information and physiological data. Mette $et\ al.$ introduced a model to predict the heart rate during postural change, whose parameters include blood pressure, barcaroles firing-rate and the combined effects of vestibular and central command stimulation of muscle sympathetic nerve activity [15]. Another recent heart rate model can predict the heart rate analysis not only by exercise physiology, but also by the areas of cardiovascular health and rehabilitation [24].

Most of the heart rate prediction models are based on strenuous exercise such as running and cycling, which is easy to observe the heart rate changes. In this paper, one obvious challenge is that we need to predict the heart rate from the data of body movement in daily activities, which is difficult to observe the heart rate changes and based on our knowledge heart rate modeling has not been studied by deep learning methods.

3 Methods

In this section, we will introduce some approaches we used in this project. Kalman Filter is a famous optimal estimator, which is recursive so that new measurements can be processed as they arrive. The heart rate could be a result of a complex function whose input parameters could be individual information, activity degree, emotion, temperature and so on. RNN is a hot topic of discussion in deep learning domain, which performs excellent for unsegmented handwriting recognition, speech recognition and time series regression.

3.1 Kalman Filter

Kalman Filter is a linear quadratic estimation algorithm, proposed by Rudolf Kalman in 1960. After that, Apollo program in the navigation system applied Kalman Filter algorithm to its main application. Kalman Filter uses the measurement of the system in the time domain to predict the state of some unknown variables at the next point in time, including noise interference and other uncertainties. Kalman Filer has higher accuracy than other prediction algorithms. It can be used for any dynamic system containing unknown information, and then adaptively predict the next state of the system. Even if the environment of the system is very complicated, Kalman Filter generally can calculate accurate results. Kalman Filter has plenty of advantages in the continuous change of the system. Its advantage is that it is not necessary to record many states of the systems in the prediction, and only needs to know the error of the previous state and covariance matrix of the error. Kalman FIlter is very fast, which can be used for real-time dynamic system.

3.2 Heart Rate

Heart of the human body is always a hot topic about medical treatment and health maintenance. The role of the heart is to provide adequate blood flow for the cells, which could supply oxygen and various nutrients, and take the metabolism products away, so that cells maintain normal metabolism and function.

The heart always delivers the blood by beating. The heart rate is the number of heart beat per minute. We cannot control the heart rate by ourselves. There are some basic factors influencing the heart rate:

- Autonomic nervous system regulation. The increasing of sympathetic activities will cause the heart rate growth, while parasympathetic activities's increasing will lead the heart rate slow down.
- Body fluid regulation. Adrenaline, nor-epinephrine and thyroid hormones can increase the heart rate.
- Body temperature. The heart rate will increase 12 to 18 times per minutes, when the human body temperature increase $1°C$.
- The woman heart rate is faster than man and the older heart rate is slower than kids.
- The emotional state can also influence the heart rate through the limbic system.

These factors are the basic points to influence the heart rate and other aspects such as exercise affect the heart rate through these basic factors. For example, the heart rate would be higher when people exercise, because exercise could cause the activities of the autonomic nervous system changing.
It means there could be a complex function to predict the heart rate, whose input parameters could be individual information, activity degree, emotion, temperature and eating. This paper is an early research to predict the heart rate by activity degree, which could help us to have a look at the heart rate function. For example, we could subtract the predicted the heart rate from the actual the heart rate to locate episodes where the heart rate is influenced by other input parameters (e.g. stresses).
The methods of the heart rate measuring and monitoring is a significant point of the heart rate prediction. In the early research, information lost during analog filter, which is a part of the heart rate measuring, is a serious problem [22]. A faster the digital filter, which is implemented by 4-bit micro-controller, is introduced by Thomas $et\ al.$ [14]. "In the present embodiment, a mass-produced general-purpose single-chip 8-bit microprocessor is adequate to accomplish all the required digitization and calculations". [22] In recent research, "the electrocardiogram (ECG) and Holter monitoring devices are accurate, but they are not appropriate for use in field settings due to cost, size and complexity of operation. Lightweight telemetric heart rate monitors equipped with conventional electrodes have been available since 1983 and have been shown to be accurate and valid tools for the heart rate monitoring and registering in the field"[14]. Basically, Equivital devices which SwitchBox study uses process electrocardiogram (ECG) information to the heart rate calculation.

3.3 Recurrent Neural Network (RNN)

Neural network (neural network, abbreviation NN), is a calculation model to simulate the structure of a biological neural network. The neural network is obtained by a large number of layers and units. Mostly, artificial neural network is an adaptive system, which could update its structure by learning data. Recurrent Neural Network is part of the most popular neural network models. Unlike traditional FNNs (Feed-forward Neural Networks), RNNs introduce directional loops that can deal with the problems related to those inputs before and after.
The aim of Recurrent Neural Networks is to deal with time series data and recognition of the language . For the normal neural networks, the layers are connected but the nodes in same layer are not connected in the network. Nonetheless, this kind of neural network is not powerful for many problems related to time and sequence. For example, when you try to predict the next word of a sentence, the performance of the network would be not satisfied if you just use the current word, because the words in a sentence are related, which means we need to use the previous words to predict what the next word is. RNNs will store the preceding data point and apply it to the calculation of the current output, which means the nodes in same layer are connected in RNNs compared to the normal neural networks. Figure 1 taken from [10] displays the typical structure of RNNs. We could see that the output o_t of current time t is calculated by the current input x_t and the previous state s_{t-1}. U represents the connection between input layers and hidden layers; V represents the connection between hidden layers and output layers; W represents the connection between different hidden layers.
In this paper, we will apply RNNs to heart rate modeling and have a look at the meaning of components in its structure.

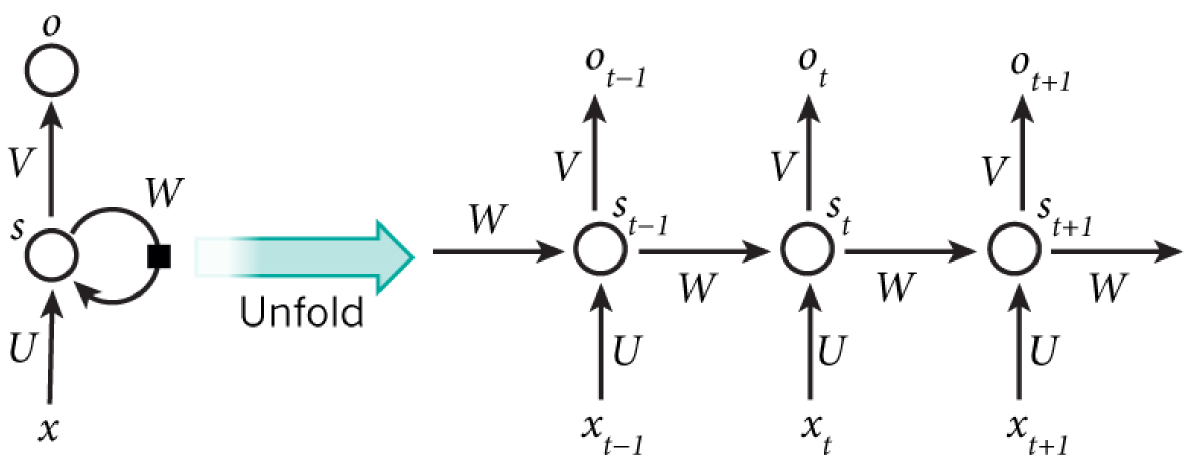


Figure 1: The typical structure of Recurrent Neural Networks

4 The SwitchBox Data Description and Preprocessing

4.1 Data Description

In this paper, we used the heart rate and accelerator data which are a part of the SwitchBox Study and Gotov Research. The data were collected from a total of 132 participants. The size of raw data about heart rate and accelerator was about 4TB, which was stored in csv format. Additionally, the participants included had to meet certain criteria, which could be checked in Figure 2 token from [11].

