An Internet of Things (IoT) Application for Predicting the Quantity of Future Heart Attack Patients

Fizar Ahmed
Ph.D Student
Business Information Department
Corvinus University of Budapest
Hungary

ABSTRACT

Now days the heart disease is the leading cause of death worldwide. It is a complex task to predict the heart attack for a medical practitioner since it is required more experience and knowledge. However, heart rate monitoring is the most important scale of measurement that is the influence factor for heart attack with other health fitness like blood pressure, serum cholesterol and level of blood sugar. In the era of rapid revolution of Internet of things (IoT), the sensors for monitoring heart rate are growing in availability to patients. In this paper, I explained the architecture for heart rate and other data monitoring technique and I also explained how to use a machine learning technique like kNN classification algorithm to predict the heart attack by using the collected heart rate data and other health related perimeter.

Keywords

Internet of Things, heart attract, machine learning, k nearest neighbour

1. INTRODUCTION

Internet of Things (IoT) signifies a general concept for the ability of network devices to sense and collect data from the world around us, and then share that data across the Internet where it can be processed and utilized for various curious purposes. Predictions for the impact of IoT on the Internet and economy are remarkable, with some expecting as many as 100 billion connected IoT devices and a global economic impact of more than \$11 trillion by 2025. IoT for healthcare is playing significant role for monitoring some health activities and diagnosing the problem related with these activities. The devices connected with internet have been hosted to patients in various forms. Whether data comes from connected blood pressure monitoring devices, electrocardiograms, temperature or blood sugar levels checking, tracking regular health information is vital for some patients. Many of these measures require follow-up interaction with a healthcare professional. These data always goes to the remote cloud, many type of analyses report are produced on the cloud and send the relevant results to the patient or healthcare personnel by using mobile phone or other active devices.

An enormous healthcare data generated by the Internet of Things (IoT) are considered of high business value. Data mining algorithms can be applied to IoT to expose unknown information from these data. Using data mining knowledge, technique and application, including classification, clustering, association analysis, time series analysis and outlier analysis can make also a prediction report for future diseases of a patient. Government or respective authorities may use this analysis and can take the further step to cure or prevention for the diseases. In this paper, I have explained the process of data collections, necessary tools and methods for using these data for data analysis. For this instance, a demo data analysis report also given here. A popular data mining algorithm k Nearest Neighbour (k-NN) used here for predicting the future patients of heart diseases using the applicable heart related data.

2. RELATED WORKS

Zhao, Wang, & Nakahira, 2011 analysed the possibility and related matters of providing advanced services of human health management and they have given also a research direction of medical technology on IOT [1]. Different health related sensors and protocol they have discussed and raised some issues need to be solved. The Decision Support and Home Monitoring System were designed by Chiuchisan and Geman in 2014 [2]. This system contributes in diagnosis, home monitoring, medical treatment, medical prescriptions, rehabilitation and progress of his patients with Parkinson's disease. Their paper suggested for future extended research on Neurological Disorders with the help of eHealth and Internet of Things technology.

15 attributes are related with predicting heart attract and data mining technique like ANN, time series, clustering and association rules can be approached for this prediction (Soni & Sharma, 2011) [3]. For accuracy of the Decision Tree and Bayesian Classification they suggested to apply genetic algorithm to reduce the actual data size to get the best subset of attribute appropriate for heart disease prediction. Durairaj & Ranjani, 2013 also compared many approaches and different tools and its impact on the healthcare sector [4]. They used the data mining application for identifying and publishing relevant healthcare information. For diagnosing breast cancer, data mining technique can be used by reducing the number of features of diagnosis. Ahmad, 2013 applied the Adaptive Neuro Fuzzy Inference system (ANFIS) with the reduced dataset. He found 98.24% accuracy from his approach [5].

Heart rate variability has substantial potential to judge the role of autonomic nervous system fluctuations in normal healthy individuals and in patients with various cardiovascular and non-cardiovascular disorders. HRV studied by Task Force of the European Society of Cardiology (1996) and boost up the understanding of physiological phenomena, the actions of medications, and disease mechanisms [6]. This Task force suggested - "Longitudinal studies are needed to determine the sensitivity, specificity, and predictive value of HRV in the identification of individuals at risk for subsequent morbid and mortal events". Stein & Conger (1994) approached to measurement of heart rate variability (HRV), time and frequency domain [7]. Their calculations are based on the analysis of inter beat intervals of normal beats determined from a routine 24 hours ambulatory electro cardiogram,

