Physical Activity Recommendation for Exergame Player Modeling using Machine Learning Approach

Zhao Zhao
Institute of Communication, Culture, Information and Technology
University of Toronto
Toronto, Canada
zhao.zhao@utoronto.ca

Ali Arya
School of Information Technology
Carleton University
Ottawa, Canada
arya@carleton.ca

Rita Orji
Faculty of Computer Science
Dalhousie University
Halifax, Canada
rita.orji@dal.ca

Gerry Chan
School of Information Technology
Carleton University
Ottawa, Canada
gerrychan@cmail.carleton.ca

***Abstract*— Exergames are effective tools to motivate and promote daily physical activities. However, previous studies indicated that many people who start any type of exercise drop out of the program before establishing new habits. Research has shown that personalization is key to effective game-based interventions. Player modeling and recommender systems are used for personalizing contents and services in many applications. In exergames, we believe it is important to continuously recommend personalized and appropriate types of physical activity and contents in order to improve the effectiveness of the game. In this paper, we proposed and validated the design of a personalized physical activity recommender system for exergames based on a study of participant's preferred activities. The proposed approach resulted in more accurate recommendations when comparing to an existing model in predicting users' preference toward physical activity types.**

Keywords—exergame, player modeling, recommender system

I. Introduction

The term exergame can be used to refer to gamification of physical activity (PA) and using games to persuade players to be more physically active. Exergames have rapidly emerged over the past years to promote healthy behaviors and keeping an active lifestyle [1]. Researchers have studied various gameplays and game features to make exercise and PA more engaging and attractive [2][3]. However, most existing work on gamification and persuasive games in health and wellness are limited in that they are typically designed using a "one-size-fits-all" approach, which has been shown to be suboptimal [4]. In particular, one of the main research gaps in this area is how to properly recommend suitable and interesting PA to each player. There are few models that allow such recommendations [14], however their suggested activities are not based on any empirical evidence. Therefore, an assiduous combination of a detailed model with features such as personality types, modeling-based personalization, and recommendation with adaptive gamified elements in the area of exergames is needed.

To address these research gaps, in our previous work [34] we proposed a comprehensive model for gamified fitness recommender systems that used detailed and dynamic player modeling and wearable-based tracking to provide personalized game features and activity recommendations. The results showed the feasibility and effectiveness of using player modelling for recommending PA. However, the parameters for building the player model were selected intuitively based on our limited knowledge of the problem domain. Therefore, parameter selection – the process of selecting a subset of relevant player information that can be used to recommend PA and game features [32] to improve effectiveness – can be optimized in our player model.

As shown in Fig. 1 below, the player model we used in our previous research consists of four sub-models: (1) activity recognition model, (2) general model, (3) exerciser type model, and (4) player type model, which covered parameters of the user's age, gender, weight, height, recognized daily activities, steps, active calories, walking/running distances, calendar events, location, as well as, the user's player type (Hexad model [33]) and exerciser type (8-colors model [14]). From the existing literature, there were some potential parameters that could be added to our system to enhance the accuracy of the model for recommendation, including users' lifestyle information such as sleeping habits, type of occupation, measures of stress, etc., [5].

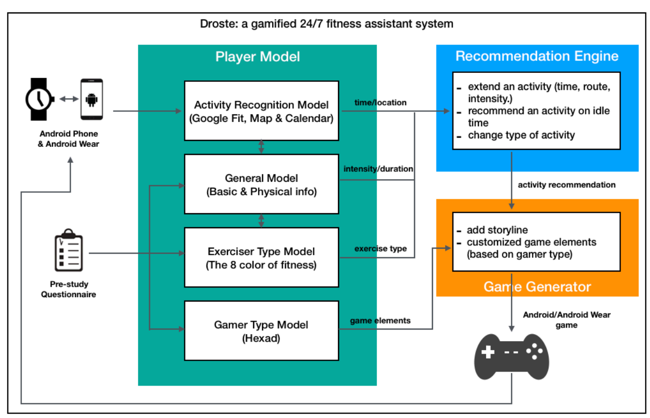

Fig. 1. The exergame player modelling system architecture

On the other hand, the 8-color-of-fitness system was used as a model to suggest activities. Thus, the recommendations of activity types that the system in our previous work generated were mostly based on personality traits. This was due to the fact that there exist no other alternatives and we needed to rely on a fairly acceptable method. In this paper, we report on an alternative design for recommending different types of PA based on questionnaire results. Our contributions are through providing answers to the following research questions:

- What parameters (player information) are necessary in an exergame player model?
- How could the model be used for activity recommendation?
- Which existing models and approaches are effective for recommending specific activities to different users?

Our proposed activity recommendation model is validated within our own general exergame architecture, but it can be implemented into other systems that needs to recommend PA. The remainder of the paper is organized as follows. We provide a brief review of related literature, explain our questionnaire-based methodology and findings, and then describe our new model-based activity recommender design.

