International Journal of Advanced Science and Technology
Vol.95 (2016), pp.89-96
http://dx.doi.org/10.14257/ijast.2016.95.08

Health Diagnosis Based on Analysis of Data Captured by Wearable Technology Devices

Haolun Zhang

Harbin No.3 High School, Harbin, China
zhanghaolun566@gmail.com

Abstract

Modern technology is developing by leaps and bounds, and more and more people begin using wearable technology devices. Recently, users have been using this kind of devices such as Fitbit, Apple Watch and Samsung wrist trackers so as to keep track of their health data such as consumed calories, running miles and steps, and even sleeping time. Many users wear their devices nearly 24/7, providing a thorough weekly health analysis in the devices' applications installed in their mobile phones. However, few people really use wearable devices to diagnose or identify common diseases which can be captured by the fluctuations or major changes in data captured by the devices. Hence, integrating with machine learning technology, we attempt to figure out a solution to detect and diagnose some diseases based on the daily health data collected by wearable devices. Aiming at this, we collected data and experimented using a classification-based machine learning method, namely Support Vector Machine, to simulate a verisimilar ambient to monitor certain users' health conditions.

Keywords: *Health Diagnosis, Wearable Technology Devices, Classification, Machine Learning Method, Support Vector Machine*

1. Introduction

Technically, various wearable devices can keep track of users' daily activities such as eating, workout, working and sleeping. Meanwhile, with the upgrade of these devices, people increasingly rely on them to get diet instructions or calories consuming recommendation. Moreover, since some of these wearable trackers can detect their users' heart rate, users can sometimes be aware of their own health conditions so as to decide whether to continue workout or not. Clearly, these devices have made our lives more convenient and guided to a great extent [1-8].

Meanwhile, common diseases like flus, cardiovascular diseases, heart attack or Inflammatory Bowel Disease (IBD) keep torturing a wealth of patients. If we can actually utilize the database provided by wearable devices to detect, and diagnose such diseases earlier, most patients can better prepare for the diseases by taking pills in advance, taking rest or asking for help from doctors. Indeed, we need a mechanism to synthesize the data tracked by wearable devices, analyze them and eventually figure out the most likely diseases that the users catch. However, there is currently no mature methods for disease analysis based on health data of wearable devices.

In a nutshell, wearable devices provide us with sufficient resources, which can be used as the repertoire of our data to conduct our experiments. Following this, we proposed a methodology based on wearable technology devices to analyze data and diagnose diseases. We use the data provided by wearable devices and classify them to several common diseases.

This paper is organized as follows: Section 2 presents our data's sources, types of health data and types of diseases to be diagnosed. Section 3 introduces our classification of data and successive experiments' procedures. Section 4 discusses some flaws existing in our

ISSN: 2005-4238 IJAST

experiments and corresponding future solutions. Along with these, our core code for experiment will also be attached.

2. Disease Diagnosis Based on Data Analysis of Wearable Devices

2.1. Types of Health Data Captured by Wearable Devices

The health data that are able to collect include age, heights, weights, consumed calories, walking steps & velocity, heart rates, body temperatures and sleeping hours. These nine kinds of data roughly provide the overall condition of wearable device users' health. A wealth of data will be collected by the wearable devices daily, so daily collected data will suffice to our purpose.

There are many wearable devices products that can capture users' health data in market. Among all of them, we chose Fitbit tracker, the most popular wearable devices in market are Fitbit. The idea of this kind of device is really ingenious that those trackers will keep capturing and recording users' health data such as calories, walking steps and heart rates. The devices will automatically integrate these data and synchronize them with users' mobile phones where Fitbit's apps are installed. Afterwards, the users can vividly view their own health data on their mobile phones. Furthermore, with this broader platform, the application can also analyze the data of a relatively longer term (*e.g.,* a week, a month) which is able to show the users' collective health conditions.

2.2. Types of Disease to Be Diagnosed

We consider three kinds of common diseases: cold, cardiovascular disease, and intestine disease. We chose these three diseases because they are common in people's daily life. Besides, these three diseases have relatively distinct symptoms and remarkable body data changes which are highly related to our data specimen. Hence, based on these two reasons, we decide to focus our target mainly on these three sorts of diseases.