Machine learning is now a day important in IoT. Different type of data mining results like prediction or others from IoT data can be measured by machine learning tools [8]. In his paper Yue Xu (Recent Machine Learning Applications to Internet of Things) described some technique and example scenarios regarding IoT data and its use on the machine learning tools. Masethe & Masethe, 2014 used Classification algorithm for predicting of heart diseases [9]. They used some data mining algorithms such as J48, Naïve Bayes, REPTREE, CART, and Bayes Net for predicting heart attacks. Their research result shows prediction accuracy of 99%. Beside heart diseases data analysis technique also uses for other health related issues. One of them is spinal cord injuries (SCI). Kraft, Desouza & Androwich, 2003 published some case studies related in this field. Including data cleaning, aggregation, integration, they discussed whole process to find the result from the data analysis [11]. They used artificial neural networks to predict the length of stay is presented.

Using of sensor devices for physiological monitoring of vital signs in healthcare services. Banaee, Ahmed, & Loutfi, (2013) reviewed the latest methods and algorithms used to analysed data coming from wearable sensor devices [12]. They used some common data mining technique such as anomaly detection, prediction and decision making when considering in particular continuous time series measurements. Finally, based on this literature review, a number of key challenges they defined for data mining methods in health monitoring systems. Alemdar, & Ersoy (2010) were investigated some challenges of wireless sensor network in healthcare [13]. They provided several state of the art examples together with the design considerations like - "unobtrusiveness, scalability, energy efficiency, security and also provide a comprehensive analysis of the benefits and challenges of these systems". Security and privacy of IoT data in healthcare is most important. A serious social unrest may arise due to the fear that such devices used for monitoring and tracking individuals by government agencies or other private organizations (Al Ameen & Kwak, 2012) [14].

3. THE REASON OF HEART ATTACK

Cardiovascular disease is the leading global cause of death. The statistics prepared by the American Heart Association, the Centres for Disease Control and Prevention, the National Institutes of Health and other government sources (Mozaffarian & Huffman 2015) accounting for 17.3 million deaths per year, a number that is expected to grow to more than 23.6 million by 2030 [16]. Canto & Iskandrian, 2003 analysed some risk factor of heart diseases [17]. Some traditional myths relating with heart rate and heart attack are published in health related website (http://www.webmd.com) are as follows:

- A normal heart rate is 60-100 beats per minute. However, heart rate higher than 76 beats per minute when in resting may be linked to a higher risk of heart attack.

- Having an irregular heartbeat doesn't mean having a heart attack. But if it's a new symptom, or if you have chest pains or problems breathing, may be the preliminary symptom for heart attack.

- Slow rates are only a problem if also pass out, feel dizzy, are short of breath, or have chest pain.

In this paper different health related data with heart rate used for the prediction of heart attack and reveal these myths.

4. METHOD OF DATA COLLECTION

I can divide six types of data into three different levels for predicting heart attack. First one is personal fixed data like age and sex. Second level is periodic data which comes in a day or in a week like blood sugar and Serum Cholesterol. The third data level is live data like blood pressure and heart rate.

Web service technology has been used here for first level data collection. Patients will be registered on the platform by giving all required fixed information. They will have the option to change the perimeter after a certain time.

Wearable devices and home health monitoring IoT devices will be used for monitoring second and third level data collection. The devices are capable enough to transmit data from a patient home to the remote cloud server. Some devices allow having a real time monitoring blood pressure and heart rate of patient's. Some devices have trigger system that give the result after a certain period or by using manual trigger. Wireless connected digital glucometers and cholesterol tester with triggering system use for this purpose. Devices helping in monitoring real time heart rate and blood pressure are indeed a big part of IoT. Live data always go to the remote cloud using these sensor based wearable devices.

A conceptual diagram of IoT-based heart attack prediction:
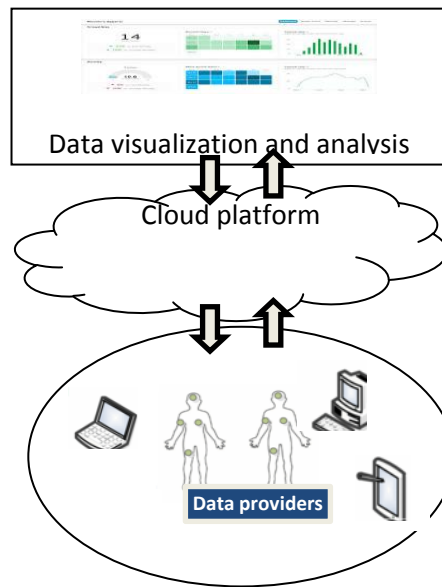
5. DATA MINING ALGORITHM

Supervised machine learning algorithm k-Nearest Neighbour (kNN) has been used for analysing the dataset. It is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). A case is classified by a majority vote of its neighbours, with the case being assigned to the class most common amongst its K nearest neighbours measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbour.