II. Related work

A. Recommender Systems

Recommender systems have been applied in several areas, especially in the context of online shopping systems (such as Amazon), as well as music and movie recommender systems (such as Spotify, Netflix, etc.). Typically, the recommendation is generated on the basis of a user's former behavior in a system. Related historical data is stored per user along with his/her account information in order to form a user profile [25].

As for building recommender systems, various methods exist. Burke [26] distinguishes four different categories of recommendation techniques based on knowledge source:

1. *Collaborative*: The system generates recommendations using only information about rating profiles from different users.
2. *Content-based*: The system generates recommendations from two sources: the features associated with products and the ratings that a user has given them.
3. *Demographic*: A demographic recommender provides recommendations based on a demographic profile of the user.
4. *Knowledge-based*: A knowledge-based recommender suggests products based on inferences about a user's needs and preferences.

Hybrid recommender systems are another method and basically combine multiple recommender systems to overcome some of the shortcomings of each.

A recent review of health recommender systems shows that most investigations aim to improve the general well-being of users, such as recommending diets and exercise plans [27], rather than recommending PA. Due to some practical constraints which make it difficult to design a recommender system for exercise and fitness activities (which is discussed in the following section), very limited research has been conducted in this area.

B. Physical Activities Recommendation

When designing a recommender system for exercise and fitness activities, there were some practical constraints which made it difficult to directly use any of the well-established recommender system algorithms. For instance, most types of recommender systems work based on ratings on items. In the field of PA, no ratings exist. One of the reasons that no ratings exist is because of individuals concerning a specific exercise activity is hard to measure.

Therefore, in our search of finding closely related research, we came across a limited amount of literature that discusses the use of recommender systems in sports and exercises. However, the idea of personalization in PA has been studied by many researchers. For example, Guo [17] proposed a system that recognizes different types of exercises and interprets fitness data (motion strength and speed, etc.) to an easy-to-understand exercise review score, which aims to provide a workout performance evaluation and recommendation. Though it achieved 90% accuracy for workout analysis, it focused only on recognizing fitness activities and not personalizing or gamifying them.

In a different study, He et al. [18] introduced a system designed to be context-aware for PA recommendations. It focused on selecting suitable exercises for individualized recommendations. A smartphone application was developed that could generate individualized PA recommendations based on their database of PA. The focus of their work was to recommend different types of activities but does not consider personal details such as proper time, location, and intensity, or offer any gamified elements.

Broekhuizen et al. [19] proposed a framework called PRO-fit, which is another example that employs machine learning and recommendation algorithms to track and identify a user's activity by collecting accelerometer data, synchronizes with the user's calendar, and recommends personalized workout sessions based on the user's and similar users' past activities, their preferences, as well as their physical state and availability. They highlighted that many applications nowadays are more focused on tracking user's activities, but do not provide a recommender system that would help users choose from activities based on their interests and accomplishment of goals. Therefore, they were motivated to design the personalized fitness assistant framework that acts as a motivator and organizer for fitness activities, making it easier for users to create and follow their workout plan and schedule the sessions according to their availability and preference. However, their system was based on prefixed recommendations which do not involve a player model.

Rabbi [20] introduced a smartphone application "MyBehavior" that generates personalized health feedback from PA and food log data. A 14-week study showed improvement in PA and a decrease in food calorie consumption when using the application compared to a control condition. This work was a novel approach that provides personalized suggestions by learning the user's behaviors, which was the closest to our proposed idea. However, MyBehavior was not an exergame. It did not look at the impact of gamification elements or any feature release method. Their approach in generating recommendation was different from player modeling.

C. Exerciser Models

When designing recommendations on PA, personality type plays an important role in determining people's fitness tastes [31]. Some people may prefer swimming laps solo while others enjoy attending a rowdy group-cycling class. These preferences have less to do with people's physical characteristics and are affected more by personalities. Matching PA to personality type has been shown to have real-world relevance [12]. Research suggests that people who engage in personality-appropriate activities will stick with the activities longer, enjoy their workout more and have a better overall fitness experience [13].

Brue [14] created a system, which is a personality centered approach to exercise grounded in the personality type framework popularized by the Myers-Briggs Type Indicator (MBTI) instrument. She took the MBTIs - Introversion (I) or Extraversion (E), Intuition (N) or Sensing (S), Thinking (T) or Feeling (F), and Judging (J) or Perceiving (P) - and reworked them into an easily maneuverable color-coded fitness personality model, which aims to help people discover the best approach to exercise based on personality type. Brue indicated that knowing more about the personality and people's likes and dislikes can make it easier to plan and, more important, be satisfied with exercise, which makes it more likely people will continue to engage in it over the long-term.

