Keeping People Active and Healthy at Home Using a Reinforcement Learning-based Fitness Recommendation Framework

Elias Tragos[1], **Diarmuid O'Reilly-Morgan**[1], **James Geraci**[3], **Bichen Shi**[1], **Barry Smyth**[1], **Cailbhe Doherty**[1,2], **Aonghus Lawlor**[1], **Neil Hurley**[1]
[1]Insight Centre for Data Analytics, University College Dublin
[2]School of Public Health, Physiotherapy and Sports Science, University College Dublin,
[3]Samsung Electronics Co., Ltd., Seoul, Republic of Korea
{elias.tragos, diarmuid.oreillymorgan, bichen.shi, barry.smyth, aonghus.lawlor, neil.hurley }
@insight-centre.org, cailbhe.doherty@ucd.ie, james.geraci@samsung.com

Abstract

Recent years have seen a rise in smartphone applications promoting health and well being. We argue that there is a large and unexplored ground within the field of recommender systems (RS) for applications that promote good personal health. During the COVID-19 pandemic, with gyms being closed, the demand for at-home fitness apps increased as users wished to maintain their physical and mental health. However, maintaining long-term user engagement with fitness applications has proved a difficult task. Personalisation of the app recommendations that change over time can be a key factor for maintaining high user engagement. In this work we propose a reinforcement learning (RL) based framework for recommending sequences of bodyweight exercises to home users over a mobile application interface. The framework employs a user simulator, tuned to feedback a weighted sum of realistic workout rewards, and trains a neural network model to maximise the expected reward over generated exercise sequences. We evaluate our framework within the context of a large 15 week live user trial, showing that an RL based approach leads to a significant increase in user engagement compared to a baseline recommendation algorithm.

1 Introduction

The United Nations (UN) agenda for sustainable development includes 17 actions and 169 goals towards shared global peace and prosperity. Sustainable development goal 3 aims to "Ensure healthy lives and promote well-being for all at all ages", while target 3.4 [Cf, 2015] aims to reduce the burden of non-communicable diseases upon the global population, in the hope of reducing world wide morbidity and promoting mental health and well-being. The key indicators for this target include the mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease, and the mortality rate from suicide. Within economically developed societies, non-communicable diseases such as diabetes and cardio-vascular diseases are strongly related to poor physical fitness [Bassuk and Manson, 2005; WHO, 2022]. Additionally, many studies have identified a strong connection between heightened physical activity and good mental health [Mikkelsen *et al.*, 2017]. In 2013, the Council of the European Union published recommendations promoting health-enhancing physical activity (HEPA), highlighting the benefits for physical and mental health and recommending at least 75 minutes of intense activity per week [CoEU, 2013]. Many EU countries have adopted WHO's recommendations establishing national strategies promoting HEPA [DoH, 2016; FORM, 2008; Barthélémy *et al.*, 2016].

Despite the strong link between fitness and morbidity, the prevalence of exercise in affluent societies remains low. As argued by Smyth [Smyth, 2019], this phenomena constitutes an evolutionary paradox, whereby the behaviour for which individuals have an adaptive preference (increased relaxation and leisure time), is easily supported by their affluent circumstances, but detrimental to their long-term health. While economic growth leads to material comfort, individuals receive less motivation to engage in exercise benefiting their health. This was only exacerbated by the COVID-19 pandemic. With gyms and public recreation spaces closed due to health restrictions, physical and mental health was seen to suffer widely [Levy *et al.*, 2021; Maugeri *et al.*, 2020].

In recent years, various mobile apps promoting a healthier lifestyle have been released [Venkatachalam and Ray, 2022], with some also exploiting wearable technologies (smartwatches) for monitoring users' heart rate [Fitness, 2023]. The element of gamification inherent in many smartphone platforms often helps to encourage participation [Lister *et al.*, 2014], with users competing as an incentive for exercising more. Though the frequent closure of gyms throughout the COVID-19 pandemic curtailed people's regular patterns of exercise, this also spurned a growing interest in exercise at home [Nyenhuis *et al.*, 2020]. However, within the domain of exercise recommendation, most existing solutions confine themselves to recommending predefined workout plans, potentially foregoing the benefits of a more granular degree of personalisation [Fitness, 2023; Ifivolve, 2023].

