Documentation

Saishradha Mohanty

July 12, 2018

Contents

1 Introduction

This article is a subpart of an IOT based research work based on the paper titled, **"A Simulation Study of Response Times in Cloud Environment for IoT-based Healthcare Workloads"**. This paper suggests a method to employ sensors in the form of wearables on the patients (or the general public) in order to continuously monitor their health conditions.
This article would mainly focus on the part where the particular health parameter values retrieved from the sensors are mapped to some values in some universal range and how these parameter values are assigned weights according to the level of severity of the health problem in order to obtain a health score.

2 Brief overview

For monitoring the health condition, we take into account three different parameters: Temperature, Heart Rate and Blood Pressure.
Since, these parameters have different units, these have to be brought into a unitless common scale, so we propose a method to map these values to a unitless scale, from where they can be rescaled into the desired common range.
The health condition has been divided into four categories: Normal, Low Severity, Grave and Critical.
The next task is to assign weights to these parameter scores according to the level of their severity so as to obtain a health score for the patient. For that we have proposed a geometrical model that makes use of the concept of a sphere.
The two methods have been explained in detail in the next upcoming sections of the article.

3 Getting health score values

First we need to process the data from the sensors in some way such that they can be used for obtaining scores. So, before going for the score calculation, we need to do the following:

3.1 Mapping health values to their respective scores

1. Get many values of a given parameter for the same patient.
2. Obtain the mean of the obtained data.
3. Obtain the deviation of the data points from the set normal value for the parameter.
4. Name those obtained deviations for the concerned parameter as σ_x where x is the label of the parameter (out of 1, 2 or 3).

3.2 Formula for categorization of the parameters into severity cases and their respective mapping

1. A relation can be drawn out between σ_x and the standard deviation (σ) of the distribution of the respective parameter.
2. This we derive by looking at the range of modified standard deviation of the parameter observations where the deviation will be seen from the supposed mid point of normal values of parameters.
3. The supposed mid point are:
   - 99°F for Temperature
   - 75(/min) for Heart Rate

- 120(mmHg) for Blood Pressure

4. The relation can be obtained by writing:
$$\Sigma(x_i - t)^2/N$$
(can be known as σ'^2) (where t is the supposed mid point and x_i are the observations of concerned parameters) as
$$\Sigma((x_i - \mu)^2 + (\mu - t)^2)/N)$$

5. By expanding the above equation we get 3 terms,
$$\Sigma(x_i - \mu)^2/N + \Sigma(x_i - \mu)(\mu - t)/N + \Sigma(\mu - t)^2/N$$
(where μ is the calculated mean of the sensor data).

6. By the properties of mean the 2nd term of the 3 terms gets canceled out.

7. The final obtained equation appears as
$$\Sigma(x_i - \mu)^2/N = \Sigma(x_i - \mu)/N + \Sigma(\mu - t)^2/N$$
which can be written simply as
$$(\sigma'_x)^2 = (\sigma_x)^2 + \Sigma(\mu - t)^2/N$$

8. So, the required equation for mapping is:
$$(\sigma'_x)^2 = ((\sigma_x)^2 + \Sigma(\mu - t)^2/N)$$
$$x\_score = \sqrt{(\sigma_x)^2 + (\mu - t)^2}$$
where σ'_x is the required mapped value.

9. We can do the mapping according to the value of $(\sigma'_x)^2$ by obtaining appropriate range values for the severity cases.

3.3 How to Map

1. In the above obtained equation,
$$x\_score = \sqrt{(\sigma_x)^2 + (\mu - t)^2}$$
since all the values are data dependent, it is difficult to directly calculate the range limits, also since the spread of the observations about the mean(variance) can be anything, upto any extent, it is tough to determine the maximum limit for the concerned severity case directly.

2. But since standard deviation and variance are all positive terms, the minimum value of the concerned case can be determined easily.

3. We need to keep the ranges continuous to avoid missing out any possibility, therefore by this the lower limit of a given case can be made the upper limit of the previous case.

4. This method shall be explained with example in the upcoming steps which would provide more clarity to the idea.

5. In the reference paper, we have the required ranges in terms of the observation values.

HEALTH STATUS	TEMP(°F)	HEART RATE (/min)	BP (mmHg)
Normal	98-100	65-85	101-140
Low Severity	97, 101	85-100, 50-65	141-170, 71-100
Grave	96, 102	100-110, 40-50	171-200, 40-70
Critical	95, 103	110-120, 30-40	200-220, 20-40

Table 3.1: Shown above is the table showing ranges of parameters for different severity cases for original data

6. For minimum limit value, the minimum value of $(\sigma_x)^2$ can be zero and for $\Sigma(\mu - t)^2/N$, the least value for the concerned severity case will be obtained when μ is equal to the upper bound of the previous range with t equal to the supposed mid value of the previous range.