The 8-Colors is based on eight preference pairs each corresponding to a color. By understanding Fitness Personality, people gain an understanding of their motivational patterns, preferred interactions, and environments, and can more effectively choose specific forms of PA that are best for them and they will stick with. For instance, some people are traditional and conservative in their approach to exercise while others seek variety and cutting-edge information. Some enjoy being solitude and consider exercise a moving meditation while others prefer a fast-paced class with energetic music.

In the 8-color system, eight types of fitness colors were introduced in which reds are quick responders in the physical world. Whites like to plan and are visionary types who like calm spaces and don't like to be rushed. Greens are nature lovers who like to be outdoors. Golds are traditional, conservative types. Saffrons value individual expression. Blues are safety oriented and are good at creating their own space in a gym. Purples are routine-oriented. Silvers like exercise to be disguised as fun, or at least a fun way to meet others. Overall, people with a different fitness color were motivated by different factors, as well as were linked to some suggested activities.

Therefore, the 8 Colors of Fitness model was initially used in our previously proposed system for suggesting different types of activities. This model is one of the few that use personality type as the basis for activity recommendation and is suggested by other researchers and practitioners [15][16].

However, because the 8-color-of-fitness system was used initially as a model to suggest activity types, the recommendations of activity types that our system used to generate were mostly based on personalities. This was due to the fact that the research group did not find any other alternatives and needed to rely on a fairly acceptable method. Therefore, the method of mapping player model to the right type of activities for recommendation can be further investigated.

III. METHODOLOGY

In this section, we describe the method used to evaluate exercise habits and the design of the recommender system. Since existing models, such as the 8-color model, have not been empirically tested, we designed a questionnaire including questions regarding people's exercise preference and their basic/lifestyle information, etc. We used the questionnaire data to train a binary predictive model to predict whether a user would like a new type of exercise or not. We aimed to use this predictive model to replace the exerciser type model (the 8-colors) in our exergame player modeling system, in order to generate more feasible PA recommendations to prolong user engagement. In the following sections we talk about the questionnaire, data analysis, and the recommender system design, respectively.

A. Questionnaire

Firstly, we conducted a questionnaire with 178 participants to gather data regarding people's exercise preferences. Potential parameters (such as their getup/bed time, purpose of exercise, favorite types of video game, etc.) were selected based on our literature review and the goal was to find the relationship between various personal parameters and preferred personal activity, In addition to regular demographic and health information like age, gender, weight, height, etc., lifestyle information, as well as other potential parameters were also collected from our participants, for example, diet constraints, sleeping patterns, activity tracker owned, etc. Our questionnaire consisted of 30 items. A brief list of questionnaire items is shown in Table I.

TABLE I. LIST OF QUESTIONNAIRE ITEMS

Basic info	**Lifestyle info**	**Objective info**	**Other parameters**
Age	Get-up time	Exercise importance level	No. of children
Gender	Bedtime	Purpose of exercise	Pets owned
Height	Nutritional supplement	Amount of exercise	Favorite exercise music
Weight	Exercise frequency	Enjoyment level of different exercises (details of different types of exercises were included in Appendix A)	Cell phone system
Occupation	Exercise partner		Wearable activity trackers owned
Medical	Exercise trainers		Favorite type of soft drink
Limitation for exercise	Gym mate		Favorite types of video game
Residence	Vegetarian		Favorite sport brand
	Daily commute methods		8-colors of fitness type
	Typical daily routine		

The detail of the questionnaire items is shown in Appendix A. The questionnaire was approved by the institutional research ethics board and conducted online from March 10, 2019, until April 30, 2019. An invitation to participate in the questionnaire was sent by e-mail from our research group to our colleagues and friends, as well as our previous research participants. We collected a total of 178 responses.

B. *Questionnaire Data Processing*

For classification purposes, we preprocessed the data by labeling the user's preference for each type of exercise. For each exercise, we defined: -1: negative experience (not at all enjoyed), 0: never tried (N/A), and 1: positive experience (including slightly enjoyed, very enjoyed and extremely enjoyed). We trained a binary prediction model with only non-zero data and using the Support Vector Machine (SVM) method [28] for each individual PA. We excluded those zero values because those were the PAs that users have no experience with. Therefore, we considered them as invalid values for training the predictive model. We chose the SVM method, which is a supervised learning algorithm, because it is effective in high dimensional spaces and works especially well on small datasets. There are many other classification algorithms that can be used in this prediction model for comparison and optimization. The choice of the learning and prediction method was not one of our research questions and as such, we chose only one possible option to develop our prototype. Improving this part of the system is one of the future directions of the research. We divided our data into 80% training data and 20% testing data. For those PA with testing accuracy (TP + TN)/total sample size >= 70%, we conducted Permutation Feature Importance (PFI) experiment [29] to compute feature importance scores, as explained in the following paragraph. PFI is an algorithm that computes importance scores for each of the feature variables of a dataset. The importance measures are determined by computing the sensitivity of a model to random permutations of feature values. In other words, an importance score quantifies the contribution of a certain feature to the predictive performance of a model in terms of how much a chosen evaluation metric deviates after permuting the values of that feature [30].