After we have data and disease categories, we need to clarify the relationships between symptoms and body data. Matching them is comparatively straightforward. For example, when people catch cold, their walking steps will definitely decrease correspondingly; when people get intestine disease, people will lose appetite, so it is not hard to find out that individuals will consume less calories. Extrapolating from this, we can easily summarize the relationship between body data and diseases symptoms.

When catching a cold, people will experience elevated heart rate. The heart functions differently when the body feels stressed or when the body is fighting against an infection so as to help the body fight the stress and infection [9]. To do so, the heart accelerates the rate of beat so as to facilitate the circulation of oxygen and immune cells which are needed to initiate the healing process. Bacteria or infection that causes a disease and is accompanied with fever often causes the heart rate to rise.

Cold caused by virus (like flu), will cause people's body temperature to rise. People's body temperatures rise having the flu as a way of fighting off the influenza virus; viruses propagate at 98.6 degrees Fahrenheit which is normal body temperature. When people's bodies heat up, then it will be tougher for virus to survive. Fevers also trigger people's immune systems to eradicate the illness.

Heart attack: Depending on the type of heart attack patients are suffering and where the blockages are, heart rates are prone to be slower or faster than normal levels.

Gastritis (Intestine): Loss of appetite, which means less calories taken from food. This can be easily monitored as a significant drop in calories from food.

Patients' walking steps will decrease in any cases no matter the type of disease due to the decrement of mobility.

2.3. Introduction of SVM Classifier

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis [10]. Given training data assigned to several separate classes, SVM algorithm can model these data and collectively categorize them to different classes. With this model, the SVM itself will then be able to tell which class the new data should belong to.

We use supervised learning to train the mode, *i.e.*, the data must be labeled and clustered before being imported into the SVM classifier. The idea of SVM classifier is straightforward: teaching computer to recognize things. This can be illustrated by a simple example, provided that we have hundreds of pictures. One cluster of them are all cat pictures, and the other cluster are all dog pictures. So our job is to use the SVM classifier to let the computer study these pictures and figure out a distinctive way to discern dogs from cats and vice versa. Here in our experiment, we would use SVM to differentiate the changes of some data correspond to which kind of disease. For sake of this, we think that the SVM classifier will suffice to our aim and the requirements of experiment.

3. Experiments

3.1. Division of Health Data

Our first step is to collect various data so as to construct a database for machine to learn. Hence, we looked up different resources about human heights, weights and ages division and divide different groups of health conditions as follows:

(1) Height stratification:

Short: Male < 4’11’’ Female < 4’7’’

Medium: Male 4’11’’— 6’3’’ Female 4’7’’— 5’11’’

Tall: Male > 6’3’’ Female > 5’11’’

(2) Age group stratification:

Teenagers: <18

Middle Age: 19-54

Senior: 55+

(3) Weight stratification associated with heights

Female Height to Weight Ratio				Male Height to Weight Ratio			
Height	Low	Target	High	Height	Low	Target	High
4' 10"	100	115	131	5' 1"	123	134	145
4' 11"	101	117	134	5' 2"	125	137	148
5' 0"	103	120	137	5' 3"	127	139	151
5' 1"	105	122	140	5' 4"	129	142	155
5' 2"	108	125	144	5' 5"	131	145	159
5' 3"	111	128	148	5' 6"	133	148	163
5' 4"	114	133	152	5" 7"	135	151	167
5' 5"	117	136	156	5' 8"	137	154	171
5' 6"	120	140	160	5' 9"	139	157	175
5' 7"	123	143	164	5' 10"	141	160	179
5 '8"	126	146	167	5' 11"	144	164	183
5' 9"	129	150	170	6' 0"	147	167	187
5' 10"	132	153	173	6' 1"	150	171	192
5' 11"	135	156	176	6' 2"	153	175	197
6' 0"	138	159	179	6' 3"	157	179	202

Height = Feet and Inches - Weight = Pounds.

