Cloud Based Predictive Data Analysis Framework for Wearable Device Health Alert System using Semantic Web Services

R.Sethuraman, T.Sasipraba

Abstract*: Rapid Innovation in Digital Technology achieved its frontier with fitness wearable technological devices. The ubiquitous tracking devices currently available in the market only monitor the amount of calories burnt by the user. They do not predict nor encourage users. This paper intends to provide prediction of calories burn based on users' physical activities, and encourage them to achieve more of their fitness goals, with the help of machine learning algorithms and ontology. The proposed framework has two different ontologies used for semantic synchronization. Fitness activities ontology deals with the predicted calories burn value and cloud Telephony ontology provides multi-channel alert services to the end user. FitBit Wearable fitness devices user data are analyzed from the cloud storage via cloud API, is proposed to interact with the user continuously with calories burn value for the improvement of their physical Activities like walking, jogging and step count. A custom model is constructed for predicting the calories burn value using Linear Regression Analysis through Machine Learning Algorithm. The proposed novel framework interacts with semantic web service registry through OWL API with the obtained predicted calories burn value from the prediction models. When compared to the existing system, the proposed framework produces enhanced insights on amount of calories burn to the user based on their activities through cloud telephony alerts like SMS, IVR, Mobile App and Email. The end user improves their activities from the obtained predicted value insights.

Keywords: wearable technological device, ontology, fitness activity, ondemand cloud telephony, web service registry, OWL API.

I. INTRODUCTION

FitBit wearable fitness gadgets are accessible in the market through different pioneers, giving general bits of knowledge to the clients towards their physical exercises [1-3]. Performing Prediction from the observed data and motivating the customers through multiple channels are very challenging. Hence we need novel framework comprises of Prediction model and ontology repository for useful data insights through multiple channels.

In this paper, a prediction model is developed for predicting the amount of calories burn based on the physical activities performed [4]. Two ontologies fitness activity ontology and cloud telephony services ontology are created for holding the predicted calorie values and multiple channels of communication respectively. Upon applying the semantic similarity on both ontologies, the values are stored in ontology service repository. Then the insights are provided to the user through suitable communication channels.

Figure 1 explains the proposed framework for predicting the amount of calories burn and communicating through multiple channels. [5] Prediction is made from the data received by the FitBit fitness wearable devices. The extracted data from cloud API is split in to Training data and Test data. Linear regression methodology is applied to this data and the model is predicted based on the analysis, which is not present in the existing system.Additional services like communicating the predicted values via multiple channels is also value added in this framework. The sum of squares and standard error of estimates are performed to eradicate the missing data input and the prediction of Calories burn Model is computed with this regression analysis. The fitness activity ontology has this predicted data and interacts with the cloud telephony services ontology[6, 7].

Revised Manuscript Received on March 08, 2019.

R.Sethuraman, Research Scholar, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India