Because our dataset is relatively small, we chose the accuracy bigger than 70% to be an acceptable prediction. Note that directories of different types of sports and exercises [21] are used in this research as potential PA for the recommendation. Questions 30 and 31 in the supplementary document (Appendix A) lists the PA used in this questionnaire.

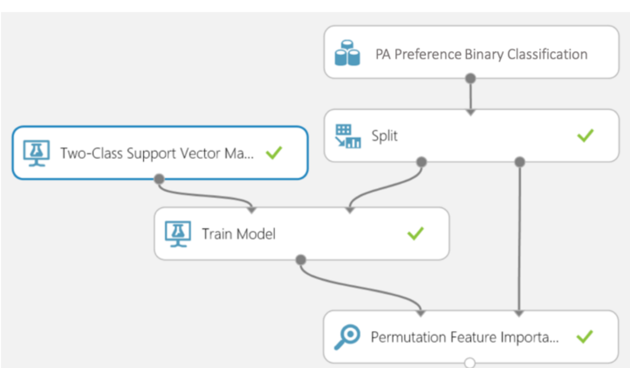

Fig. 2. Overall Workflow of the Feature Importance Experiment in Azure (screenshot)

The experiment of training of the predictive model, and evaluating which features were more important was done in Microsoft Azure Machine Learning Studio. For each individual PA, we set up the following experiment (see Figure 2 above):

1. Add the PA Preference Binary Classification to the experiment.
2. Add a Split module to create a training and test datasets.
3. Add a Two-Class Support Vector Machine module to initialize the SVM classifier.
4. Add a Train Model module to train the classifier, and connect the SVM module to the left input port and the training dataset to the right input port.
5. Add a Permutation Feature Importance module and connect the trained model and the test dataset to the left and right input ports respectively. Set the Metric for measuring performance property to Classification - Accuracy.

The above 5-step experiment was done in Azure, and we visualized the output port of the PFI module. Fig. 3 shows an example of the list of features sorted in descending order of their permutation importance scores for playing basketball.

We conducted the experiment for each individual PA, then computed the average importance score for each feature. Note that for different PA, the corresponding rank of the important features varied, but the overall goal of this experiment is to find the important features for building a player model in general, so we used the average score here. The top 10 features with relatively higher average feature importance score were selected as the features for the final classification model.

view as

Feature	Score
Age	0.038462
Gender	0.034837
Partner	0.029738
Video game	0.027364
Getup time	0.021753
Occupation	0.018464
Residence	0.013746
Wearable Trackers	0.009364
Importance	0.008263
Children	0.006234
Trainer	0.003347
Pets	0.001832

Fig. 3. Example of Feature Permutation Importance Scores (screenshot for Basketball)

C. *Recommender System Design*

The overall recommendation process of the system is outlined in Fig. 4 below. For those PA with accuracy >=70%, we directly used the prediction model to decide whether to recommend a certain type of PA to a new user or not (see the left branch).

For other types of PA, we employed a collaborative filtering method with K-means clustering [22]. Unsupervised cluster analysis is performed using the questionnaire data (for the only PA with accuracy <70%) (see the right branch). The K value of 4 was generated with the elbow method (to define clusters such that the total within-cluster sum of square (WSS) is minimized [23]. The K-value was calculated using fviz_nbclust() function in factoextra R package [24]. For each cluster, we found their top five (out of 13) popular PA.