Exclusion criteria Switchbox participants	
Laboratory results	Fasting Plasma glucose > 7 mmol/L Hemoglobin < 7.1 mmol/L TSH < 0.3 mU/L or > 4.8 mU/L fT4 < 10 pmol/L or > 24 pmol/L
Disease history	Any significant chronic disease; renal, hepatic or endocrine disease
Medication use	Hormone therapies Use of medication known to influence lipolysis
Lifestyle	Recent weight changes (> 3 kg weight gain/loss within last 3 months) Extreme diet therapy Alcohol consumption of more than 28 units/week Smoking addiction
Others	Severe claustrophobia Difficulties to insert IV cannula Blood donation (< 2 months) Participation in other research project (< 3 months or >2 within 1 year)

Figure 2: Exclusion criteria for the SwitchBox participants

The experiments in the SwitchBox study took 5 days. The electrocardiography, core body temperature, breathing rate and physical activity were recorded by Equivital devices and the core body temperature was collected by a special capsule which was swallowed by each participant. Meanwhile, wearing the GENEActive devices was another requirement for participants, which could record the movements of their wrist and ankle. The records of participants' glucose were collected by another monitor [11].

4.2 Data Preprocessing

The purpose of data preprocessing was to transform raw data into an understandable format. The data without preprocessing often contained many errors and could not be applied to data mining methods direct. Data preprocessing is a proven method of resolving such issues, which prepares raw data for further processing.

4.2.1 Noise

As we know, data noise is always a key point of data processing and mining, which is the same for the SwitchBox Study.
The human heart rate can physiologically only reach the values between 30 beats per minute (bpm) and 220 beats per minute bpm, however as you can see in Figure 5 (participant 11), which is the heart rate data of one participant for 4 hours in the SwitchBox Study, a lot of noise entered the data by having a lot data points less than 30 bpm and more than 220 bpm. In order to predict the heart rate correctly, we need to remove this noise. In this paper, we used Kalman Filter, which we introduced in Chapter 3, to do this job.
Sometimes, a Kalman Filter will remove some useful information about the original data. Actually, there is a trade-off between degree of smoothing a way and conserving information. In the SwitchBox Study, participants collected their data by themselves at home after the first day. Due to the reason that the Equivital devices need to be charged 2 times a day and the devices could be worn improper, the qualities of the data for different participants are different and uncontrollable.
There is a measurement in the SwitchBox Study, which could measure the data's percentage of artifact. Taken into account this measurement, we chose some participants' data which had high quality. Figure 6 shows the data of participant 24 which has high quality, we could see that it almost does not have noise compared to figure 5 (participant 11), which is helpful for us to decide the smooth degree of Kalman Filter. After discussion with LUMC researchers, we agreed that these high quality data will help us to decide the proper degree of smoothing and we also decided it together according to the biological validity. And then we applied the Kalman Filter which is tuned for the high quality data to the data from all participants. Figure 3 shows the heart rate raw data and preprocessed data by Kalman Filter. We could see that Kalman Filter removes some obvious noise data points and smooths the raw heart rate data in an excellent way in terms of a physiological nature. For example, the data points marked with red circles are the obvious noise, because the human heart rate could increase from 70 to 150 within few seconds but it cannot slow down within the same time.

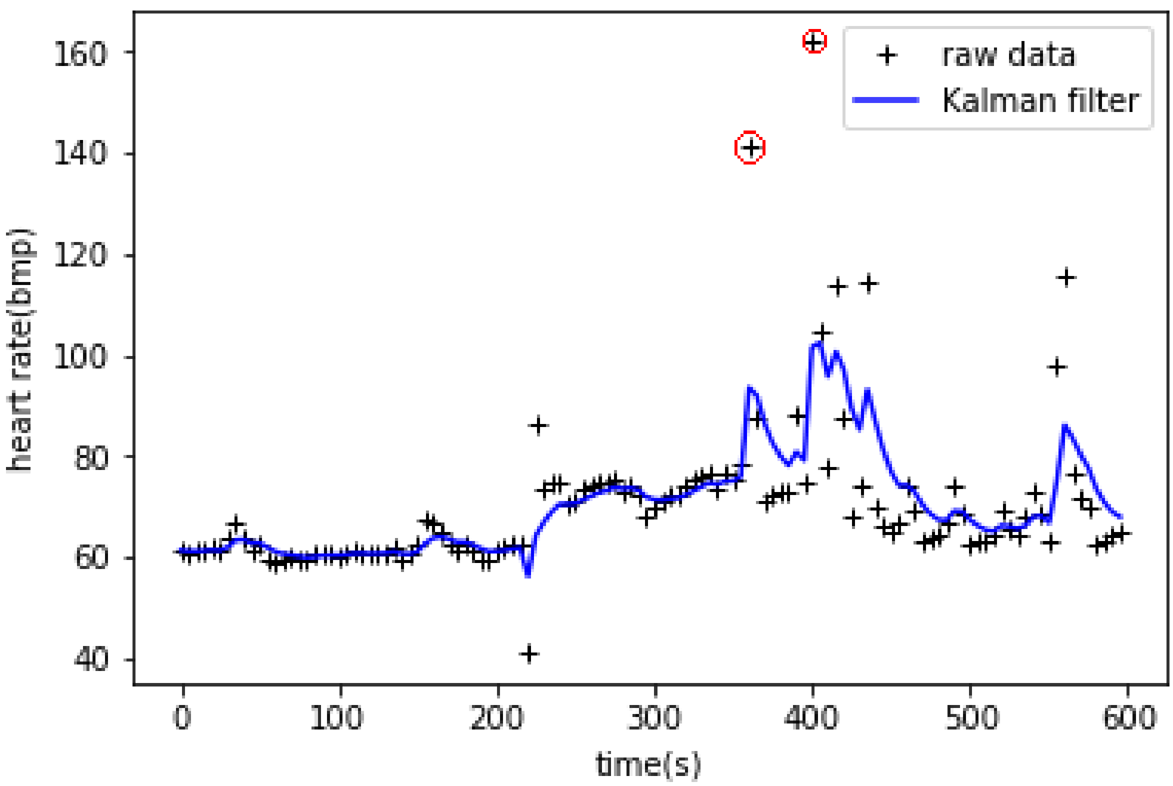


Figure 3: Kalman filter for the heart rate data of one participant for 10 minutes in the SwitchBox Study

The artifact of data also could influence the regression models' performance. The RMSE (root mean square error) of the RNN regression model for participant 11 (figure 5) is 17.63, while the RMSE (root mean square error) of the RNN regression model for participant 24 (figure 6) is 13.47, which means the quality of data is a significant factor to build heart rate prediction model. And also we did not apply Kalman Filter to this experiment, because the parameters of Kalman Filter we used is tuned for high quality data rather than the worse one, which could influence the comparison.
We also implemented the Kalman Filter on other features (breath rate and accelerator) and Kalman filter also performed excellent for these features. Figure 4 displays the effect of Kalman Filter on normalized accelerator data which has zero mean and unit variance. We could see that Kalman filter remove most of out layers and make data continuous which is more reasonable in terms of their physical nature.