7. Example for explaining point number 6 is:

    - From the table above, let us consider the temperature parameter, for example we want to find the range of score for Normal severity, then the lowest value of σ can be 0, over that for the second part of the equation, its least value can be 0, as mentioned above the highest value of σ can't be determined directly, therefore we look for the lower limit of the next severity case which can become the upper limit for its preceding severity case.
    - So, the for lower limit of the next severity case, $\sigma = 0$ and the second term of the equation would be $|97 - 99| = 2$, hence the lower limit of the low severity case is 2 and hence it becomes the upper limit of the normal case (with 2 excluded).
    - In the same way, the range for other cases can be obtained for different parameters.

8. Hence the table for parameter scores would look like:

HEALTH STATUS	TEMP_SCORE	HEART RATE_SCORE	BP_SCORE
Normal	0-2	0-10	0-21
Low Severity	2-3	10-25, 21-50	141-170, 71-100
Grave	3-4	25-35	50-80
Critical	above 4 (ideally 4-5)	above 35 (ideally 35-50)	above 80 (ideally 80-110)

Table 3.2: Shown above is the table showing ranges of parameter scores for different severity cases for original data after processing

9. As it is evident from the above values that because of different scales of the parameters, the scores obtained are also of different scores, hence we need to bring them into a common scale.

10. This can be done by minmax scaling of the scores between a range of 0 to 333, as we choose the total scale consisting of all the 3 parameters to be from 0 to 1000.

11. Hence the table for parameter scores after rescaling would look like:

HEALTH STATUS	TEMP_SCORE	HEART RATE_SCORE	BP_SCORE
Normal	0-133.2	0-66.6	0-63.6
Low Severity	133.2-199.8	66.6-166.5	63.6-151.4
Grave	199.8-266.4	166.5-233.1	151.4-242.2
Critical	266.4-333.0	233.1-333.0	242.2-333.0

Table 3.3: Shown above is the table showing ranges of parameter scores for different severity cases for rescaled data after processing

3.4 Selection of method for mapping

The method making use of standard deviation or deviation from the normal values of the health values has been chosen because:

- We need to show that as the severity increases, the distance of the health value must also be more away from the normal values, for which the concept of standard deviation fits the best here.
- Also, as standard deviation is the square root of variance, which is a squared term, therefore, the effect observed will be uniform for points both above and below the normal values i.e as distance of the values from the normal point increases, the value of the mapping increases as per the requirement irrespective of whether the point lies below or above the normal value.

4 Assigning weights to parameters

Now, the next job is to assign weights to the concerned parameters: T-score, HR-score, BP-score according to the following equation.

$$Healthscore = \alpha * T - score + \beta * HR - score + \gamma * BP - score$$

We resort to some geometrical method to solve this problem instead of going for any other existing methods like Machine Learning (to be explained why machine learning won't be applicable in the upcoming sections).

4.1 How to assign weights

We would be using sphere to assign weights to the parameters.
Since weights, α, β, γ are assigned to individual parameters, we would try to isolate T, HR, BP values from the data point obtained from sensors that has temperature, heart rate and blood pressure values by using the concept of sphere.

Method:

1. First choose the parameter to which weight has to be assigned, out of α, β, γ, (i.e T-score, HR-score or BP-score).
2. The chosen parameter's score has divisions from 0 to 333.0 for different severity cases.
3. Construct four cuboidal regions for the four different severity cases with the chosen parameter's score to be in the range of required severity case and other two parameter's scores to be from 0 to 333.0 (complete range).
4. The above point signifies that we are paying attention only to the required parameter since only that particular parameter is being taken into account according to the severity cases, and the other two parameters are being taken from the complete allowed range.
5. This is done for all the three parameters.
6. The next step is to get the data point and rescale it to be in the range of 0 to 333.0 by the method of Min-Max scaling of each feature score.
7. Weights should be assigned in a way such that as score values increase, the concerned weight value also increases.
8. For scores lying in higher severity cases, the corresponding weight value must also be higher than those scores that lie in lower severity cases.
9. For this, first we see how far does the data point lie from the normal value (or the averaged normal) for the parameter that needs to be assigned weight.
10. Normal value or the averaged normal for the concerned parameter would be the centroid of the cuboid with the corresponding parameter lying in the normal range. (i.e the first of all the cuboids in all the three figures)
11. The centroid acts as the normal value for the concerned parameter because we are taking into account all the possibilities of normal scores for the concerned parameter and all the possible values for the other two remaining parameters.
12. Now we obtain the distance of the data point from the normal value.
13. This distance we can consider as the radius of a sphere whose center is the normal value of the concerned parameter.
14. From now on we take into account the concept of sphere because the data point contains all the three values: T, HR, BP. But we need to isolate the parameters to assign weights to them individually. i.e we need to transform the data point into the forms of (T, 0, 0); (HR, 0, 0) and (BP, 0, 0) so that the respective weights are solely assigned to only the concerned parameter.