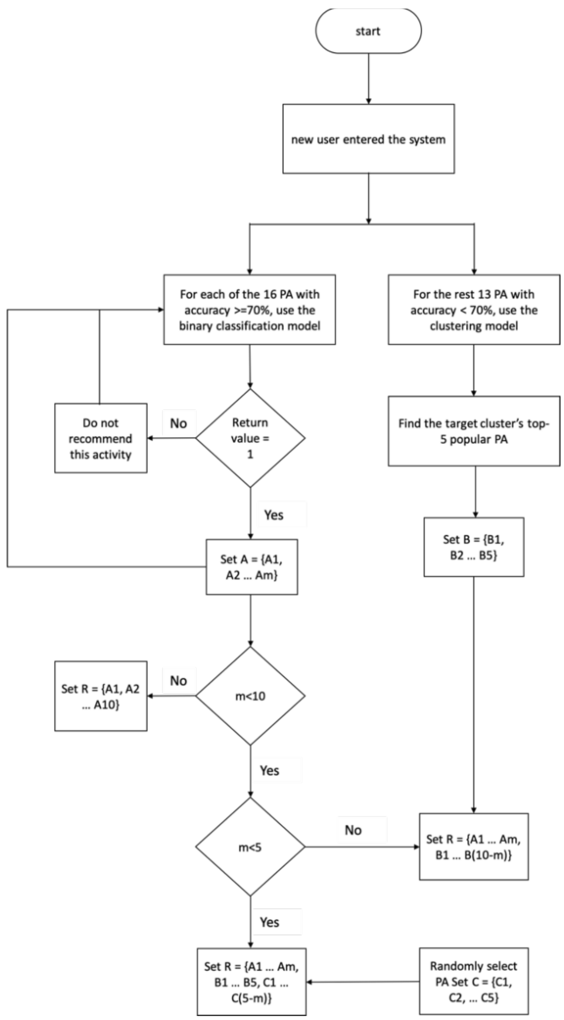

Fig. 4. Overall Recommendation Flowchart

We aimed to recommend ten PA to each individual user. As shown in Fig. 4, set R represents the result of the final ten PA for recommendation, Set A represented those PA with positive results from the binary classification model, Set B represented the top five popular PA of the cluster which the target user belongs to, and Set C included five randomly selected PA (excluded those PA in Set B), as an alternate set in case the number of item in Set A is less than five. At this stage, the PA in Set R was recommended to our users in a randomized order, because we did not consider the weight of different PA in the prediction model.

Furthermore, the PA prediction model in this proposed work kept updating itself. Thus, when any new data was entered into the system, it will be added to the dataset to re-train the prediction model. For instance, if a PA was recommended to a user, but the user provided negative feedback to the recommendation afterwards, the system used this feedback to re-label the corresponding PA (changing the original label 0 to -1), and the prediction model was re-trained to include this data. An example of how the system updated itself is shown in Fig. 5 below. In order to achieve this, we added a feedback system into the original application. Users could provide feedback on any of the recommendations provided by the system, by either tapping "I liked it" (1) or "I didn't like it" (-1) buttons, or leave any specific comments in the comment box. This feedback was continuously used to retrain the model, refined, and adapted the recommendation to reflect the users current state. This account for possible changes in the user's preferences over time. Ideally, with more users and data entered into the system, the accuracy of the prediction model, and the quality of the clustering improved.

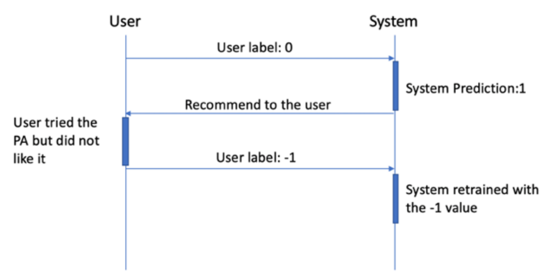

Fig. 5. Example of self-updated system

IV. Result and Analysis

Overall, we collected 178 questionnaire responses. Out of the 178 participants, 95 were males and 83 were females, with 17.4% age between 18 to 24 years old, 34.8% age between 25 to 34 years old, 26.4% age between 35 to 44 years old, 12.9% age between 45 to 54 years old, 6.2% age between 55 to 64, and 2.2% age between 65 to 74 years old. This study covered 29 different types of PA. Our player model consisted of 30 features (29 questions from the questionnaire and an additional Body Mass Index (BMI) calculated from height and weight).

A. Feature Importance

There were 16 out of 29 types of PA resulted in a prediction accuracy >=70%, the remaining 13 types of exercises resulted in relatively lower accuracy, as summarized in Table II below. As an example, if we look at the PA basketball, in our 178 questionnaire responses, 121 of them have either positive experience (1) or negative experience (-1) for basketball. The rest 57 replied N/A (0) because they never actually tried it. So, as we can see in Table III the response rate for basketball is 121/178 = 68%. For those 121 valid data, we use 80% of it (97) for training the model and the rest 20% (24) for testing the model (as described in section 5.2.2). We compared the predictive result with the actual label (1 or -1), then calculated the accuracy. For the 24 testing data, we predict 20 of them correctly and 4 of them wrong, so the accuracy is 20/24 = 83%.

TABLE II. ACCURACY FOR THE BINARY CLASSIFICATION PREDICTION MODEL

	Type of exercise	Response rate (non-0 value)	Accuracy
Accuracy >= 70%	Walking or jogging	100%	81%
	Running	100%	74%
	Hiking	97%	80%
	Swimming	85%	73%
	Riding a bike	86%	76%
	Dancing	74%	85%
	Skiing	58%	72%
	Ice-skating	69%	79%
	Golf	32%	81%
	Soccer	41%	78%
	Hockey	13%	74%
	Basketball	68%	83%
	Horse-riding	21%	73%
	Snooker or billiards	89%	78%
	Yoga	42%	81%
	Conditioning exercise	37%	71%
Accuracy < 70%	Aerobics	94%	54%
	Weight lifting	33%	66%
	Pilates	8%	62%
	Roller skating	72%	68%
	Extreme sports	9%	48%
	Martial arts. Boxing or wrestling	4%	46%
	Tennis or badminton	86%	64%
	Bowling	63%	64%
	Football	20%	57%
	Rugby	37%	55%
	Volleyball	56%	62%
	Fishing	43%	68%
	Sailing, wind-surfing or boating	16%	55%