T.Sasipraba, Professor, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

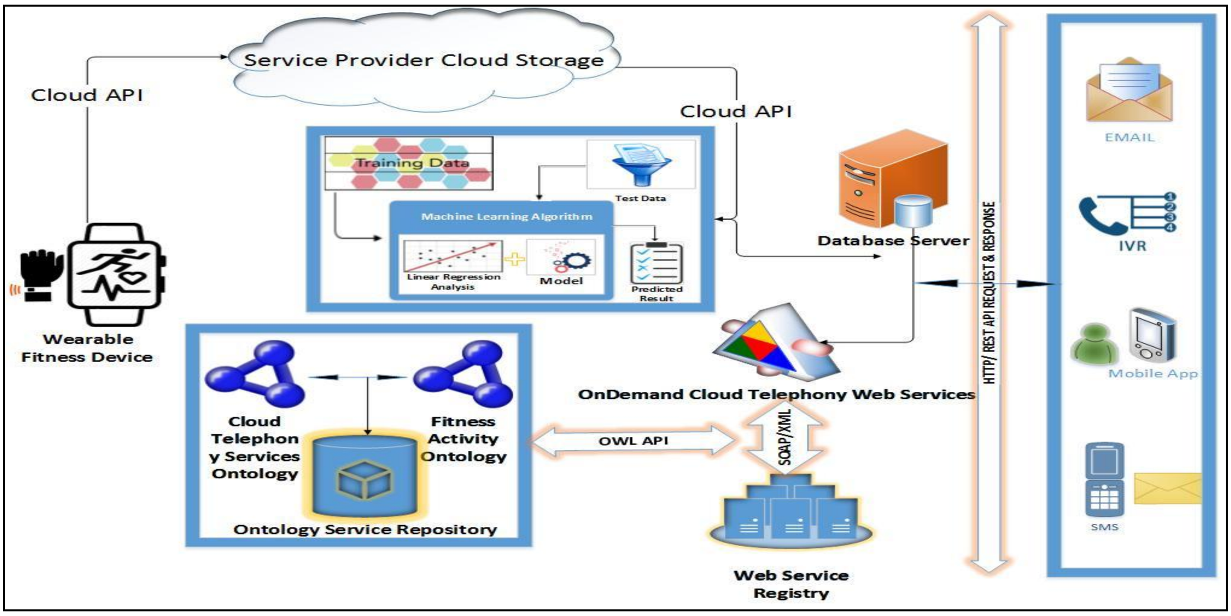

Fig. 1 Overall Architecture of the proposed prediction framework

These predicted analysis were intimated to the user from Ontology service Repository through On Demand cloud telephony web services via SMS, IVR, Mobile Application and Email.The web services registry maintains this communication through OWL API and SOAP/XML protocol[8-10].

The motive behind this research is to provide the predicted calories burn value as insights to the customers continuously from their physical activities through multiple services using telephony ontology [11-13]. Whereas in the existing system prediction is not done from the collected data and telephony ontology services are not in practice. Rather they project the collected data in pictorial representations which lags in the productive information to the customers [3]. In the proposed framework, collected data are visualized to understand the activities without ambiguity and helps in formulating the Regression relation between the Dependent and Independent variables.

The degree of Relations (r) between the dependant variable activities performed duration (in minutes) and the independent variable amount of calories burn are calculated. Obtained "r" value denotes the nature and strength of association known as correlation coefficient. We infer the association type as direct or indirect from the obtained "r" value and then appropriate regression models and graphs are implemented for prediction [4,5].The experimental datasets containing 1631 records were retrieved from zenedo.org. The size of the dataset is 403.8KB [13]. From the dataset Duration on activities performed (in Min) are considered to predict the amount of calories burn. The computation of Regression model from the Regression Analysis is based on mathematical concepts like correlation coefficient, slope determination, linear Regression line and Regression Equation. Correlation

Coefficient factor r is calculated and the Association is confirmed as Direct or Indirect using equation (1).

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}} \quad \ldots\ldots\ldots\ldots \quad (1)$$

The correlation coefficient value (r) for the experimental data using the Pearson's method of correlation is evaluated as 0.76 using Equation(1).Since r(value) ≥.75, implies that there exists a Strong Direct Correlation[6]. From the correlation value, the regression value is calculated. Correlation describes the strength of linear relationship between two variables, where as the Regression describes the average relationship between set of independent variables and dependent variable. The regression analysis does the process of predicting values from the known variables [8]. In the simple linear Regression we use only one Independent variable(X). Linear function describes the relationship between X and Y using the equation (2).

$$\hat{y} = a + bx \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

The changes in X cause the changes in Y also. From this we can form the regression equation, by plotting all the points in XY pairs. After plotting, the line is drawn in such a way that its slope best fits all the plotted points. From that line slope and intercepts were measured for regression equation.

