Design of Sports Selection Architecture using Recursive Neural Network and K-Means Clustering Algorithm

Dr. Mahalakshmi[1], Ms. K. Vaishnavi[2]
[1]*Principal i/c, R.B. Gothi Jain College for Women, Chennai*
[2]*Assistant Professor, Department of Information System Management, Mahalashmi Women's College of Arts and Science, Chennai*

Abstract **- Sports are essential part for healthy life for all human being, as it keeps the person fit and agility. Sports also helps person for healthy mind. Practice is very essential for the person in sports. If human body or health is not suitable for the particular sport, then the person cannot lead the sport cycle to the successful part. Selecting the sport is very essential for the success. This paper proposed the system for selecting the correct person to the sports which the person is matched in the boundary of practice and healthy life according to the several perceptions of calculation-based speed, weight, height and heart rate. This implementation is developed using MATLAB software. This system was implemented with FitRec dataset with the base of Json file read with pyspark. For the selection process, this system was designed by Recursive Neural Network (RNN) and for grouping the collected data is completed by K-means clustering algorithm. Result is produced Exploratory Data Analysis (EDA). This system provides the 89% of accuracy rate among the sample size N chosen by the user.**

Index Terms **- FitRec dataset, Jsion, Pyspark, RNN, EDA**

1.INTRODUCTION

Sport pertains to the various forms of the competitive in the physical activity that aims to pertain, necessity or improvement fitness that gives happiness to the participants and, but perceptions. Sports can, through different zone of the health sector. Some sports allow the result as draw that means there is no one person winner; provides draw breaking methods to give the success to one winner and to one loser. This paper is explained with complete system of analysis and selection architecture with the design of implementation. Therefore, it is best to use a sports coach to design a special program and see the appropriate sports for the person in proportion to the individual's capabilities and health so that the latter does not suffer any possible injuries in the future or negatively affect the body health. And because we saw that every person who practices sports needs to know what sport is appropriate for his body and health. [1]

2.METHODS

2.1 DATA SOURCES:

FitRec DATASETS:

These datasets contains users sports record on the Endomondo. Data includes many source files of ordered sensors data's like pulse rate, speed, GPS moreover as sports types, user gender and atmospheric phenomenon (i.e. temperature, humidity). After achieving the filters datasets (i.e. endomondoHR_proper.json), the user can able to process the datasets by normalization methods, the measurements data's into Z-Scores. During test, we further strain those users with but 10 workouts. Supports the processing datasets above the normalized version, we use interpolation to get a dataset with same sampling intervals gap of 10s during each and every workout.

Types:

- Measurement data for the physical test like pulse, rate speed in time, distance covered, etc.
- Depending data for the power test like latitude, longitude, gender, etc.

3.EXPERIMENTAL SETUP

Measurement data and depending data are collected by using FitRec data set, and these are data are grouped by using K-means clustering based on the grouping as heart rate vs. altitude, heart rate vs. speed, etc. These grouped data are uploaded to the MATLAB software

for the selection procedure. Selection process is computed by the Recursive Neural Network algorithm.

3.1K-MEANS CLUSTERING ALGORITHM:

RNN is used for selecting the appropriate person for the correct sports of FitRec dataset with the calculation based. A RNN is also a method of deep neural network. So, with this, user can expect and acquire a structured prediction by applying the identical number of sets of weights on structured inputs. The above algorithm provides the functional data sets value for selection procedure.

1. Specify number of clusters K.
2. Initialize centroids by first shuffling the dataset so randomly selecting K data points for the centroids without replacement.
3. Keep iterating until there's no change to the centroid.

3.1 Computing the sum of the squared distance between data points and every one centroid.

3.2 Assigning each and every information to the closest clusters.

3.3 Computing the centroids for the clustering datasets by taking the typical of the all data points that belong to every clusters.