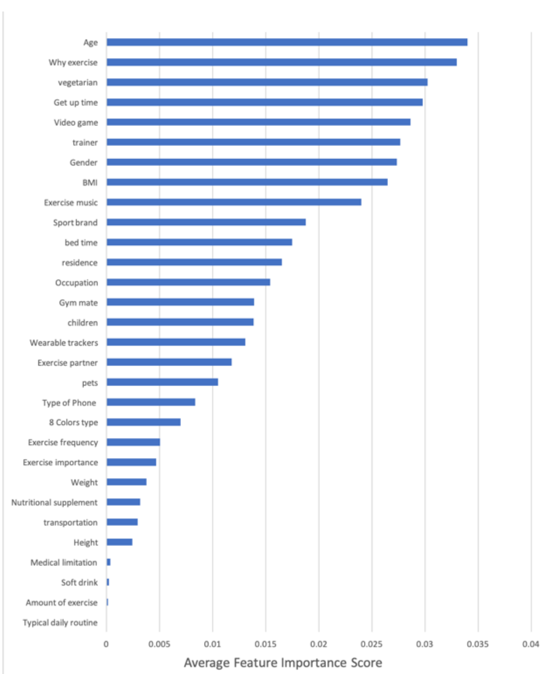

Fig. 6. Average feature importance score for all model parameters (in descending order)

TABLE III. AVERAGE FEATURE IMPORTANCE SCORE FOR ALL MODEL PARAMETERS

Feature	Average feature importance score
Age	0.033984
Gender	0.027345
Weight	0.003762
Height	0.002397
BMI	0.026455
Occupation	0.015428
8 Colors type	0.006964
Get-up time	0.029765
Bedtime	0.017492
Nutritional supplement	0.003148
Exercise importance	0.004654
Exercise frequency	0.005054
Why exercise	0.032987
Exercise partner	0.011794
Medical limitation	0.000349
Amount of exercise	0.000130
Trainer	0.027671
Residence	0.016504
Vegetarian	0.030217
Children	0.013856
Pets	0.010479
Exercise music	0.023982
Type of Phone	0.008344
Wearable trackers	0.013067
Soft drink	0.000228
Video game	0.028630
Gym mate	0.013893
Sport brand	0.018764
Transportation	0.002905
Typical daily routine	0.000031

The results of the average feature importance score are summarized in Table III above. Fig. 6 visualized the result in descending order. As we can see from the result, the most important parameters (top ten) in this prediction model were the user's age, purpose of exercise, if vegetarian or not, get up time, video game preference, if workout with a trainer or not, gender, BMI, favorite type of music when exercising, and favorite sport brand.

B. *Comparison with the 8-Color Model*

For those twenty-nine different types of PA covered in this study, we did not find any significant correlations between the 8-colors of fitness results and our participants' PA preference. Thus, we believe that personality type has only a limited influence on people's preference toward PA. On the other hand, other parameters such as an individual's lifestyle, physical conditions, and social connections, may play more important roles in determining their exercise taste. Table IV below shows an example (result of one of our participant UserID #97) comparison between our prediction model and the 8-colors model.

The average accuracy of the 8-colors model among our participants was 36.7% with a *SD* of 10.3%, compared to the average accuracy of our prediction model was 70.5% with a *SD* of 8.7%.

TABLE IV. EXAMPLE COMPARISON OF PREDICTIVE RESULTS BETWEEN THE 8 COLOR MODEL AND OUR PREDICTION MODEL FOR USER #97

UserID 97 (Reds)	**8-colors model prediction**	**Our prediction model**	**Actual label**
Running	-1	1	1
Walking	-1	1	1
Yoga	-1	1	-1
Swimming	-1	1	1
Tennis	1	-1	1
Hiking	-1	1	1
Basketball	1	1	1
Skating	1	-1	-1
Skiing	1	-1	-1
Soccer	1	-1	1
Dancing	-1	1	1
Cycling	-1	-1	-1
Pilates	-1	1	1
Strength training	-1	-1	1
	5 hit, 9 miss. Accuracy = 35.7%	10 hit, 4 miss Accuracy = 71.4%	

C. *System updates and the Pilot Study*

Our exergame player modeling system introduced in the introduction section (see Fig. 1) was updated with the sub-model - Exerciser Type Model (8-Color) - replaced by the new recommender system introduced above. The replacement of this sub-model could improve the accuracy of the PA prediction, and has the potential to engage exergame user for longer term.