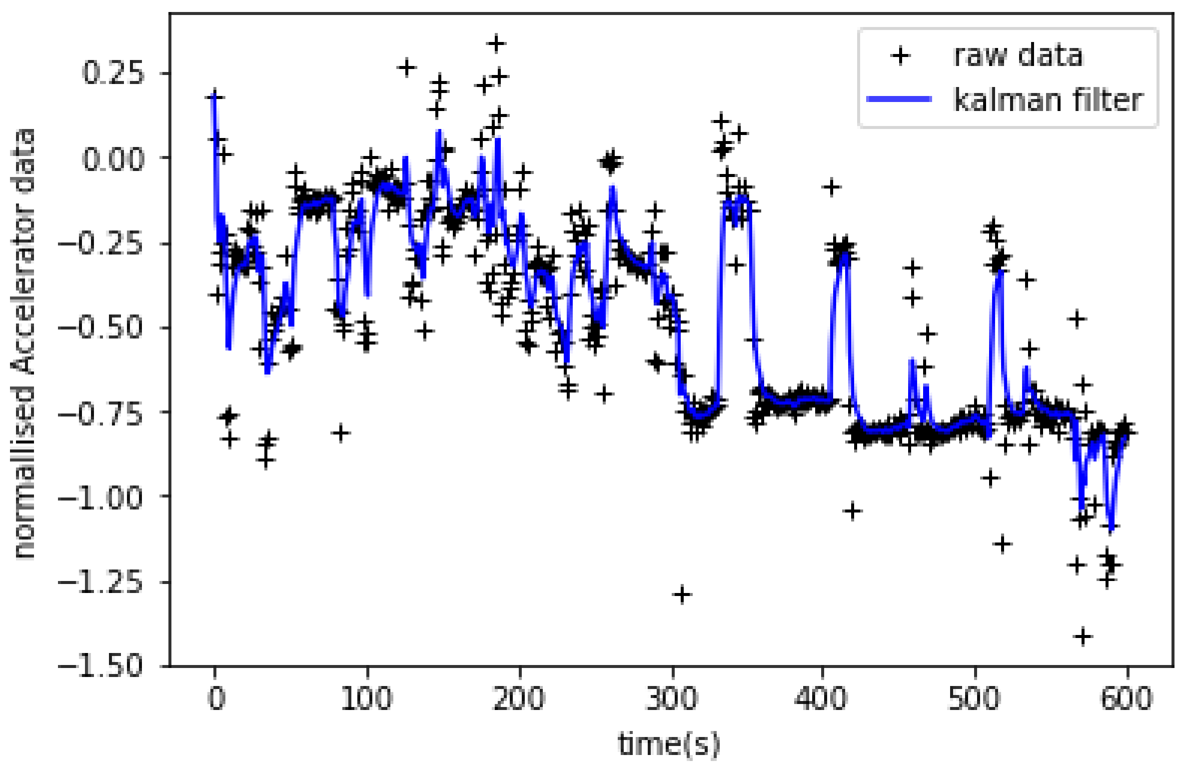


Figure 4: Kalman filter for accelerator data of one participant for 10 minutes in the SwitchBox Study

In order to validate that a Kalman Filter is useful for building heart rate prediction model, we applied some general regression models to one participant's 1 Hz data to test the raw data and the preprocessed data. The result is on the Table 1 and we could see that the RMSE(Root Mean Square Error) of data which is preprocessed by Kalman Filter is lower than raw data' RMSE. This means Kalman Filter is helpful to decrease the error of heart rate prediction models.

Methods	Ridge Regression	KNN(K Nearest Neighbors)	Decision Tree
Raw Data(RMSE)	21.03	50.00	57.16
Kalman Filter(RMSE)	19.70	46.61	47.93

Table 1: RMSE(Root Mean Square Error) of some general regression models for 1Hz raw data and preprocessed data(Kalman filter)