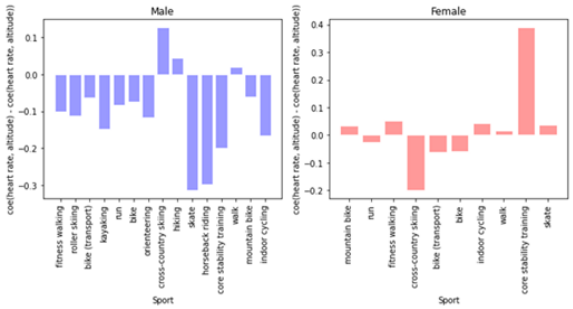

Fig 1 – Pearson Coefficient Difference Between (Heart Rate, Altitude) vs. (Heart Rate, Speed)

Figure 1 is explained about the clustering structure based on the male and female on the perception of their heart rate, altitude and speed. This system proposed the sports related to the gender wise classification. On the gender wise selection, male can be included in the list of sports as fitness walking, roller skilling, bike, kayaking, run, bike race, orienteering, cross-country skilling, skate, hiking, horseback riding, core stability training, walk, mountain bike and indoor cycling. Out of these sports, the algorithm proposed that the male can be in the sport cross-country skilling. The selection based on the person's height, weight, speed, state of mind and heart rate. This system proposed that the male can play more good in cross-country skilling suggested by the Pearson's coefficient calculation resulted as 0.1. Next section, female can be included in the list of sports as mountain bike, run, fitness walk, cross-country skilling, bike race, bike, indoor cycling, core stability training, skate and walk. The selection based on the person's height, weight, speed, state of mind and heart rate. This system proposed that the female can play better in core stability training suggested by the Pearson's coefficient calculation resulted as 0.39.

COMPARISON:

Based on the comparison, bike race – male can perform -0.04 and female can perform -0.02 and mountain bike – male can perform -0.07 and female can perform 0.02 and skate – male can perform -0.3 and female can perform 0.03 and fitness walking – male can perform -0.1 and female can perform 0.05 and indoor cycling – male can perform -0.13 and female can perform 0.02

4.RECURSIVE NEURAL NETWORK ALGORITHM (RNN)

RNN is employed for choosing the acceptable person for the right sports of FitRec dataset with the calculation based. A RNN may be a style of deep neural network. So, with this, user can expect and acquire a structured prediction by applying the identical number of sets of weights on structured inputs.

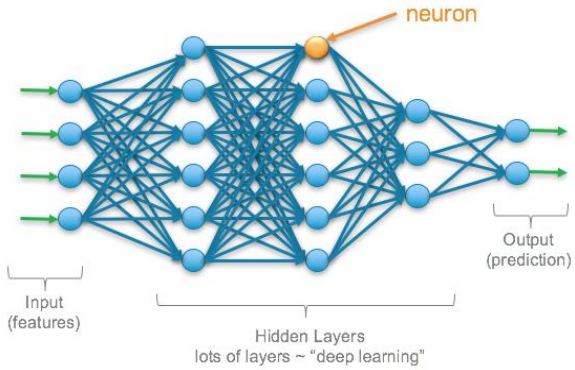

Fig 2 – An Architecture of RNN

These RNN's are even costlier in any respect computational learning stages and phases.

Step 1: Create a variable to carry the input images as N

Step 2: Set a filter matrix (randomly generate it). With the filter, images are convolved.

$$Z1 = N * f$$

Step 3: On the result, use the Sigmoid activation function. A = sigmoid(Z1)

Step 4: Transform the values with a linear transform. Z2 = WT.A + b

Step 5: Run the info through the Sigmoid function. the ultimate product will appear as if this.

$$O = \text{sigmoid}(Z2)$$

There is also differing types of branch that consists of the nodes, but branch nodes of the identical type have tied weights.

	id	sport	userId	gender	mark	perAvgHeatRate	perAverageAltitude	perAverageSpeed	enegy level	Time taken in mins	Humidity	temperature in F
0	396826535	bike	10921915	male	1	152.650	43.0712	26.162158	50	6	37	98.795355
1	392337038	bike	10921915	male	2	147.710	35.7248	27.218369	40	2	46	98.400333
2	389643739	bike	10921915	male	1	140.554	62.4256	26.050774	68	5	41	98.656770
3	386729739	bike	10921915	male	1	147.020	32.6104	26.877838	42	8	40	98.978446
4	381185226	bike (transport)	10921915	male	1	166.084	13.3640	29.592281	61	10	44	97.093543

Table 1 – Cleaned and Selected Feature Data

Table 1 states the selection of bike race sports person by the perception based on gender, heart rate, altitude, speed, energy level, time taken, humidity and temperature. This table picks the male gender person as the sports are based on bike race. Each person's average heart rate is approximately from 152 to 167 and average altitude is based on the person's bike. This system is monitored the bike speed of each person and the time taken for completing the task by monitoring each person's body temperature. Average speed is from 26 km/min to 29 km/min and the time taken is between 6 min to 10 min.