Distance functions

$$\text{Euclidean} \quad \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

$$\text{Manhattan} \quad \sum_{i=1}^{k}|x_i - y_i|$$

$$\text{Minkowski} \quad \left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$$

All three distance measures are only valid for continuous variables. In the instance of categorical variables the Hamming distance must be used. It also brings up the issue of standardization of the numerical variables between 0 and 1 when there is a mixture of numerical and categorical variables in the dataset.

Hamming Distance

$$D_H = \sum_{i=1}^{k}|x_i - y_i|$$

$$x = y \Rightarrow D = 0$$
$$x \neq y \Rightarrow D = 1$$

X	Y	Distance
Male	Male	0
Male	Female	1

Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise but there is no guarantee. Cross-validation is another way to retrospectively determine a good K value by using an independent dataset to validate the K value. Historically, the optimal K for most datasets has been between 3-10. That produces much better results than 1NN. (Source: http://www.saedsayad.com/k_nearest_neighbors.htm)

6. EXAMPLE DATA SOURCE

An example dataset from UCI Machine Learning Repository [https://archive.ics.uci.edu/ml/datasets/Statlog+(Heart)] from Centre for Machine Learning and Intelligent Systems has been used here. 270 patients dataset has been restructured such a way that keeping in mind the results which are generally obtained from healthcare cloud platform.

The data set consists of 270 observations and 8 variables (out of which 4 numeric variables, 3 binary values and one is id) which are as follows: Patient id, Sex, Age, Resting Blood Pressure, Serum Cholesterol, Fasting blood sugar if more than 120 mg/dl, maximum heart rate within 24 hours, Heart diseases (Prediction)

In real life, more important parameters may be needed to measure the patients who are in danger period for heart attack. I can classify this data field on the base of collection procedure. Here it is in three ways. Two live data blood pressure and heart rate will be collected by using wearable sensor devices. These data will be stored in a separate temporary storage and from there highest heart rate and average blood pressure in a certain period will be recalculated from this temporary storage and finally stored in main database. Main database will consist with patients' static information like age, sex and other periodic information such as blood sugar and serum cholesterol. In this dataset binary values have been used for fasting blood sugar measurement. If the blood sugar level more than 120 mg/dl, this result is treated as one and if this is 120 mg/dl or less than that, this is defined as normal or zero. The class value real of heart disease also measured in binary and noted P for present and A for absence.

Here's the data set looks like:

	A	B	C	D	E	F	G	H
1	P_id	age	sex	resting_blood_pressure	serum_cholestoral	fasting_blood_sugar_more_120	max_heart_rate	heart_disease
2	1	70	1	130	322	0	109	P
3	2	67	0	115	564	0	160	A
4	3	57	1	124	261	0	141	P
5	4	64	1	128	263	0	105	A
6	5	74	0	120	269	0	121	A
7	6	65	1	120	177	0	140	A
8	7	56	1	130	256	1	142	P
9	8	59	1	110	239	0	142	P
10	9	60	1	140	293	0	170	P
11	10	63	0	150	407	0	154	P
12	11	59	1	135	234	0	161	A
13	12	53	1	142	226	0	111	A
14	13	44	1	140	235	0	180	A
15	14	61	1	134	234	0	145	P
16	15	57	0	128	303	0	159	A
17	16	71	0	112	149	0	125	A

7. RESULT AND ANALYSIS

The kNN algorithm is applied to the training data set and the results are verified on the test data set. For this, I would divide the data set into 2 portions in the ratio of 150: 120 for the training and test data set respectively. The value for k is generally chosen as the square root of the number of observations. I have checked the accuracy of the predicted values in prc_test_pred as to whether they match up with the known values in prc_test_labels.

The test data consisted of 150 observations. Out of which 87 cases have been accurately predicted (TN->True Negatives) as Absence (A) in nature which constitutes 58%. Also, 66 out of 150 observations were accurately predicted (TP-> True Positives) as Present (P) in nature which constitutes 44%. Thus a total of 66 out of 150 predictions where TP i.e, True Positive in nature.

There were 22 cases of False Negatives (FN) meaning 22 cases were recorded which actually are present in nature but got predicted as benign. The FN's if any poses a potential threat for the same reason and the main focus to increase the accuracy of the model is to reduce FN's.