Figure 1. Relations between Stratified Heights and Weights (based on ratio index)

(4) Walking steps (per day) stratification

Insufficient: <5000 steps

Medium: 5000-25000 (average level: 15000)

Excessive: >25000

(5) Velocity stratification:

Slow: <3km/h

Normal: 3-7km/h (average velocity: 5km/h)

Fast: >7km/h

(6) Heart rate stratification:

Slow: <55 times/min

Normal: 55-100 times/min (average HR: 75 times/min)

Fast: >100 times/min

(7) Body temperature stratification:

Low: <36.3

Normal: 36.3-37.2 (average: 37.0)

High: >37.2

(8) Sleeping hours stratification:

Little: <5

Normal: 5-10 (average: 7.5)

Excessive: >10

(9) Consumed calories stratification:

Insufficient: <500

Normal: 500-3000 (average: 1500)

Excessive: >3000

With such classifications, we can utilize them to represent and simplify the relationship existing among different health data.

Basically, we categorize all data into three levels: low, medium, and high. Besides, we designate three digits 0, 1, 2 to represent these three levels for sake of convenience.
We collected following data: ages, genders, heights, weights, walking steps, walking velocities, heart rates (HR), body temperatures, sleeping time, and consumed calories. As mentioned before, three digits 0, 1, 2 are designated to all these types of data.

3.2. Data Collection

We collected health data from 100 people and 25 of them are totally health, 25 of them are diagnosed with flus, 25 of them got chronic cardiovascular diseases, and the rest 25 had intestine disease. These eight kinds of data were collected from these 100 people, and as mentioned before, we represented the heart rate and other sorts of data with the same mechanism of designation and definition. For instance, we can define all heart rates data less than 55 times/min to be 0, 55-100 times/min to be 1 and larger than 100 times/min to be 2. Then we iterated this process with all patients' data and eventually we got a table composed entirely of numbers 0, 1, and 2 representing the classification of different people's different health data.

3.3. Modeling of SVM Classifier

In 4.2, we mentioned that we transferred all health data into three numbers 0,1, and 2. So here we have a number table, which can be used as our training data. The training data aim to let our computer study the pattern of them and figure out a way to classify new data into three existing classified diseases. While doing this, we need to import this number table into Eclipse. But this table was transferred from Excel whose data layout and format are incompatible with those in Eclipse. Under this circumstance, we need to write codes that can read each line of our number table and revise its format to import into Eclipse. The codes to read lines and change data format are as follows.

Algorithm 1. Training algorithm

Input: health data, penalty parameter C
Output: health diagnosis model

```
Samples={}
Foreach person
    X[i]←extract features for health data
    Samples=Samples+X[i]
Initialization: W⃗ set 0 b=0
For each sample X in Samples
    Updata W⃗
    Update b
return W⃗ , b
```

Then the format of the training data matches the requirements of SVM Classifier. We can import data files into the Eclipse. Among these files is a Java file called libsvm.jar. This file is basically the prerequisite of the whole experiment, and with this file, we are able write and run the code for the SVM classifier.

Algorithm 2. Predict health diagnosis algorithm

Input: health data, $\vec{W}$, b
Output: health diagnosis result

Samples={}
Foreach person
    X[i]←extract features for health data
    Samples=Samples+X[i]
For each sample X in Samples
    Calculate $\vec{W} \cdot X + b$
return predict result

3.4. Experimental Results

The learning process in our experiment is shown in Table 1.

Table 1. The Learning Process of SVM

	nu	obj	rho	nSV	nBSV
#iter = 38	0.5054	-16.5905	0.4482	30	19
#iter = 55	0.5423	-17.1933	-0.6871	30	18
#iter = 35	0.5287	-16.8928	-0.3538	29	22
#iter = 43	0.4331	-13.8320	-0.7219	27	19
#iter = 41	0.4281	-13.2709	-0.1674	26	15
#iter = 53	0.3413	-10.5388	0.0766	23	12

After 53 iteration, we get the total nSV = 75, and 93.9394% on the main evaluation metrics Accuracy.