For the experimental data slope value b is calculated using the linear regression equation (3).

$$b = \frac{\sum XY - \frac{\sum X \times \sum Y}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} \quad \ldots\ldots\ldots\ldots (3)$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The obtained slope value denoted by b is 5.27. Now the value of $\bar{x}$ and $\bar{y}$ are calculated using equation (4) for the Formation of linear regression equation and the same is represented in equation (5) for the experimental data.

$$\bar{x} = \frac{\sum x}{n} \text{ And } \bar{y} = \frac{\sum y}{n} \quad \ldots\ldots\ldots\ldots \quad (4)$$

The $\bar{x}$ and $\bar{y}$ values are 71.9 and 429 respectively. The general equation of linear regression for the extracted data using the equation (5) is given in equation (5.1).

$$\hat{y}_{(x)} = \bar{y} + b(x - \bar{x}) \quad \ldots\ldots\ldots\ldots \quad (5)$$

$$\hat{y}_{(x)} = 49.95 + 5.27x \quad \ldots\ldots \quad (5.1)$$

Two estimated values for Y are calculated against their X component and the regression lines are drawn from these plotting points.

$$\hat{y}_{(53)} = 49.95 + 5.27 * 53 = 329.26 \ldots\ldots\ldots \quad (5.1a)$$

$$\hat{y}_{(99)} = 49.95 + 5.27 * 99 = 571.68 \ldots\ldots \quad (5.1b)$$

The values are plotted based on the above linear equation. In reality the linear regression component comes with random error component called Residuals and is given by equation (6).

$$y = \beta_0 + \beta_1 x + \varepsilon \quad \ldots\ldots\ldots\ldots \quad (6)$$

Based on the dependant variable y, the estimated value of y is calculated with regression intercept, slope along with the independent variable from equation (7).

$$\hat{y}_i = b_0 + b_1 x \quad \ldots\ldots\ldots\ldots\ldots \quad (7)$$

Here the values of b_0 and b_1 are obtained as the sum of squared differences from the plots. b_0 provides the estimated average value of y, when x is 0. b_1 provides the estimated change in the average value of y as a result of unit change in x. Based on the factor which behaves unpredictably a random component (ε) is added and the equation (8) is obtained.

$$y = \beta_0 + \beta_1 x + \varepsilon \quad \ldots\ldots\ldots \quad (8)$$

The plotted points are compared against the mean line. Better fit the line to the data, if the sum of squared differences are small. The slope estimation and the least square line estimation is done using equation(9) and (10).

$$b_1 = \frac{SS_{xy}}{SS_{xx}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

$$b_0 = \bar{y} - b1\bar{x} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (10)$$

Table. 1 A comparison table with other Fitness Tracker's dashboard report services and communication

The above values are estimated Coefficients. Simple linear regression measures between the dependent variable(Y) and independent variable (X). The obtained simple linear regression line is computed using Scatter plot and regression analysis. The regression line in the regression model is obtained using least square method. All the methods used to assess the model are based on Sum Of Squares for errors(SSE), which provides the sum of differences between the points and the regression line. Also it measures how well the line that fits the data and is calculated using equation(11).

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (11)$$

$$s_\varepsilon = \sqrt{\frac{SSE}{n-2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (12)$$

The slope which is not equal to zero implies different input(x) produces different output(y) in linear relationship. The strength of linear relationship is calculated using the coefficient of determination which is denoted by r^2.

$$R^2 = \frac{SSR}{\sum (y_i - \bar{y})^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (13)$$

R^2 takes values between zero and one. zero denotes no linear relationship exists between X and Y where one denotes the exact match between the line and data points. The obtained value of R^2 for the experimental data is 0.63.For the two data points (x_1,y_1) and (x_2,y_2) of any sample, the variation in Y is obtained as the sum of variation explained by the regression line and the errors maintained as unexplained variation. In case of errors being very small, implies they are close to mean error and its estimator of σ_ε is given by s_ε.