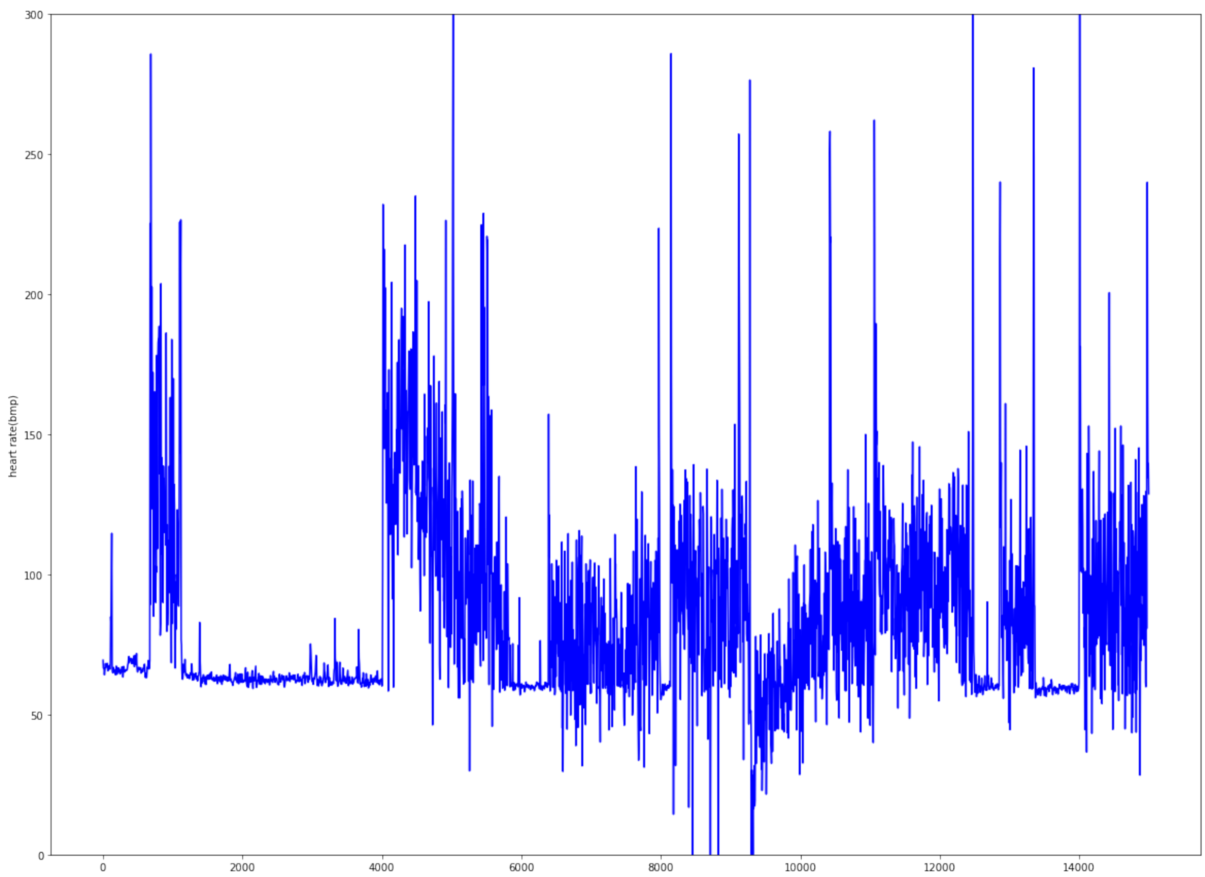


Figure 5: heart rate data of one participant for 4 hours in the SwitchBox Study

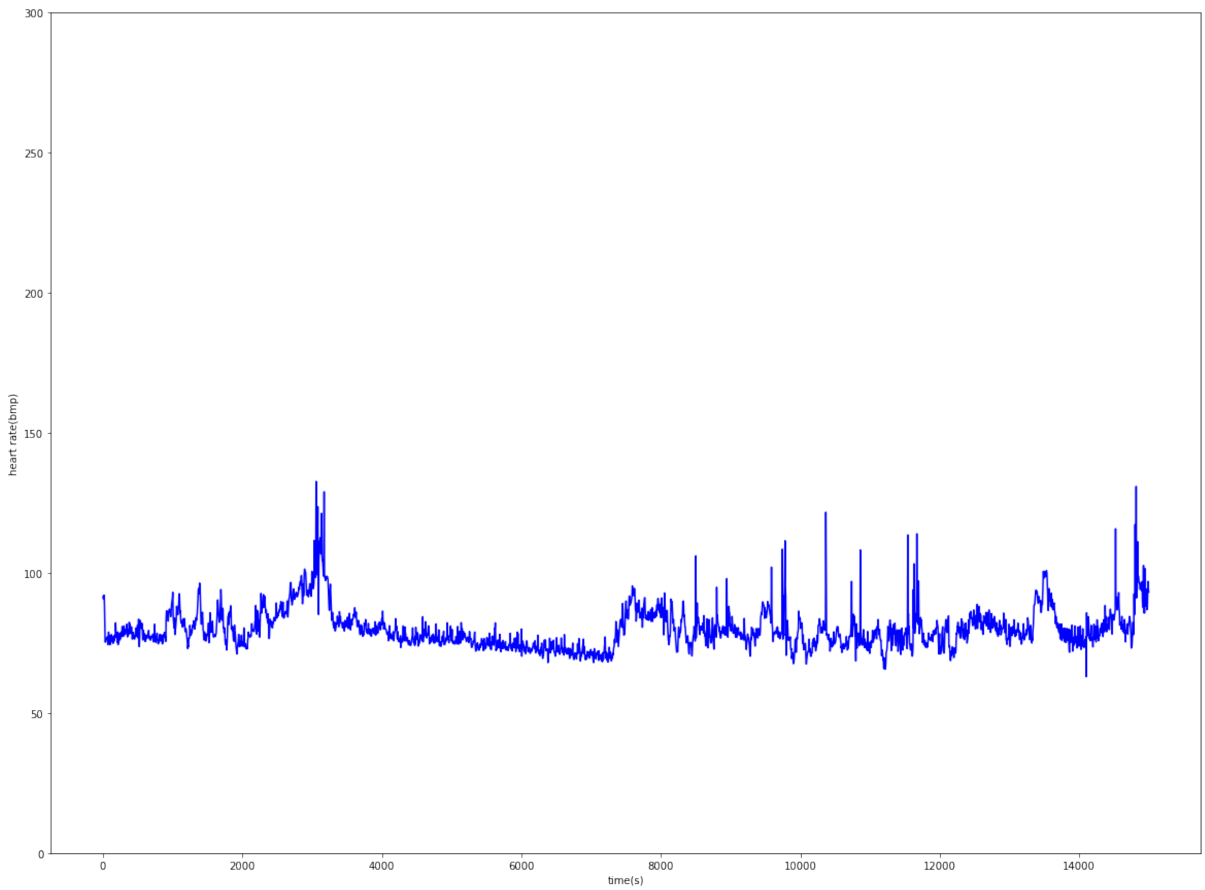


Figure 6: high quality heart rate data of one participant for 4 hours in the SwitchBox Study

4.2.2 Data Frequency

In the SwitchBox Study, the different features of the data were collected in different frequency. In this paper, we used the heart rate data and breath rate data(0.2 Hz), chest accelerator data(24 Hz), Ankle and Wrist accelerator data(83 Hz). Due to the reason that most general regression models require the data to have the same frequency, we need to merge these data in same frequency. Usually higher frequency data contains more information, which is more helpful for most regression models. However, because the SwitchBox Study data is large for building regression especially for RNNs, we need to make a trade-off between data frequency and time/space consumption.

We picked up the first data point of high frequency data in a specific period as the represent active sample when we reduced the frequency of the data. For example, if we want to reduce 25 Hz data to 1 Hz data, which means the the 25 Hz data has 25 data points and 1 Hz data just has 1 data point in 1 second period, we picked up the first data point among the 25 data points as the data point of 1 Hz data. We applied linear interpolation methods when we increased the frequency of the data. Linear interpolation is an approach to curve fitting, which applies 1D polynomials to create new data points between exited data points, just like figure 7.

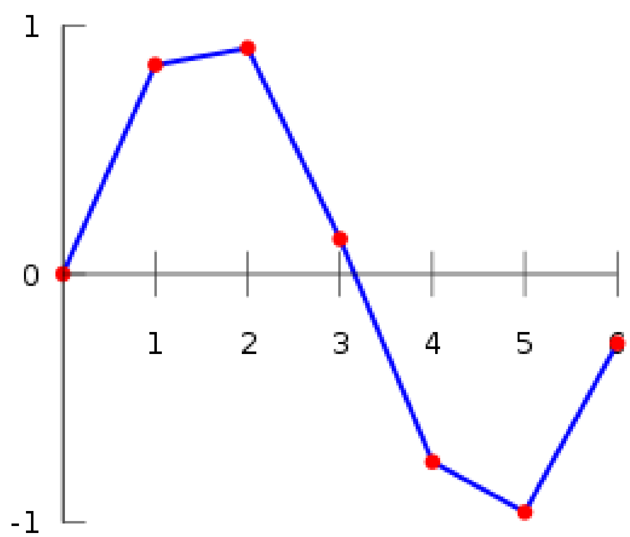


Figure 7: A example of linear interpolation

In order to decide which frequency we should use, we made experiments to validate it. We merged the data of participant 11 in different frequencies and built the same regression models to test it. Table 2 presents the results of the experiment. We could see that the RMSE of 1Hz data and 25Hz data are similar for general regression models. As a result, we decided to use 1Hz data to do the experiments in this paper, which could reduce much time to preprocess data and build models. The reason why we do not consider 0.2Hz data is that the frequency is too low to record the details of body movements, , which is insufficient for further research such as energy expenditure.

In the SwitchBox study, the devices would still record the data during charging time which leads to useless data, which should be removed. So the participants of SwitchBox study were called upon to record the time of devices putting on or taking off the devices. According to

Methods	Ridge Regression	KNN(K Nearest Neighbors)	Decision Tree
0.2Hz (RMSE)	22.01	34.45	63.67
1Hz (RMSE)	21.03	46.62	57.16
25Hz (RMSE)	25.82	41.48	54.61

Table 2: RMSE(Root Mean Square Error) of some general regression models for 0.2Hz, 1Hz and 25Hz raw data

this record, we cut the data properly.

5 Experiments

In this chapter, we introduced the main experiments in this project, which included the performance of regression models for all participants, the influence of the number of hidden layer and the number of hidden units in each layer, different validation methods, to the influence of data size for regression models.
In these experiments, we regarded first four days' data as training data and the last day's data as testing data for the data of every participant, which were preprocessed by Kalman filter and frequency alignment. And then we merged them together. All experiments were run on duranium service , a distributed system in LIACS (The Leiden Institute of Advanced Computer Science) data science group, which has 20 CPUs (Intel Xeon E5-265ov3 @2.30GHz), 8 GPUs (6 NVIDIA GTX 980 Ti each with 6GB memory, 2 NVIDIA Titanium each with 12GB memory) and 3TB local memory. All experiments were programmed by Python 3.6 and the libraries this paper used are Pandas, Numpy, Scikit-Learn and TensorFlow.

5.1 RNN regression model

One of this paper's aims was to compare RNN regression models to state of art regression models such as ridge regression model, K nearest neighbors and decision tree regression. In this experiment, we picked up first four days of data to train the models and apply the last day's data to test the models due to the reason that the data is time series data.
In the neural network, the hidden layers transformed the inputs to useful features or information for outputs, which is the same with RNNs. In this paper, we designed an experiment to look at the influence of the number of hidden layers for RNNs. Figure 8 displays the relation between the number of hidden layers and the errors of RNN models. The orange line is RMSE of testing data and the blue line is RMSE of training data. We found that both training data error and testing error decreased firstly and then increased with the number of hidden layers increasing in RNN. The RNN model which has three hidden layers performed best.