15. Now, these new points (T, 0, 0); (HR, 0, 0) and (BP, 0, 0) have to be chosen such that they also lie at the same distance as the original data point for invariance of the relation of this new data point from the normal even after processing the data point.

16. Also, we need the original data point and the new corresponding processed data point to be related in some way, which can be achieved only if both the points lie on a sphere whose center is the normal value data point for the parameter.

17. Sphere would be the most appropriate choice because both the points lie at the same distance from the center.

18. For obtaining the corresponding point for the corresponding parameter we need to first check whether the sphere cuts the corresponding parameter axis or not. This can be checked by the equation of sphere.
$$(t - TN)^2 + (h - HN)^2 + (b - BN)^2 = r^2 \text{.................} eqn.(4.1.1)$$
(where TN, HN and BN are the normal data point values respectively for the parameter which has to be assigned weights and t, h and b are the new data point values for the original data point and r is the radius of the sphere.)

19. r can be obtained by finding the euclidean distance between (TN, HN, BP) and the original data point.
$$r^2 = (TN - T)^2 + (HN - H)^2 + (BN - B)^2 \text{.................} eqn.(4.1.2)$$
(where T, H and B are the original data point values of temperature, heart rate and blood pressure respectively.)

20. In the above equation, out of t, h and b, only one of the parameters will be nonzero, the one that has to be assigned weights.

21. Let's take an example, say we want to assign weight to T-score, then according to eqn. (4.1.1), we have h = 0, b = 0 and only t is nonzero.So, the equation becomes
$$(t - TN)^2 + (0 - HN)^2 + (0 - BN)^2 = r^2 \text{.................} eqn.(4.1.3)$$
which appears as
$$(t - TN)^2 + HN^2 + BN^2 = r^2 \text{...................................} eqn.(4.1.4)$$
r can be obtained from equation (4.1.2)

22. Now for obtaining the new parameter value (t value according to equation 4.1.4) we need to first check whether the sphere cuts the corresponding parameter axis or not.
Continuing from equation (4.1.1), we obtain another equation:
$$(t - TN)^2 = r^2 - ((h - HN)^2 + (b - BN)^2) \text{................} eqn.(4.1.5)$$
which becomes:
$$(t - TN) = \pm\sqrt{r^2 - ((h - HN)^2 + (b - BN)^2} \text{............} eqn.(4.1.6)$$

23. From the above equation, t can only exist if the equation inside the radical sign is positive.

- If the equation inside the radical sign is positive, then we can obtain 2 values of t as:

$$t = TN + \sqrt{r^2 - ((h - HN)^2 + (b - BN)^2} \text{.....} eqn.(4.1.7)$$

or

$$t = TN - \sqrt{r^2 - ((h - HN)^2 + (b - BN)^2} \text{.....} eqn.(4.1.8)$$

- In case the equation inside the radical sign is negative, then we approach some other way to obtain the corresponding new parameter value, which shall be discussed later.
- So, according to equations 4.1.4, 4.1.7 and 4.1.8, we get,

$$t = TN + \sqrt{r^2 - (HN^2 + BN^2)} \text{..........................} eqn.(4.1.9)$$

and

$$t = TN - \sqrt{r^2 - (HN^2 + BN^2)} \text{...........................} eqn.(4.1.10)$$

if the equation under radical sign is positive.

24. Now we shall discuss how to get the new parameter value if the equation under radical sign is negative.

25. If the equation under radical sign is negative, it means that the sphere does not intersect the corresponding parameter axis. In such a case, we simply take the new corresponding parameter value as the parameter value of the original data point (as for such points complete isolation is not possible).

26. After obtaining the new parameter value, we have to check whether this value lies in the range of the required parameter severity range i.e if T in the original data point lies in the normal range, then the obtained t value must also lie in the normal range on the Temperature axis and so on.

27. The significance of the above point is that, through the weights directly we should be able to determine the severity level of the parameters, even if we don't look at the corresponding parameter scores, that is possible only if α, β and γ lie in some specified ranges for different severity cases, those ranges can be chosen from the already defines ranges for the parameter scores.

28. Since there are 2 possible values for the new parameter value (say x_1 and x_2), there will be 6 possible cases as follows:

    - When both x_1 and x_2 are greater than the upper limit of the range.
    - When x_1 is greater than the upper limit of the range and x_2 lies in the required range.
    - When x_1 is greater than the upper limit of the range and x_2 is less than the lower limit of the range.
    - When both x_1 and x_2 lie in the required range.
    - x_1 lies in the required range and x_2 is less than the lower limit of the range.