5IMPLEMENTING METHODOLOGY

5.1 JSION FILE READ IN PYSPARK:

Spark SQL can automatically infer the schema of a JSON dataset and load it as a Dataset[Row]. This conversion can be done using SparkSession.read.json() on either a Dataset[String], or a JSON file. Note that the file that is offered as a json file is not a typical JSON file. Each line must contain a separate, self-contained valid JSON object. For a regular multi-line JSON file, set the multiline option to true. SPARK SQL provides spark.read.json("path") to read a single line and multiline (multiple lines) JSON file. Reading the json file is actually pretty straightforward, first, create an SQLContext from the spark context.

5.2 EXPLORATORY DATA ANALYSIS:

Table 1 shows the statistics of the dataset in respect of the total number of sports, workouts, genders, speed, and heart rate. As shown in the table, the majority of participants is males, while the minority is females. There are also 17,450 unknown gender participants. The speed mean ±SD = (47.783 ± 16.867 MPH) and the speed range is (min = 0.0, max = 148.164 MPH). The speed standard deviation is small, meaning that speed values of the participants are centralized around the mean. The min value of speed rate illustrates that there are sports that do not require fast movement. The heart rate mean ±SD = (266.14 ± 36.925 BPM) and the heart beat rate range is (min = 0.0, max = 468.0 BPM). The large value of the SD of the heart rate indicates that the heart rates values are scattered, meaning there are a notable difference between the users' heart rates. This could be due to their gender, weight, diseases, etc. These differences confirms that, the heart rates do not follow a specifies the pattern and they are varying from participant to another according to his/her personalized health. Therefore, the heart rate issues should be considered during the workouts for each participant alone, this is the value of our proposed recommendation system. The females heart rates are usually higher than the males specially when they skate or play gymnastics. At the same time, the standards show that males are much faster than the females. However, their heart rates are negatively correlated with the speed due to athletic heart syndrome. It is a phenomenon that explains the natural changes that take place in the hearts of people

participating in vigorous athletic training. Eventually, our recommendation model will address these two issues, it recommends participants the suitable sports. Plus, the participant's heartbeat will be observed during the workout to notify him if his heartbeat reach dangers threshold to take suitable action, such as slowing done or changing his path if there are altitudes.

Classification Report :

	precision	recall	f1-score	support
0	0.52	0.96	0.67	4661
1	0.00	0.00	0.00	284
2	0.00	0.00	0.00	5
3	0.00	0.00	0.00	7
4	0.00	0.00	0.00	8
5	0.00	0.00	0.00	1
7	0.00	0.00	0.00	201
8	0.00	0.00	0.00	1
9	0.00	0.00	0.00	304
10	0.00	0.00	0.00	1
11	0.00	0.00	0.00	8
12	0.00	0.00	0.00	3512
13	0.00	0.00	0.00	2
14	0.00	0.00	0.00	19
accuracy			0.50	9014
macro avg	0.04	0.07	0.05	9014
weighted avg	0.27	0.50	0.35	9014

Table 2 – Classification Report

	precision	recall	f1-score	support
0	0.86	0.98	0.91	4661
1	0.00	0.00	0.00	284
2	0.00	0.00	0.00	5
3	0.00	0.00	0.00	7
4	0.00	0.00	0.00	8
5	0.00	0.00	0.00	1
7	0.00	0.00	0.00	201
8	0.00	0.00	0.00	1
9	0.00	0.00	0.00	304
10	0.00	0.00	0.00	1
11	0.00	0.00	0.00	8
12	0.94	0.99	0.97	3512
13	0.00	0.00	0.00	2
14	0.00	0.00	0.00	19
accuracy			0.89	9014
macro avg	0.13	0.14	0.13	9014
weighted avg	0.81	0.89	0.85	9014