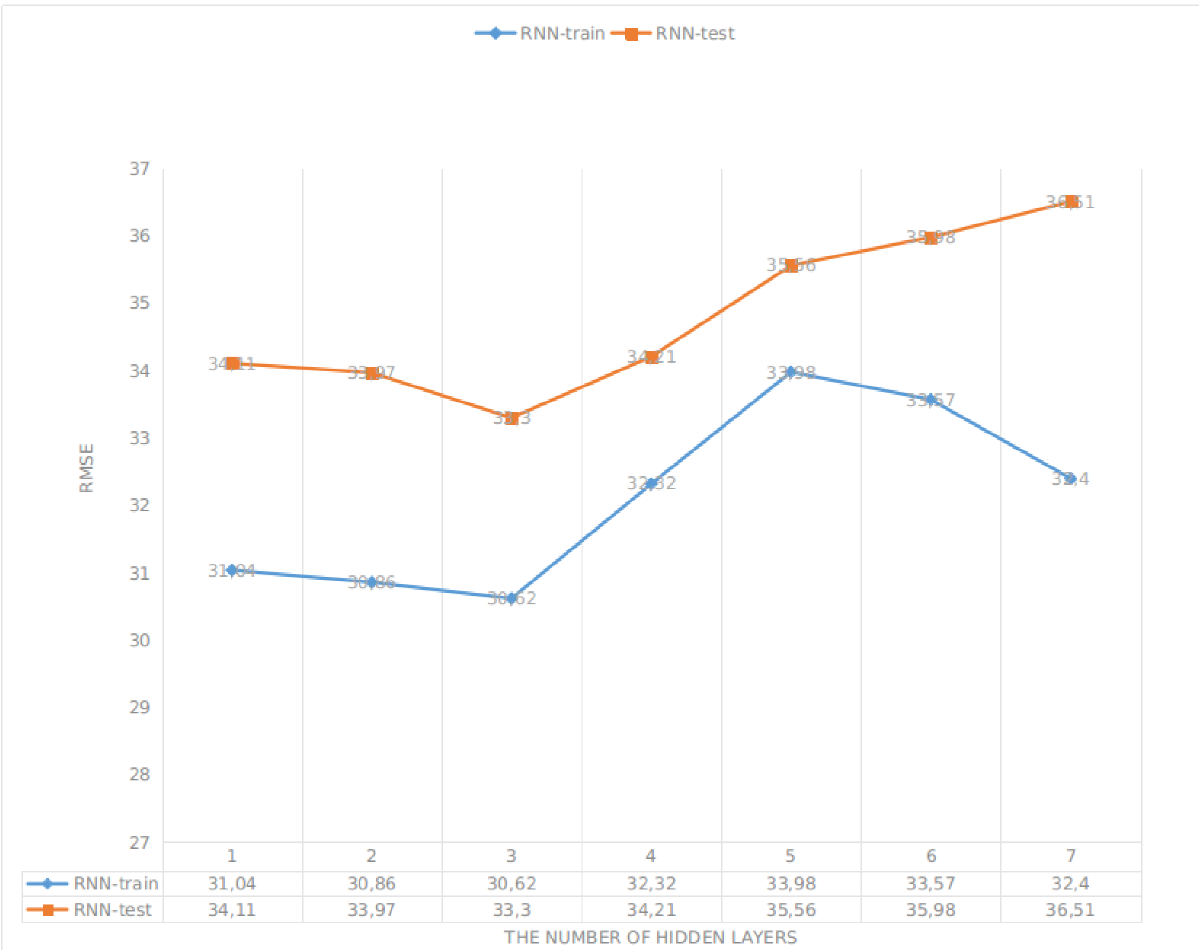


Figure 8: RMSE(root mean square error) of RNNs for different number of hidden layers

Figure 9,10,11,12 display the testing error and training error for epochs of training in different number of hidden layers. The RNN model of one hidden layer could not narrow the error to a steady level which means the error maintains in a specific value, compared to other RNN models which have more hidden layers. With the number of hidden layers increasing, the errors of the models reduce to a steady level faster.

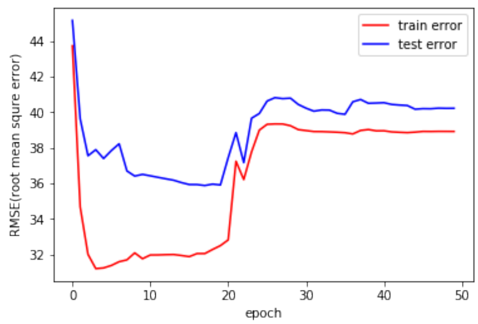


Figure 9: RMSE of RNN for one hidden layer

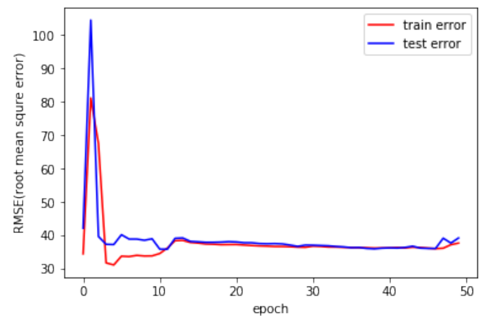


Figure 10: RMSE of RNN for three hidden layers

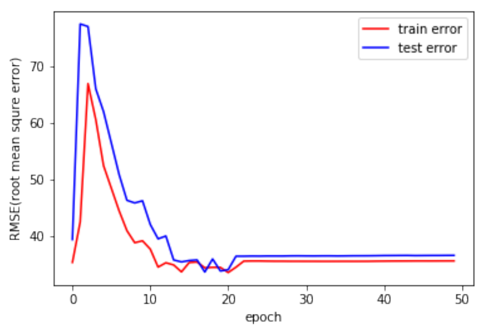


Figure 11: RMSE of RNN for five hidden layers

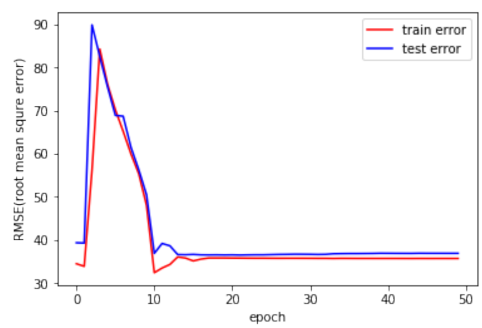
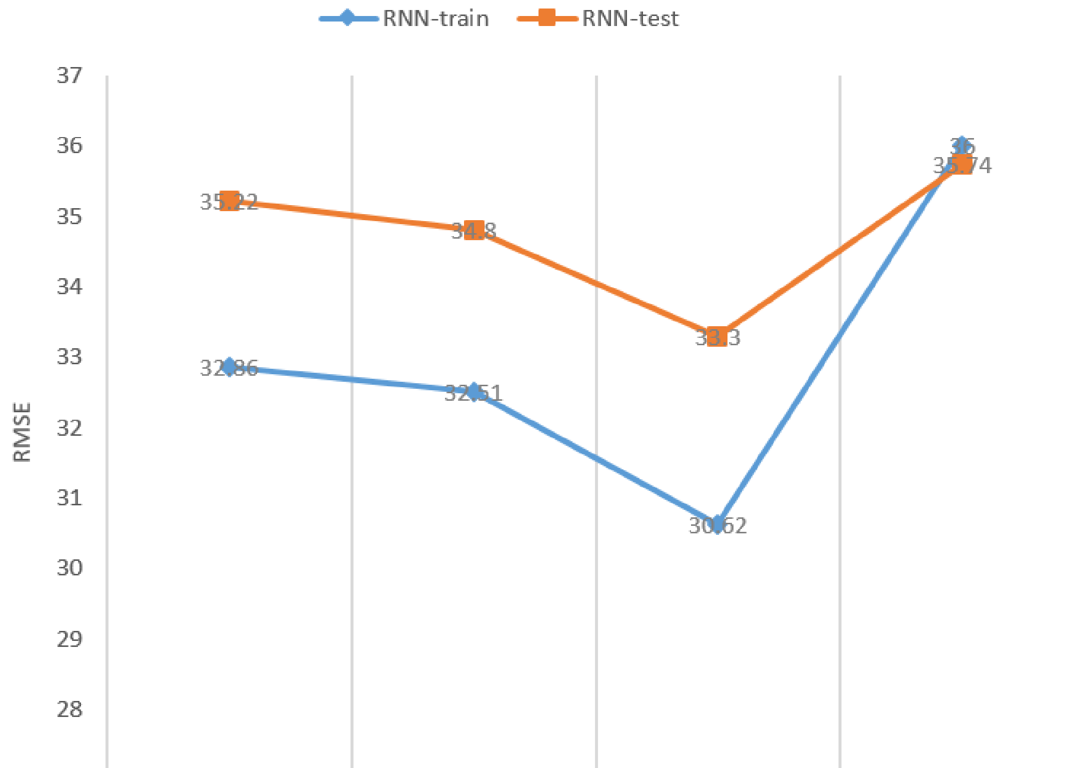