This work, inspired by the home fitness revolution during the COVID-19 pandemic, is aligned with the EU's recommendations and the national HEPA strategies and aims at promoting physical activities that can help people of all ages to maintain a good level of physical and mental health through personalised home workouts. Contrary to existing applications, which exploit machine learning only to select from predefined workout plans, this framework creates workout sessions from scratch, selecting exercises from a pool of body-weight exercises and constructing a full workout session that maps to user's goals and preferences. Given the sequential dependency of the exercises in a workout session, reinforcement learning (RL) was identified as a suitable approach for providing personalised workout recommendations. However, a major constraint on this work was the unavailability of a large amount of training data to ground a RL based system. To address this data scarcity issue, a User Simulator was developed, which generated synthetic data and simulated users with "realistic behaviours" and "realistic feedback", so that an RL model could be effectively trained to address the exercise recommendation problem. We comprehensively test our solution both within the simulation framework and through a large randomized crossover trial.

Overall, the contributions of this work are:

- An RL-based fitness framework for providing personalised exercise recommendations to home users;
- A User Simulator for generating synthetic users with realistic behaviour and feedback, enabling training an RL model with limited real data;
- A set of metrics (as rewards) to evaluate the quality of workout sessions using sports science knowledge;
- A simulation based comparison of the RL framework against a particle swarm optimization (PSO) baseline;
- A real-world trial evaluation consisting of 69 users and 559 workout sessions, comparing the performance of the RL framework against the PSO baseline.

2 Related Work

RS for health and exercise is a rapidly growing field of research. There are several works investigating the efficacy of intelligent recommendations applied to marathon running [Berndsen *et al.*, 2019; Berndsen *et al.*, 2020], which, however, remain targeted at niche markets. There are several fitness mobile applications on the market today that try to recommend a healthier lifestyle to their users, and although several of them claim to use machine learning to provide more personalised recommendations, they generally provide predefined workout programs that have been created by fitness coaches. Their "personalisation" is related to mapping user preferences (based on questionnaires) to the most suitable pre-existing program, without the user being able to modify the program or adapt it to their needs [Ifivolve, 2023; Freelitics, 2023]. Other applications recommend times when users should workout, or the type of activity that a user should engage in, selecting from a pool of activities such as running, cycling, swimming, walking, etc . [Mahyari and Pirolli, 2021; Dharia *et al.*, 2018; Sami *et al.*, 2008; Abdulaziz *et al.*, 2021; Guo *et al.*, 2017; Tran *et al.*, 2018; Gymfitty, 2023].

Recently, RL has grown in popularity in recommendations, especially within the field of session-based recommendation [Afsar *et al.*, 2022]. Within this paradigm, the recommendation problem is formulated as that of maximizing cumulative reward over a sequence of recommendations, rather than the myopic approaches popular in rating based or top-N recommendation. Jain et al [Ie *et al.*, 2019b] applied the approach to slate recommendation. Shi et al developed the concept of a gym for training recommender models on traditional RS datasets [Shi *et al.*, 2019]. In the absence of enough data to learn either a strong value function, or to enable learning from logged feedback, it is common to employ the use of a simulator [Ie *et al.*, 2019a]. Though many algorithms exist for learning RL policies, proximal policy optimization (PPO) [Schulman *et al.*, 2017] is often favored in practical settings, due to its parallelism and non-reliance on a replay memory.