There were 19 cases of False Positives (FP) meaning 19 cases were actually absent in nature but got predicted as present.

The total accuracy of the model is 96 %((TN+TP)/150), that means model performance is better.

Table: Result from kNN model

```
Cell Contents
 |-------------------------|
 |                       N |
 |         N / Row Total |
 |         N / Col Total |
 |       N / Table Total |
 |-------------------------|


Total Observations in Table:  150


                 | prc_test_pred
 prc_test_labels |          A |          P | Row Total |
 ----------------|------------|------------|-----------|
               A |         56 |         28 |        84 |
                 |      0.667 |      0.333 |     0.560 |
                 |      0.718 |      0.389 |           |
                 |      0.373 |      0.187 |           |
 ----------------|------------|------------|-----------|
               P |         22 |         44 |        66 |
                 |      0.333 |      0.667 |     0.440 |
                 |      0.282 |      0.611 |           |
                 |      0.147 |      0.293 |           |
 ----------------|------------|------------|-----------|
    Column Total |         78 |         72 |       150 |
                 |      0.520 |      0.480 |           |
 ----------------|------------|------------|-----------|
```

Generally, the k-value is the square root of the observations and in this case I took k=16 which is a near square root of 270.The k-value may be fluctuated in and around the value of 16 to check the increased accuracy of the model

8. CONCLUSIONS

Wearable devices are getting enough receptiveness because of patient mind set with these devices   and also allergy diseases need to be considered with this wearing. Energy related problems like battery failure or other issues that may be cause of power down of the devices that can be sent bad or interrupted result. This result may be the cause of non-accurate analysis report. Subsequently, a mature infrastructure for everything related with this testing and finalizing it for real calculation, must be the first priority of every task.

9. REFERENCES

[1] Zhao, W., Wang, C., & Nakahira, Y. (2011, October). Medical application on internet of things. In *Communication Technology and Application (ICCTA 2011), IET International Conference on* (pp. 660-665). IET.

[2] Chiuchisan, I. U. L. I. A. N. A., & Geman, O. A. N. A. (2014). An Approach of a Decision Support and Home Monitoring System for Patients with Neurological Disorders Using Internet of Things Concepts. WSEAS Transactions on Systems, 13, 460-469.

[3] Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-48.

[4] Durairaj, M., & Ranjani, V. (2013). Data mining applications in healthcare sector a study. International Journal of Scientific and Technology Research, 2(10), 29-35.

[5] Ahmad, M. A. B. (2013). Mining Health Data for Breast Cancer Diagnosis Using Machine Learning.

[6] Task Force of the European Society of Cardiology. (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use. Eur Heart J, 17, 354-381.

[7] Stein, P. K., Bosner, M. S., Kleiger, R. E., & Conger, B. M. (1994). Heart rate variability: a measure of cardiac autonomic tone. American heart journal, 127(5), 1376-1381.

[8] Xu, Y. Recent Machine Learning Applications to Internet of Things (IoT).

[9] Masethe, H. D., & Masethe, M. A. (2014, October). Prediction of heart disease using classification algorithms. In Proceedings of the world congress on engineering and computer science (Vol. 2, pp. 22-24).

[10] Jothi, N., & Husain, W. (2015). Data Mining in Healthcare–A Review. Procedia Computer Science, 72, 306-313.

[11] Kraft, M. R., Desouza, K. C., & Androwich, I. (2003, January). Data mining in healthcare information systems: case study of a veterans' administration spinal cord injury population. In System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on (pp. 9-pp). IEEE.

[12] Banaee, H., Ahmed, M. U., & Loutfi, A. (2013). Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. Sensors, 13(12), 17472-17500.

[13] Alemdar, H., & Ersoy, C. (2010). Wireless sensor networks for healthcare: A survey. Computer Networks, 54(15), 2688-2710.

[14] Al Ameen, M., Liu, J., & Kwak, K. (2012). Security and privacy issues in wireless sensor networks for healthcare applications. Journal of medical systems, 36(1), 93-101.

[15] Rose, K., Eldridge, S., & Chapin, L. (2015). The internet of things: An overview. The Internet Society (ISOC), 1-50.

[16] Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., ... & Huffman, M. D. (2015). Executive summary: heart disease and stroke statistics—2015 update. Circulation, 131(4), 434-441.

[17] Canto, J. G., & Iskandrian, A. E. (2003). Major risk factors for cardiovascular disease: debunking the only 50% myth. Jama, 290(7), 947-949.