Figure 12: RMSE of RNN for seven hidden layers

Another important parameter for RNNs is the number of hidden units for each hidden layers. We also applied different numbers of hidden units to RNNs in this paper. We just consider the same number of hidden units in each hidden layer. Figure 13 shows the relation between the number of hidden units and errors of RNN and LSTM models. The orange line is RMSE of testing data and the blue line is RMSE of training data. We found that both training data error and testing error decreased firstly and then increased with the number of hidden units increasing in RNNs. The RNN model which has 64 hidden units performed best.



	16	32	64	128
RNN-train	32.86	32.51	30.62	36
RNN-test	35.22	34.8	33.3	35.74

Figure 13: RMSE(root mean square error) of RNN for different number of hidden units

Figure 14,15,16,17 display the testing error and training error for epochs of training in different number of hidden units in each hidden layer. The RNN model of 16 and 32 hidden units could not reduce the error to a steady level, compared to other RNN models which have more hidden layers. With the number of hidden layers increasing, the errors of the models reduce to a steady level faster, whose situation is similar to the experiments of the number of hidden layers.

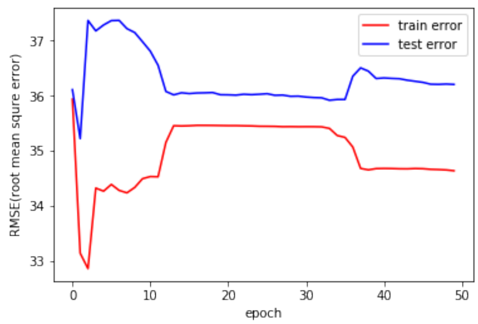


Figure 14: RMSE of RNN for 16 hidden unit

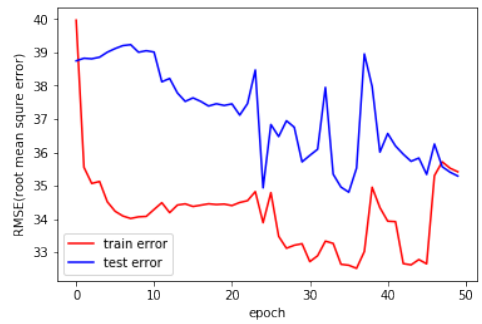


Figure 15: RMSE of RNN for 32 hidden units

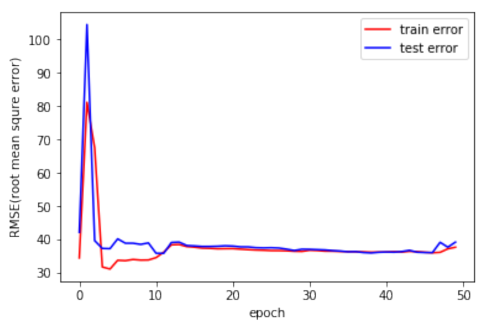


Figure 16: RMSE of RNN for 64 hidden units

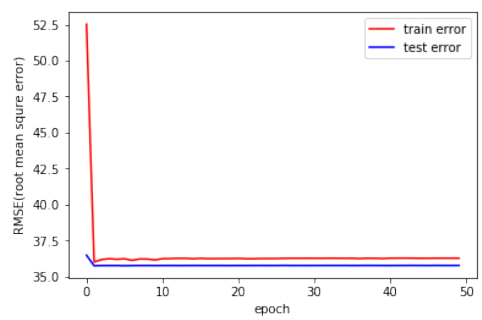


Figure 17: RMSE of RNN for 128 hidden units

The more complicated RNNs's structure (more hidden units and more hidden layers) is, the more time RNNs training take, because the more parameters need to be updated for each epoch of training.

Dropout is a popular regularization method in deep learning domain to avoid over-fitting. At each epoch of training, every hidden unit could be dropped out of the network with probability $1-p$. The dropped out hidden units do not attend this epoch of training, while the left hidden units will update weights in this epoch of training. After this epoch of training, the dropped out hidden units return to the network with their original weights. In this paper, dropout methods also improved the RNN regression models. The training set's and testing set's error of the RNN regression model which has 3 hidden layers and 64 hidden units for each hidden layers are 30.62 and 33.3 respectively with dropout method, compared to error of training set and testing set are 37.01 and 35.49 separately without dropout method.
According to the experiments we mentioned before, we picked a proper RNN regression model for heart rate regression. The LSTM regression model was also picked with same process.
The table 3 illustrates the comparison of state-of-art regression models and obviously the RNN and LSTM get lower RMSE(root mean square error) than other regression models. One important reason is that each result of RNN and LSTM is not only influenced by the current data point but also affected by the previous data point, compared to the state-of-art regression models which are only getting the results from current data point.

Methods	Ridge Regression	KNN	Decision Tree	RNN	LSTM
training error (RMSE)	33.49	45.48	51.61	30.62	29.82
testing error (RMSE)	35.72	49.00	52.16	33.3	32.73

Table 3: RMSE(Root Mean Square Error) of some general regression models for 1Hz and 25Hz raw data

Due to the fact that this paper is an initial project of heart rate prediction in SwitchBox Study. We did not set 'real' testing set to evaluate the regression models in this paper. The testing set and its error (RMSE) we mentioned before are both validation set (or tuning set) actually. However, we also validated that the regression models we used throughout this paper are meaningful and effective by five folds cross validation methods more or less. We regarded data of each day (total five days) as testing set respectively and the left data as training set to perform the cross validation methods. Figure 18 illustrates the results of five cross validation methods. The RNN regression model this experiment used is based on the experiments we mentioned before. In other words, The RNN regression model is based on the situation that the data of first four days for every participant is regarded as training set and the data of last day for every participant is regarded as testing set, which corresponds to the last fold in the five cross validation method. In figure 18, the other four folds' results are similar to the last fold's results, which means the model has an excellent performance for unknown data.