The problem of selecting the exercises to include in each session can be seen as analogous to that found in several traditional areas of RS research, such as session-based or sequential recommendations [Afsar *et al.*, 2022], playlist or shopping list recommendations [Liebman *et al.*, 2014], and slate recommendations [Mehrotra *et al.*, 2019]. However, the workout recommendation problem has some distinct features which differentiate it from traditional areas of RS. Unlike sequential recommendation, the user is presented with a complete workout session, consisting of n_s exercises, while user feedback for the previous items in the session is unknown at the time of generating the recommended list. Compared to slate recommendation, the user doesn't select only a few items from the recommended slate, but is expected to complete all of the exercises in the specified order. The workout session can include the same item multiple times, as it is common for users to repeat exercises within a workout. As shown in the sports science literature [Hoffman and others, 2011], the sequential dependency of the exercises in the session is of high importance, both to users' ability to complete a session, and their satisfaction with it. This is a marked difference with the domains of shopping list or playlist recommendation, since a user might require a steady progression of difficulty to the exercises or they might find that certain sets of exercises naturally flow from one to the other.

3 An RL-based Health and Fitness Recommendation Framework

3.1 Problem Definition

Given the positive link between regular exercise and physical and mental health, our aim is to develop a platform that both enables and encourages users to engage in workouts at home, without the need for complicated equipment. User satisfaction after a workout is assumed to be strongly correlated with their long term continued engagement [Fernández-Martínez *et al.*, 2020]. As such, the personalisation of workout programs towards the needs of individual users should play a key role in encouraging consistent use of the platform. Given the sequentially dependent nature of exercise programs, reinforcement learning appears to be a strong candidate for delivering this degree of personalisation.

The funders of this work provided a home-fitness mobile application, which delivers body-weight exercise workout programs via a series of short videos. The app allows users to create ”Workout” programs, stating their overall fitness goals, such as whether they wish to focus on improving cardio or strength, which muscles they wish to train, and how frequently and for how long they wish to exercise. During each exercise period, they are presented with a list of preselected body-weight exercises and instruction videos. During and after a session, users are encouraged to deliver feedback, either by *skipping* or *changing* the exercises that they find unsuited to them, or by answering a questionnaire at the end of the session. One limitation of this mobile application is that there is no option for the users to give any “direct” feedback per exercise to indicate if they liked or disliked specific exercises in the session. This complicates the problem of inferring user preference per exercise, as there is no explicit way to identify which exercises the user likes or dislikes, and this must be inferred from the user's satisfaction score and the user's actions as to skipping or changing exercises.

Given these dependencies, we formulate our problem set up as that of finding a policy that recommends sessions that maximise user satisfaction. We wish to choose a recommendation policy π that maximises the expected value of user satisfaction, F, with the recommended session $\mathbf{a}_\pi = (a_1, \ldots, a_{n_s})$, where each a_i is a recommended exercise and the workout consists of n_s exercises. The expectation is taken over the arrival process of users to the system. Writing the optimal policy as π^*, and assuming that the policy is parameterised by parameters θ, we write the general problem as:

$$\theta^* = \arg\max_{\theta} \mathbb{E}_u \left[F(u, \mathbf{a}_{\pi_\theta})\right]; \pi^* = \pi_{\theta^*} \qquad (1)$$

Here, we adopt the standard RL formalisation of agent, environment, action, state, observation and reward. An agent is a mobile application, which learns the model that optimises the satisfaction of the users served by the application. The environment consists of the users served by the application and the feedback that they give to the application. Rewards are a function of user feedback and session quality metrics, which have been tailored in consultation with sports scientists.

Rather than considering an action to correspond to the recommendation of a full workout session, instead an action is the recommendation of the next exercise in the workout sequence to the current active user. This is therefore an episodic RL formulation, in which each episode is terminated when a complete workout of n_s exercises has been recommended, where n_s is a fixed session size selected by the active user. Each time there is a need to make a recommendation for a user u, the initial state s_u is computed including a description of the user and their history and preferences. The agent then interacts with the environment for n_s steps, giving an action a_k and receiving an observation o_k and a reward r_k for each step k, in order to produce the sequence of actions/exercises $(a_0, a_1 ... a_{n_s})$ in the session until the end of the episode.