Due to time constraints, we did not conduct another round of long-term study with the updated system yet. Instead, we invited five of our participants (3 males, 2 females) ranging in age from 23 to 36 years old (M = 28.67 year old, SD = 5.35 years old) from the previous study [34] which was a long-term study with an exergame game using the player model (Fig. 1). We emailed all the previous participants and chose the first five who responded with interest to try the new version of the system and provide feedback. A casual two-question questionnaire was collected after their seven-day use of the same exergame system they used for two months before but with an updated exerciser type model. Participants provided their responses on a 7-point Likert scale (from 1 = strongly disagree to 7 = strongly agree). The statements were:

1) I found the recommendation quality of this new version of the application better than the old one; and

2) I prefer using this new version of the application over the old one.

For question 1, the average score was 6.40 with a standard deviation of 0.55. Question 2 resulted in an average score of 6.20 with a standard deviation of 0.84. The sample size was too small for conducting formal statistical analysis, but it indicated a visible improvement of the new recommender system.

V. DISCUSSION

Overall, this study focused on the optimization of our previously proposed exergame player modeling system [34]. Firstly, a questionnaire was carried out to gather data regarding people's exercise preferences. Then based on the questionnaire results, a feature importance experiment was conducted to find out what parameters were more important at predicting users' preference for PA. Afterward, the recommender system was re-designed and implemented based on the prediction model. We compared the proposed model to an existing model on the accuracy in predicting user's PA type preference.

The new player model is a novel approach for PA recommendation that takes into account multiple parameters simultaneously and continuously learn and adapt to any changes in players preference over time.

The result shows the feasibility of using the player model for personalizing PA, as well as the potential of using machine learning in building the recommender system for PA. The preliminary result shows considerable effect in optimizing the system.

The comparison between our proposed system and the 8-colors model shows that based on our results, it is not sufficient to generate PA recommendations based only on a user's personality type. In fact, personality type has only limited influence on people's preferences toward PA while other parameters such as people's lifestyle, physical conditions, and social connections, may play more important roles in determining people's exercise taste.

Most existing player type models, and gamified user type models (e.g. BrainHex, Hexad, etc.) were built based on personality traits. However, for personalizing exergames, we argue that our proposed comprehensive exergame player model could be considered a potential alternative that could replace the traditional personality-based models (e.g., the 8-colors model). Particularly, our proposed model resulted in better performance in generating PA recommendation.

There were certain limitations in this study, which include:

- The binary classification doesn't show the detailed preference level; a multi-class classification may be used for better accuracy and the ranking of different PA.
- For different PA, the corresponding rank of the important features varied obviously. In this study, we aimed to find those common important parameters to build the general model, so we used the average feature important score for different PA. In order to generate more specific recommendations, we could consider each PA separately regarding parameter selection in the future.
- Sometimes it might not be realistic for some users to try those new PAs that our system recommended. For example, people living in big cities may not be able to

access certain types of outdoor activities. Therefore, in our future work, we would consider more viability issues when optimizing our recommendation algorithm.

- In this system, we did not look at the distance between PA (item-based recommendation method). With a hybrid recommendation method, we could also consider the distance between PA from different perspectives (cost, calorie consumption, in-door/outdoor activity, aerobic/non-aerobic, group/individual, summer/winter sport, hardness, intensity, flexibility, pleasure, and discipline, etc.)

VI. Conclusion

Overall, this study extended the exergame player modeling system we proposed in our previous research [34]. In this add-on study, we further examined what features are more important in building a PA recommendation model. Afterward, the recommender system was updated with a new recommendation model. The proposed model resulted in better accuracy when compared to an existing model in predicting users' preference toward PA types. A follow-up pilot study conducted with five participants to further verify the precision of the model indicated a visible improvement of the new recommender system.

Our proposed PA recommendation model is used within our own general exergame architecture, but it can be generalized to any other system that requires PA recommendation, including other exergames, health/fitness recommendation applications, and personal training systems, etc.

References

[1] Papastergiou, M., 2009. Exploring the potential of computer and video games for health and physical education: A literature review. Computers & Education, 53(3), pp.603-622.

[2] Hagen, K., Chorianopoulos, K., Wang, A.I., Jaccheri, L. and Weie, S., 2016, May. Gameplay as exercise. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 1872-1878). ACM.

[3] Dharia, S., Eirinaki, M., Jain, V., Patel, J., Varlamis, I., Vora, J. and Yamauchi, R., 2018. Social recommendations for personalized fitness assistance. Personal and Ubiquitous Computing, 22(2), pp.245-257.

[4] Campbell, T., Ngo, B. and Fogarty, J., 2008, November. Game design principles in everyday fitness applications. In Proceedings of the 2008 ACM conference on Computer supported cooperative work (pp. 249-252). ACM.

[5] Sami, A., Nagatomi, R., Terabe, M., & Hashimoto, K. (2008, July). Design of Physical Activity Recommendation System. In IADIS European Conf. Data Mining (pp. 148-152).