Normally, a method can be viewed as accurate if its accuracy is over 90%. As we can see from the experimental results, the accuracy of our method to diagnose diseases is approximately 93.9394%, which indicates that the data provided to SVM classifier can predict new data targets' disease with over 93% accuracy. Hence, we can conclude the outcome of our experiment, that is, the SVM classifier is accurate enough to predict the disease based on the new data provided.

4. Conclusion

The SVM Classifier was utilized as a machine to learn the data and analyze the pattern of them. We provided the data that the classifier needed: training data was provided to teach the SVM Classifier the analysis of data, while the test data was used to verify the classifier's own analysis.

Throughout our experiment, we collected health data as the resource, and we used SVM classifier as the method to study these data and classify them into three distinctive classes. The result of our experiment is positive because the SVM classifier is accurate enough to classify test data and thus it can predict diseases with more data provided.
However, we have to acknowledge that there are some flaws existing in our methodology.

1. The experimental data are not sufficient. Our data are simply not ample enough to conclude the result of our experiments. They are not able to accurately judge the precision of the SVM classifier.

2. The types of diseases are not enough, and there should be more variables and connections in these diseases and data. We just focused on these three diseases because they are among the most common diseases, and they also have distinctive and discernable symptoms.

3. The types of health data are not enough. We were only able to collect certain kinds of data (*e.g.*, Heart Rate and calories) because these are the only ones that can be captured by wearable devices. And we mainly limited our concentration to these data rather than a broader scope.

In the future, we will focus on the following solutions to these problems as listed below:

4. We will collect more data from larger population of targets so as to enlarge our database for training SVM classifier to analyze patients' diseases.

5. Relevant search on more types of diseases in addition to these three that have already been covered in our study will be needed. And after that, we will include them into our classes of diseases.

We will use more sophisticated and advanced wearable devices that are able to capture and detect more sorts of health data.

References

[1] C.-K. Kenneth Chan and S. Halgamuge, "A New Generalized Growth Threshold for Dynamic SOM for Comparing Average Mutual Information and Oligonucleotide Frequency as a Species Signature", International Journal of Bio-Science and Bio-Technology, vol. 1, no. 1, **(2009)** December, pp. 1-10.

[2] H. Trung Huynh, J.-J. Kim and Y. Won, "Classification Study on DNA Microarray with Feedforward Neural Network Trained by Singular Value Decomposition", International Journal of Bio-Science and Bio-Technology, vol. 1, no. 1, **(2009)** December, pp. 17-24.

[3] D. Ganguly, S. Chakraborty and T.-h. Kim, "A Cognitive Study on Medical Imaging", International Journal of Bio-Science and Bio-Technology, vol. 2, no. 3, **(2010)** September, pp. 1-18.

[4] L. Chong Hoon, "Biomechanical Analysis of the Physical Effects of Riding the X-lider", International Journal of Bio-Science and Bio-Technology, vol. 3, no. 4, **(2011)** December, pp. 69-78.

[5] G. Park, H. An and S. Lee, "A Study on Effective Amplification Gain in Cochlea Dead Region with Word Recognition Score", International Journal of Bio-Science and Bio-Technology, vol. 4, no. 3, **(2012)** September, pp. 1-8.

[6] Z. Al-Ameen, G. Sulong and Md. G. Md. Johar, "Enhancing the Contrast of CT Medical Images by Employing a Novel Image Size Dependent Normalization Technique", International Journal of Bio-Science and Bio-Technology, vol. 4, no. 3, **(2012)** September, pp. 63-68.

[7] B.-H. Kim and D.-U. Cho, "Improved Liver Function Effect Analysis of Drinking the Lacquer Beverage Using Speech Signal Analysis", International Journal of Bio-Science and Bio-Technology, vol. 5, no. 5, **(2013)**, pp. 65-74.

[8] G. Kaur, G. Singh and V. Kumar, "A Review on Biometric Recognition", International Journal of Bio-Science and Bio-Technology, vol. 6, no. 4, **(2014)**, pp. 69-76.

[9] MDHealth, Elevated Heart Rate When Sick, http://www.md-health.com/Elevated-Heart-Rate-When-Sick.html.

[10] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers", Neural processing letters, vol. 9, no. 3, **(1999)**, pp. 293-300.