3.2 Modelling the Fitness RL Framework

The overall fitness RL framework is depicted in Figure 1. A user interacts with the framework through a mobile app, receiving exercise recommendations (actions) and providing their feedback and their exercise history. The RL framework consists of an RL gym and an RL agent. The *agent* is parameterised by a neural network, as described in the next subsection. The *gym* represents the user and the overall environment. The *gym* includes a User Simulator that functions to create synthetic interactions, allowing the *agent* to be trained with only limited real world data. The *gym* interacts with the *agent*, sending the current “state” of the environment, receiving recommended actions, and then providing a reward and a new state/observation. As seen in the figure, we also employ “action masking”, both for speeding up the training process of the *agent* and imposing some hard constraints. In this respect, the *gym* also sends an action mask vector to the *agent*, which zeros out constrained actions. These constraints might involve exercises that the user has disliked in the past or that the user has specified that they are unable to perform.

Below we provide further details of the RL setting:

Actions

The actions in the fitness recommendation problem represent the 161 exercises that are available within the mobile app. The exercises are standard body-weight exercises that can be easily completed at home. Exercises are grouped into 10 categories of similar exercises (i.e. Walking Burpee and Squat Burpee are examples of Burpees). The exercises have various attributes, such as their level of difficulty, the list of muscles that they train, whether they target cardio or strength, and any equipment required for their execution (i.e. chair).

State

Within the formal RL paradigm, the state includes all of the information necessary to determine both the reward and the subsequent state transition at a given point in time. Within the exercise recommendation scenario, this includes:

- *User information*: this includes various user characteristics such as their initial fitness level and summary statistics for their prior workout history within the app. The initial fitness level is expressed in terms of difficulty preference for each of the exercise categories, assessed by a questionnaire when a user registers with the app.
- *Workout goals*: this includes information for the user's particular exercise targets, i.e. if the focus is on cardio or strength training, which muscles they wish to train and what equipment they have available to exercise with.
- *Session status*: this comprises parameters describing the status of the session at any given point in time, including the exercises already recommended and their corresponding difficulties, the exercises recommended in the previous L sessions, and information regarding the target muscles trained and muscle fatigue, etc.
- *User feedback*: this includes information indirectly disclosing the user exercise preference i.e. which exercises they liked or disliked in the previous N sessions, whether they completed or skipped exercises, etc.

In the general modelling of the workout recommendation problem, the state has both “observable” and “non-observable” or “partially-observable” parts, which could lead to modelling the problem as a partially observed Markov decision process. In the current implementation of the RL

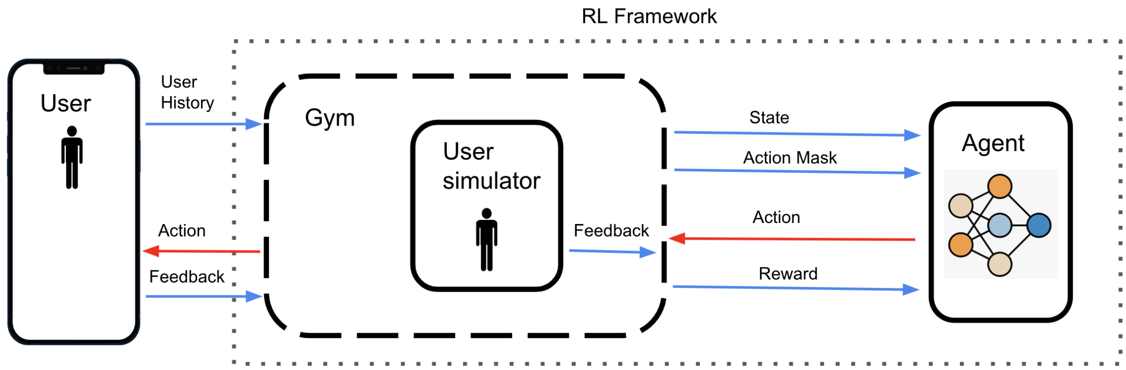

Figure 1: The overall structure of the RL framework

framework though, we assume that all gym state is observable to the RL model. Though real world users would undoubtedly have some unobservable state, such as their preference for certain exercises, within the context of a small trial there is insufficient data for this to be reliably inferred, and so we take the users' logged feedback to suffice for these preferences.