Variation in Y = SSR + SSE …………………. (14)

In our model, the prediction of Calories burn is calculated from the duration data by the regression equation (15). The regression equation is

y = 5.2719x + 49.952 ………… (15)

Here the Intercept value 49.952 and the b_0 value 5.2719 along with the parameter x predicts the calories burn for given input. SSR provides the variation that exists between the parameters X and Y in Linear relationship and SSE (Residuals) deals with variation in other factors besides the

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Fitness Tracker Wearable Devices Dashboard Services		Fitbit Flex	Jawbone Up24	Withings Pulse	Misfit Shine	Xiaomi Mi Band 2	Proposed Framework
Time Series Analysis Data Driven Daily Report	Step Counting	Yes	Yes	Yes	Yes	Yes	Yes
	Distance (km)	Yes	Yes	Yes	Yes	Yes	Yes
	Floor	Yes	No	No	No	No	Yes
	Calories Burnt	Yes	Yes	Yes	Yes	Yes	Yes
Report Type		Bar graph	Bar graph and number	Bar graph	Number	Bar graph	Bar graph and Number
Goal Based Prediction Reports on Calories Burn		No	No	No	No	No	Yes
Modes of Alert Services	SMS	No	No	No	No	No	Yes
	EMAIL	No	No	No	No	No	Yes
	VOICE(IVR)	No	No	No	No	No	Yes
	Web/ Mobile Apps	Yes	Yes	Yes	Yes	Yes	Yes

Linear relations. The residual plot model is generated for the experimental data used in this paper. The obtained predicted results are plotted in Fig 2 against time series duration and calories burn values. Determined values are plotted using regression equation Y for multiple inputs. The residual value for the predicted calories burn is calculated using Standard Residuals. Final Residual Model is plotted against the time series duration (in mts). Fig 3 shows the plotted values are close to each other implying the predicted output shows less deviation from the standard line in the proposed framework. The residual outcome of predicted value is shown in Table 2.

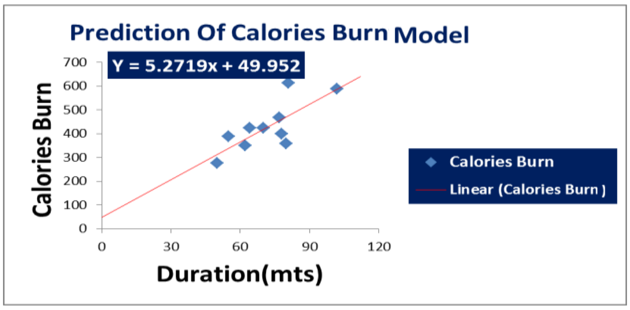

Fig. 2 Prediction of Calories Burn Model

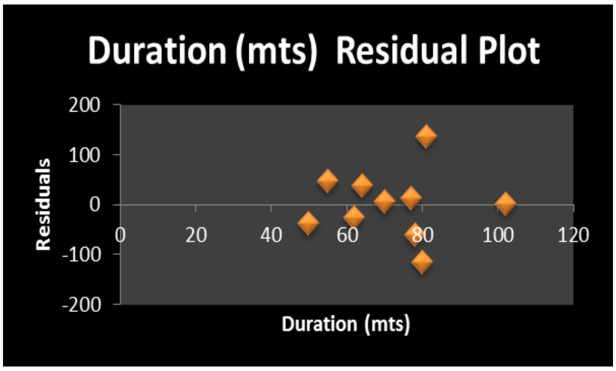

Fig. 3 Residual Plot Model for Experimental Data

Table. 2 Residuals outcome of Predicted Calories Burn

Observation	Predicted Calories Burn	Residuals	Standard Residuals
1	376.808442	-26.8084	-0.39459
2	461.1584347	-61.1584	-0.90018
3	418.9834384	6.016562	0.088556
4	455.8865602	13.11344	0.193014
5	313.5459475	-38.5459	-0.56735
6	339.9053202	48.09468	0.707894
7	587.6834237	0.316576	0.00466
8	476.9740583	136.0259	2.002133
9	471.7021838	-114.702	-1.68827
10	387.3521911	37.64781	0.554129

The impact of the proposed framework is towards customer wellness. The efficient usage of fitness wearable devices without any anomalies produces accurate prediction on calories burn value using proposed framework. In case of irregular usage of fitness tracker, the proposed prediction will change and will not be so helpful in motivating the customers.