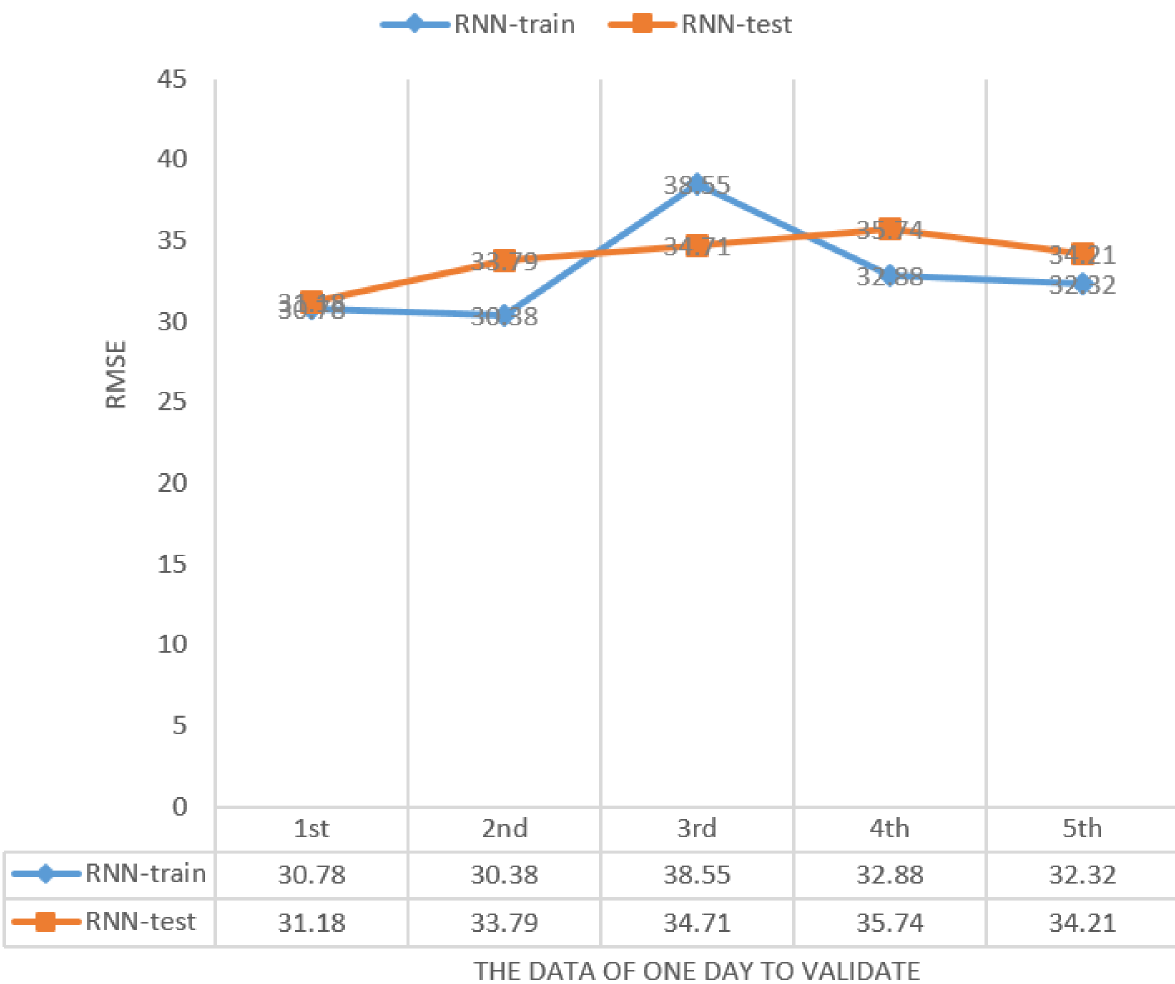


Figure 18: RMSE(root mean square error) of RNN for cross validation

5.2 Influence of Data Size

The influence of data size is also an interesting point for us. In other words, we want to find out that how much data we need to build a proper model, which could reduce the cost of collecting data. The figure 19 demonstrates the relation between errors of the RNN model and training data size. In order to let the results comparable, we used the same testing data set in this experiment. Due to the fact that the data was lost during in collection and data merging in preprocessing, the data size of each person was different, leading to the situation that the the time period of the data was approximately close to the real time period in this experiment. The training error was increasing when the data size was increasing, while the testing error was decreasing. We found that 24 hours data could build a model which performed comparable to the model built from 4 days' data for this RNN structure.

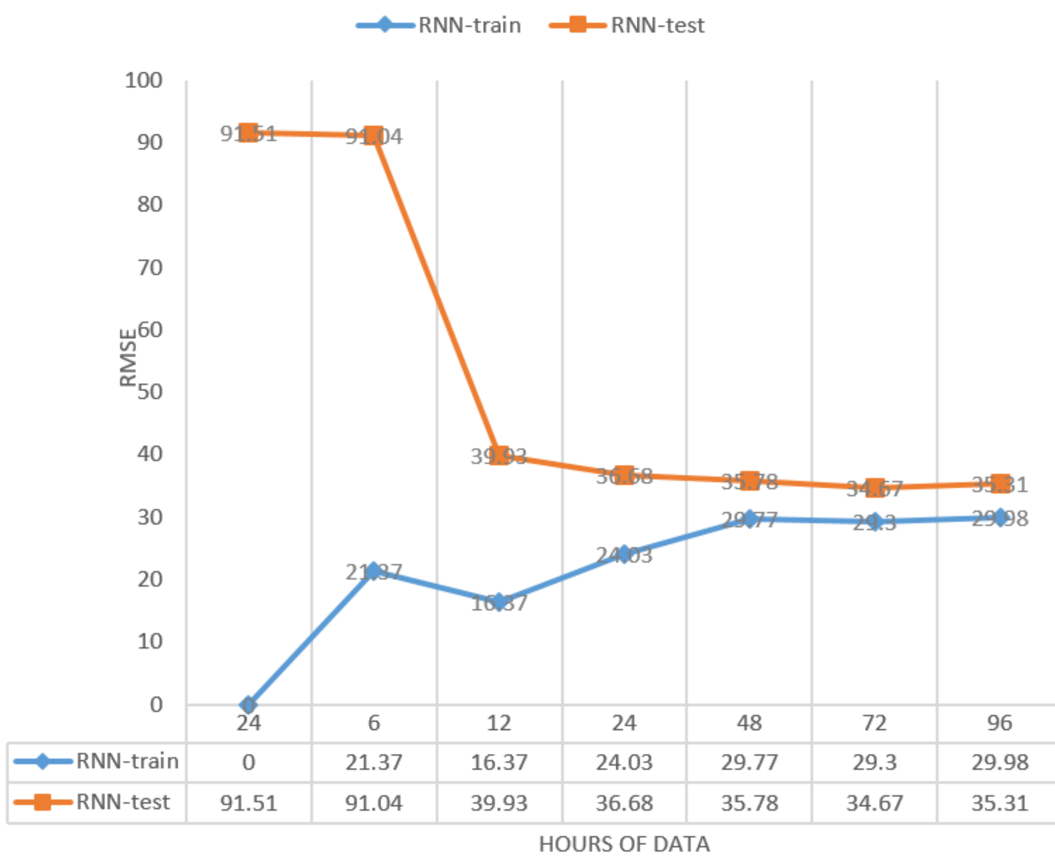


Figure 19: RMSE(root mean square error) of RNN for different number of hidden units