Table 3 – RNN Performance and Accuracy

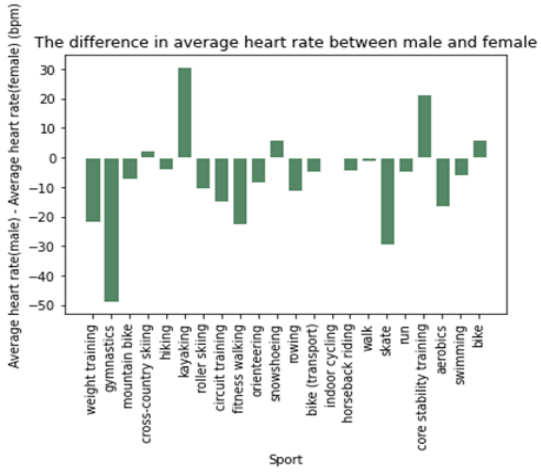

Fig 3 – EDA Analysis Report – Heart Rate Difference Between Male and Female

6.RESULTS AND DISCUSSION

Based on the comparison, bike race – male can perform -0.04 and female can perform -0.02 and mountain bike – male can perform -0.07 and female can perform 0.02 and skate – male can perform -0.3 and female can perform 0.03 and fitness walking – male can perform -0.1 and female can perform 0.05 and indoor cycling – male can perform -0.13 and female can perform 0.02.

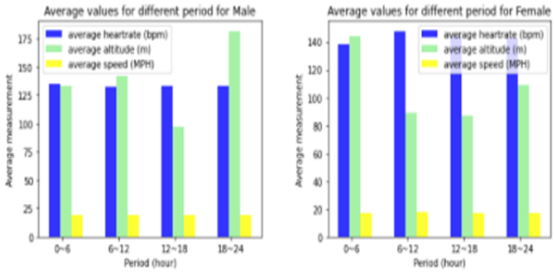

Fig 3 – Average Value of Heart Rate, Altitude and Speed

Figure 3 discuss about the total value of the different period of speed, heart rate and altitude for both male and female. On the perception, male can perform with their altitude more in the period of 18 – 24 hours and female can eventually play all the time with normal heart rate.

7.CONCLUSION

Sport pertains to any forms of the competition physical activities that aim to using, maintaining or improving the physical strength and the skills while provides joyful to the participants. This paper concludes that the system designed is for selecting the appropriate person to the matched sports under the category that is matched in the boundary of practice and healthy life according to the several perceptions of calculation based speed, weight, height and heart rate. This system was implemented with FitRec dataset and Jsion file read with pyspark. For the selection process, this system was designed by Recursive Neural Network (RNN) and for grouping the collected data is completed by K-means clustering algorithm. Result is produced Exploratory Data Analysis (EDA). This system can result 89% of accuracy precision level is produced by the datasets.

REFERENCE

[1] A. K. Chowdhury, A. Farseev, P. R. Chakraborty, D. Tjondronegoro, and V. Chandran, "Automatic classification of physical exercises from wearable sensors using small dataset from non-laboratory settings," in 2017 IEEE Life Sciences Conference (LSC). IEEE, 2017, pp. 111–114.
[2] https://doi.org/10.1016/j.neunet.2016.12.008
[3] https://elibrary.kubg.edu.ua/id/eprint/28282
[4] C. Zhang, H. Liang, and K. Wang, "Trip recommendation meets real-world constraints: Poi availability, diversity, and traveling time uncertainty," ACM Transactions on Information Systems (TOIS), vol. 35, no. 1, pp. 1–28, 2016.
[5] B. Loepp and J. Ziegler, "Recommending running routes: framework and demonstrator," in Workshop on Recommendation in Complex Scenarios, 2018.
[6] K. Bafna and D. Toshniwal, "Feature based summarization of customers' reviews of online products," Procedia Computer Science, vol. 22, pp. 142–151, 2013.
[7] K. Hammond and A. S. Varde, "Cloud based predictive analytics: text classification, recommender systems and decision support," in 2013 IEEE 13th International Conference on Data Mining Workshops. IEEE, 2013, pp. 607–612.
[8] J. P. Verma, B. Patel, and A. Patel, "Web mining: opinion and feedback analysis for educational institutions," International Journal of Computer Applications, vol. 84, no. 6, 2013.
[9] H. Liu, J. He, T. Wang, W. Song, and X. Du, "Combining user preferences and user opinions for accurate recommendation," Electronic Commerce Research and Applications, vol. 12, no. 1, pp. 14–23, 2013.
[10] S. G. Esparza, M. P. O'Mahony, and B. Smyth, "Mining the real-time web: a novel approach to product recommendation," Knowledge-Based Systems, vol. 29, pp. 3–11, 2012.
[11] V. Gurusamy, S. Kannan, and K. Nandhini, "The real time big data processing framework: Advantages and limitations," International Journal of Computer Sciences and Engineering, vol. 5, no. 12, pp. 305–312, 2017.
[12] S. Shahrivari, "Beyond batch processing: towards real-time and streaming big data," Computers, vol. 3, no. 4, pp. 117–129, 2014.