[6] Ferro, L. S., Walz, S. P., & Greuter, S. (2013, September). Towards personalised, gamified systems: an investigation into game design, personality and player typologies. In Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death (p. 7). ACM.

[7] McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. Journal of personality, 60(2), 175-215.

[8] Busch, M., Mattheiss, E. E., Hochleitner, W., Hochleitner, C., Lankes, M., Fröhlich, P., ... & Tscheligi, M. (2016). Using Player Type Models for Personalized Game Design-An Empirical Investigation. IxD&A, 28, 145-163.

[9] Nacke, L. E., Bateman, C., & Mandryk, R. L. (2011, October). BrainHex: preliminary results from a neurobiological gamer typology survey. In International Conference on Entertainment Computing (pp. 288-293). Springer, Berlin, Heidelberg.

[10] Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016, October). The gamification user types hexad scale. In Proceedings of the 2016 annual symposium on computer-human interaction in play (pp. 229-243). ACM.

[11] Orji, R., Tondello, G. F., & Nacke, L. E. (2018, April). Personalizing persuasive strategies in gameful systems to gamification user types. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (p. 435). ACM.

[12] Courneya, K. S., & Hellsten, L. A. M. (1998). Personality correlates of exercise behavior, motives, barriers and preferences: An application of the five-factor model. Personality and Individual differences, 24(5), 625-633.

[13] Rhodes, R. E., & Smith, N. E. I. (2006). Personality correlates of physical activity: a review and meta-analysis. British journal of sports medicine, 40(12), 958-965.

[14] Brue, S (2008). The 8 Colors of Fitness. Oakledge Press, Delray Beach, Florida, USA.

[15] Taylor, K. (2017). The Relationship Between Personality Types and Exercise Preferences and Behaviors.

[16] Bayne, R. (2012). The Counsellor's Guide to Personality: Understanding Preferences, Motives and Life Stories. Macmillan International Higher Education.

[17] Guo, X., Liu, J., & Chen, Y. (2017, May). FitCoach: Virtual fitness coach empowered by wearable mobile devices. In INFOCOM 2017-IEEE Conference on Computer Communications, IEEE (pp. 1-9). IEEE.

[18] He, Q., Agu, E., Strong, D., & Tulu, B. (2014, November). RecFit: a context-aware system for recommending physical activities. In Proceedings of the 1st Workshop on Mobile Medical Applications (pp. 34-39). ACM.

[19] Broekhuizen, K., van Poppel, M. N., Koppes, L. L., Brug, J., & van Mechelen, W. (2010). A tailored lifestyle intervention to reduce the cardiovascular disease risk of individuals with Familial Hypercholesterolemia (FH): design of the PRO-FIT randomised controlled trial. BMC Public Health, 10(1), 69.

[20] Rabbi, M., Aung, M. H., Zhang, M., & Choudhury, T. (2015, September). MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 707-718). ACM.

[21] Sports A-Z, (2019) retrieved from: https://www.bbc.com/sport/all-sports

[22] Wagstaff, K., Cardie, C., Rogers, S., & Schrödl, S. (2001, June). Constrained k-means clustering with background knowledge. In Icml (Vol. 1, pp. 577-584).

[23] Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. International Journal, 1(6), 90-95.

[24] Kassambara, A., & Mundt, F. (2017). Package 'factoextra'. Extract and visualize the results of multivariate data analyses, 76

[25] Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems*, *46*, 109-132.

[26] Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, *12*(4), 331-370.

[27] Pincay, J., Terán, L., & Portmann, E. (2019, April). Health Recommender Systems: A State-of-the-Art Review. In *2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG)* (pp. 47-55). IEEE.

[28] Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, *9*(3), 293-300.

[29] Casalicchio, G., Molnar, C., & Bischl, B. (2018, September). Visualizing the feature importance for black box models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 655-670). Springer, Cham.

[30] Putin, E., Mamoshina, P., Aliper, A., Korzinkin, M., Moskalev, A., Kolosov, A., ... & Zhavoronkov, A. (2016). Deep biomarkers of human aging: application of deep neural networks to biomarker development. *Aging (Albany NY)*, *8*(5), 1021.

[31] Sutin, A. R., Stephan, Y., Luchetti, M., Artese, A., Oshio, A., & Terracciano, A. (2016). The five-factor model of personality and physical inactivity: A meta-analysis of 16 samples. Journal of Research in Personality, 63, 22-28.

[32] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), 16-28.

[33] Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016, October). The gamification user types hexad scale. In *Proceedings of the 2016 annual symposium on computer-human interaction in play* (pp. 229-243).

[34] Zhao, Z., Arya, A., Whitehead, A., Chan, G., & Etemad, S. A. (2017, May). Keeping Users Engaged through Feature Updates: A Long-Term Study of Using Wearable-Based Exergames. In *CHI* (pp. 1053-1064).