6 Discussion

This paper is an early research of heart rate modeling as a function of body movement. There are still many aspects of this model, which are limited but could be improved.
One limitation of this paper is that the heart rate data was calculated by the Equivital device, which was not very reliable. Because we did not know how the Equivital devices got it and the heart rate data' frequency calculated by Equivital device was just 0.2Hz. In this paper, we chose 1Hz as the target frequency and merged all data we used in SwitchBox study together.As the results, the Low frequency of the heart rate data needs interpolation methods (we used linear interpolation method in this paper) to be aligned with other features, which adds noise to the heart rate data. Although, it is not an issue if we use 0.2 Hz as the target frequency. 0.2 Hz is too low to record body movement details for accelerator data. 25 Hz and 83 Hz are another choice for target frequency. In the experiments of this paper, we validate that the models' performances of 25 Hz data are similar to the models of 1Hz data. Considering the cost of space and time, 1 Hz gets a better choice. 83 Hz could record most details of body movement in accelerator data, but it also would introduce more noise for the low frequency data, such as heart rate and breathe rate data. It is a better option to acquire the heart rate data by calculating the heart rate data from ECG data which is collected by the SwitchBox study. In this way, we could control the heart rate data frequency and understand the heart rate data clearly, because each algorithm of calculating the heart rate from ECG has its own advantages and disadvantages, which could cause the heart rate data differently.
On the other hand, when the high frequency data need to be down sampled, we regarded the first value of the period as the representation of target frequency data, which we introduced before. We believed that this method could represent how the sensor work in low frequency, which is helpful for the regression models performance and adaptability. In other words, if the frequency of raw data is not 25 Hz rather than 1 Hz, the regression models do not have to be changed or fixed. Taking average of the period as the representation of target frequency is another method to down sample. The advantage of this method is that it represents the average level of body movement, which is advantageous for some specific domain, such as body energy consumption.
As we mentioned before, the quality of every participant's data is distinct, which leads to the performances of the regression models for every participant are different. In this paper, we decided the degree of a smooth way based on the some best quality data and then apply the smooth way to all participants' data. The advantage of this method is that the regression models built on this method are universal, which means it would still work well for the new data. Adaptive smooth way is another creative idea. The basic thought of this method is to design three degrees of smooth way, which corresponds to three levels of data quality. This method would be more flexible for the distinct quality data than the original method, because all data could be smoothed in a more proper way by the three specific filter. Due to the reason that this paper is an initial research, we did not consider this method in this project.
In this paper, we built heart rate prediction models from human body movements. However, the heart rate is not simply influenced by activity degree but also affected by other factors such as emotion and illness. One of original aims of this paper is to find episodes where the heart rate deviates from just an exercise based model, which is useful for some specific medical domain. For example some special patients need to be detected whether the heart speeds up caused by physical activity. Another target of this paper is to add the other factors, such as emotions and individual differences to the models, which should improve the model stronger.

So that, we could have a look at the complex heart rate function.

7 Conclusion

This paper built RNN and LSTM regression models to predict the heart rate according to human body movement, which have better performance than the state-of-art regression models, such as ridge regression, K nearest neighbors and Decision Tree.
We found that applying different part of the data as the testing set and the left part as the training set did not not influence the error of RNN and LSTM models, when we applied five folds cross validation methods to evaluate the RNN and LSTM regression models.
Another finding of this paper was that when the size of training data grew and the size of testing data did not change, the training error became bigger and testing error became smaller. One day of training data is enough to build a proper RNN regression models.
Thus, we strongly believed that deep learning methods RNNs and LSTMs are better regression models than some state-of-art regression models in term of predicting the heart rate according human body movement. The RMSE of RNNs and LSTMs are 2.42 and 2.99 less than ridge regression models respectively which is the best one among the state-of-art models this paper used.
For this paper, there is some future work, which we recommend to do. One is to calculate the heart rate from the raw ECG data rather than use the consequences of Equivital devices. Another one is to use an adaptive Kalman filter instead of one fixed Kalman filter, which we mentioned in the discussion section. The last one is to optimize the structure of RNNs and LSTMs instead of using regular structure.

References

[1] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[2] Wanpracha Art Chaovalitwongse, Ya-Ju Fan, and Rajesh C Sachdeo. On the time series k-nearest neighbor classification of abnormal brain activity. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37(6):1005–1016, 2007.

[3] Teddy M Cheng, Andrey V Savkin, Branko G Celler, Steven W Su, Lu Wang, et al. 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.

[4] RS De Fries, M Hansen, JRG Townshend, and R Sohlberg. Global land cover classifications at 8 km spatial resolution: the use of training data derived from landsat imagery in decision tree classifiers. *International Journal of Remote Sensing*, 19(16):3141–3168, 1998.

[5] Philippe Esling and Carlos Agon. Time-series data mining. *ACM Computing Surveys (CSUR)*, 45(1):12, 2012.

[6] Mark A Friedl and Carla E Brodley. Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment*, 61(3):399–409, 1997.

[7] Tak-chung Fu. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1):164–181, 2011.

[8] Matthias Füller, Ashok Meenakshi Sundaram, Melanie Ludwig, Alexander Asteroth, and Erwin Prassler. Modeling and predicting the human heart rate during running exercise. In *International Conference on Information and Communication Technologies for Ageing Well and e-Health*, pages 106–125. Springer, 2015.

[9] Pierre Geurts. *Contributions to decision tree induction: bias/variance tradeoff and time series classification*. PhD thesis, University of Liège Belgium, 2002.

[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[11] Wilhelmina Maria Jansen et al. *Pituitary hormone secretion in familial longevity: The Switchbox Study*. Department of Gerontology and Geriatrics, Faculty of Medicine, Leiden University Medical Center (LUMC), Leiden University, 2016.

[12] D Kugiumtzis, OC Lingjærde, and N Christophersen. Regularized local linear prediction of chaotic time series. *Physica D: Nonlinear Phenomena*, 112(3):344–360, 1998.

[13] Robert K Lai, Chin-Yuan Fan, Wei-Hsiu Huang, and Pei-Chann Chang. Evolving and clustering fuzzy decision tree for financial time series data forecasting. *Expert Systems with Applications*, 36(2):3761–3773, 2009.

[14] Thomas Ying-Ching Lo and Yuh Show Tsai. Ekg based heart rate monitor with digital filter and enhancement signal processor, March 2 1999. US Patent 5,876,350.

[15] Mette S Olufsen, April V Alston, Hien T Tran, Johnny T Ottesen, and Vera Novak. Modeling heart rate regulation—part i: sit-to-stand versus head-up tilt. *Cardiovascular Engineering*, 8(2):73–87, 2008.

[16] Hans-Peter Piepho. Ridge regression and extensions for genomewide selection in maize. *Crop Science*, 49(4):1165–1176, 2009.

[17] Lior Rokach and Oded Maimon. *Data mining with decision trees: theory and applications*. World scientific, 2014.

[18] S Rasoul Safavian and David Landgrebe. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3):660–674, 1991.

[19] Steven W Su, Weidong Chen, Dongdong Liu, Yi Fang, Weijun Kuang, Xiaoxiang Yu, Tian Guo, Branko G Celler, and Hung T Nguyen. Dynamic modelling of heart rate response under different exercise intensity. *The open medical informatics journal*, 4:81, 2010.

[20] Hongyu Sun, Henry X Liu, Heng Xiao, Rachel R He, and Bin Ran. Short term traffic forecasting using the local linear regression model. In *82nd Annual Meeting of the Transportation Research Board, Washington, DC*, 2003.

[21] Olga Troyanskaya, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein, and Russ B Altman. Missing value estimation methods for dna microarrays. *Bioinformatics*, 17(6):520–525, 2001.

[22] Drouet Warren Vidrine, Jack Gordon Kisslinger, and Joseph Michael Brown. Heart rate monitor and method, January 25 2000. US Patent 6,018,677.

[23] Xinying Yu and Shie-Yui Liong. Forecasting of hydrologic time series with ridge regression in feature space. *Journal of Hydrology*, 332(3):290–302, 2007.

[24] Maria S Zakynthinaki. Modelling heart rate kinetics. *PloS one*, 10(4):e0118263, 2015.

[25] Min-Ling Zhang and Zhi-Hua Zhou. Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, 40(7):2038–2048, 2007.