- When both x_1 and x_2 are less than the lower limit of the range.

  (Let x_1 be given by the equation (4.1.7) and x_2 be given by the equation (4.1.8), therefore, x_1 will always be greater than or equal to x_2).

29. Referring to cases from point number 28:
    - Case 1: In this case, put x_1 equal to the upper limit of the required range.
    - case 2: In this cases, put x_1 equal to x_2.
    - Case 3: In this case, put x_1 equal to that value from x_1 or x_2 which lies near to the boundary of the required range.
    - Case 6: In this case, put x_1 equal to the lower limit of the required range.
    - In rest of the case, we would accept the value of x_1 as it is.
30. Then obtain the distance of this new parameter value from the origin (0, 0, 0) which would be same as the new parameter valued data point.
31. Since we need the health score value between 0 and 999 (as all the three parameter scores have been scaled between 0 to 333), hence the values of α, β, γ must lie between 0 to 1.
32. This can be obtained by dividing the obtained new parameter value by 333.0, since the ratio obtained will also lie in different ranges according to the severity cases.
33. Hence, by following the above steps we can obtain the values of α, β, γ.
34. by the following table, we can check whether the weights lie in the correct range or not. (The values have been represented only till 3 decimal places here)

SEVERITY CASE	α	β	γ
Normal	0.0-0.399	0.0-0.199	0.0-0.190
Low Severity	0.399-0.600	0.199-0.5	0.190-0.454
Grave	0.600-0.799	0.5-0.7	0.454-0.727
Critical	0.799-1.0	0.7-1.0	0.727-1.0

Table 4.1: Shown above is the table showing ranges of weights for different severity cases.

4.2 Some interpretations

1. Some of the above points might not be very intuitive, but they have been implemented such they can be interpreted in some specific way.

2. In many cases, it might be possible that even though the required parameter value is higher in the original data point, but in the new parameter valued data point, its value is lower. That can be interpreted as follows:

    - Although the new parameter point gives the absolute value of the concerned parameter free of the effects of the other parameters, but it also reflects the extent of relative effects of the other two parameters.
    - That means, if the new parameter value has high value, then at least one of the other two parameters lie in lower severity cases than that of the concerned parameters, and if the new parameter value is low (according to the range), then at least one of the other two concerned parameters lie in higher severity case than the concerned parameter.

3. If we look at point 29 of the previous section, it would be a very common question that why did we choose only x_1 value over x_2 in the rest of the cases.

4. The reason for that can be interpreted as follows:

    - We know that x_1 is always greater than or equal to x_2, also, the original data point contains all the three parameter values. If we are changing the data point into a form where only one of the parameters is non negative, that means we are removing the relative effect of other two parameters, in that case the effect of that particular parameter should tend to be more evident, hence we choose the higher value of x_1 and x_2 in case 4.

5. For obtaining the weights, we first take the distance of the new parameter from the origin can be explained as follows:

    - Since, the new data point has only effect of one parameter value, for weights, we need to consider a point from which the calculated distance (or value) has the effect of only 1 parameter. Hence, the best possible choice for such a point is the origin, where as the distance is calculated along the required parameter axis , the effect of other two parameters will remain zero (eg. Also, T = 0 in (0, 0, 0) is also point in the normal range of T).

4.3 Possible directions for weight assignment

What we discussed above are not the only methods to accomplish the required task, there can be other and even better solutions to solving the problem. Basically, for solving these kinds of problems, no particular approach exists, since the interpretations of different computer scientists is different, so different people tend to apply different methods for doing so which can also be seen in many different research works. There may be better probabilistic methods, some people can go in the direction of applying Machine Learning techniques also, but for now we have resorted to the solution that seems more feasible and logical. But in future this method can be altered to obtain more accurate results.

4.3.1 Why Machine Learning won't be much applicable here

A possible thought for solving the weight assignment problem could have been the application of KNN (K - Nearest Neighbours) using the weighted Euclidean distance.

- The main motto of Machine Learning is to predict the future values of some parameter based on some fixed or chosen parameters.
- Here in this problem, we already know the severity cases of the individual parameters based on their scores, hence we don't need to predict anything when we already have the results with 100% accuracy.
- Moreover, Machine Learning is used to predict the future values of a particular parameter based on some past values. In this case, even if we try to apply KNN, we would be having values (weights or categorization) for health values of different patients, even if we try to obtain weights for the respective parameter scores, that would be driven by the past values of other patients and not of the concerned patient, which would be a violation to the definition and purpose of Machine Learning and would yield incorrect results.

Hence, it is better to opt for other methods than machine learning techniques.

5 Conclusion

As said before, there may be even better possible solutions for solving the above two problems, but we have resorted to those methods that seem to fit the best according to the problem statement. It is definitely possible that in future different approaches may be used for solving and implementation if the above task.