The proposed framework is analyzed with existing system in Table 1. The proposed framework provides the goal based prediction reports on calories burn values, where as the existing system provides the data driven daily report without any prediction. The existing system provides the reports as dashboards available in their web pages and they don't support multiple modes of communications. The proposed framework provides four modes of alert services like SMS, IVR web/Mobile Apps and Email. Also this proposed work can be extended to provide the services as chatbot application using Artificial Intelligence there by increasing the motivation and satisfaction of the customers towards their Physical wellness.

ACKNOWLEDGEMENT

We wish to acknowledge the Department of Science and

Technology,India and School of Computing, Sathyabama Institute of Science and Technology,Chennai for providing the facilities to do the research under the DST-FIST Grant Project No.SR/FST/ETI-364/2014.

REFERENCES

1. Judit Takacs, Courtney L. Pollock, Jerrad R. Guenther, Mohammadreza Bahar, Christopher Napier, Michael A. Hunt "Validation of the Fitbit One activity monitor device during treadmill walking." J Sci Med Sp.DOI:http://dx.doi.org/10.1016/j.jsams.2013.10.241
2. Keith M. Diaz,1 David J. Krupka,1 Melinda J Chang,1 James Peacock,1 Yao Ma,2 Jeff Goldsmith,2 Joseph E. Schwartz, and Karina W. Davidson An accurate and reliable device for wireless physical activity tracking. Int J Cardiol. DOI : 10.1016/j.ijcard.2015.03.038
3. Lisa A. Cadmus-Bertram, Bess H. Marcus, Ruth E. Patterson, Barbara A. Parker, Brittany L. Morey Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women Am J Prev Med DOI: 10.1016/j.amepre.2015.01.020
4. Jeffer Eidi Sasaki , Amanda Hickey , Marianna Mavilia , Jacquelynne Tedesco , Dinesh John, Sarah Kozey Keadle , Patty S. Freedson Validation of the Fitbit Wireless Activity Tracker for Prediction of Energy Expenditure J phys. Activity Health DOI: http://dx.doi.org/10.1123/jpah.2012-0495
5. Slootmaker SM, Schuit AJ, Chinapaw MJ, Seidell JC, van Mechelen W. Disagreement in physical activity assessed by accelerometer and selfreport in subgroups of age, gender, education and weight status. Int J Behav Nutr Phys Act. DOI: 10.1186/1479-5868-6-17
6. Comparison of FitBit® Ultra to ActiGraph™ GT1M for Assessment of Physical Activity in Young Adults During Treadmill Walking R.J. Gusmer, T.A. Bosch, A.N. Watkins, J.D. Ostrem, D.R. Dengel Open Sp. Med. J DOI: 10.2174/1874387001408010011
7. M. Shamim Hossain and Ghulam Muhammad Cloud-Based Collaborative Media Service Framework for HealthCare Int. J Distrib. Sens. Netw. DOI: http://dx.doi.org/10.1155/2014/858712
8. Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining (2015 5th Edition) Introduction to Linear Regression Analysis.
9. Vinh Bui, Weiping Zhu, Antonio Pescape Long Horizon End-to-End Delay Forecasts: A Multi-Step-Ahead Hybrid Approach. 2007 12th IEEE Symposium on Computers and Communications.DOI: 10.1109/ISCC.2007.4381513
10. K. Salamatian and S. Vat On. Hidden markov modeling for network communication channels. In Proc. ACM SIGMETRICS 2001, pages 92-101, USA
11. Kanitthika Kaewkannate and Soochan KimEmail A comparison of wearable fitness devices BMC Public HealthBMC series – open, inclusive and trusted201616:433.https://doi.org/10.1186/s12889-016-3059-0
12. Harish, P., Subhashini, R., Priya, K Intruder detection by extracting semantic content from surveillance videos 2014 Proceeding of the IEEE International Conference on Green Computing, Communication and Electrical Engineering, ICGCCEE 2014
13. Fitbit Fitness Wearable Device Experimental Dataset from zenedo.org https://zenodo.org/record/14996#.WfxTJFuCzZ4.