Reward

With a large enough amount of data, and using modern techniques such as RL for human feedback [Ziegler *et al.*, 2019], it should be possible to empirically learn a reward function matching the preferences of individual users. With limited data however, to create a personalised reward we must take account of multiple relevant signals, such as user fitness level, workout goals, and previous feedback given to the application. Thus, exploiting sports science knowledge [Hoffman and others, 2011], we assume that the objective of maximizing user satisfaction can be split into multiple sub-objectives, each one characterised by an individual reward. These sub-objectives are related with the quality of the recommended session and how well the session fits users preferences and goals. In contrast with works where the RL mode receives reward after each step of the training process, in this work we try to emulate the "real world" session data and only return a non-zero reward at the end of an episode, thus mimicking the real life situation in which explicit feedback is only available at the end of the session. The individual reward components generalise a number of metrics tracking which aspects of a session matter to a user. Specifically, the components are:

1. **Intra-session Diversity**, which measures the diversity of the exercises recommended within a session. The similarity of two exercises is computed via comparing the muscles trained by each movement. The assumption is to have enough diversity in the session so that users train several muscles and aren't bored from repetition.

$$R_{\text{intra}} = 1 - \frac{\sum_{i \neq j} w^{|i-j|-1} \text{sim}(a_i, a_j)}{\sum_{i \neq j} w^{|i-j|-1}} + p\,,$$

   where $\text{sim}(a_i, a_j)$ is the similarity of two movements computed as the mean similarity of the muscles they train and p is a penalisation factor which is -1 if the same exercise is repeated consecutively.

2. **Inter-session Diversity**, which measures the extent to which the exercises recommended in the current session differ from those recommended in the previous L sessions. As above, the assumption is that users should complete a variety of different exercises over sequences of workout sessions, so that their muscles get enough rest between workouts.

$$R_{\text{inter}} = \frac{1}{L} \sum_{j}^{L} w_j \left(1 - \text{corr}(\mathbf{s}_i, \mathbf{s}_{i-j})\right)\,,$$

   where $\mathbf{s}_i$ is a vector containing the counts of each movement in session i, corr is the Pearson correlation and w_j is a weight that decreases with the age of session j

3. **Session matching Fitness level (FL)**, which measures the extent to which a given set of exercises matches the fitness level of the user. This is measured using the assessed difficulty of the exercises (mdiff) and the difficulty preference of the user (pfdiff) per exercise category, assuming that these should be relatively close, so that the user isn't recommended very easy or very difficult exercises, allowing though room for attempting more difficult exercises in the case that the user's fitness level has progressed.

$$R_{\text{FL}} = \text{rescale}\left(\sum_{i}^{n_s} \frac{1 - \mid \text{mdiff}(a_i) - \text{pfdiff}(a_i) \mid}{\max(\text{mdiff}(a_i), \text{pfdiff}(a_i))}\right)$$

   where rescale is a Gaussian Kernel function (scaled between [0,1]) that rescales the score using the normal distribution across a mean (0.9 here) and a standard deviation (0.15 here), to ensure that the exercises are close to the user's FL, but with some room for variation.

4. **Session matching user workout Goal**, which measures the extent to which the exercises in a session match the user's workout goal (cardio or strength).

$$R_{\text{goal}} = w\text{perc} + (1 - w)\text{consec}$$

   where,

$$\text{perc} = \text{rescale}\left(\frac{1}{n_s} \sum_{i=1}^{n_s} \mathbb{1}(\text{style}(a_i) = \text{target}), 0.625\right)$$

is a score which shows how close the percentage of cardio/strength exercises in the session is to a target value (i.e. 62.5% for the user set goal) and

$$\text{consec} = \text{rescale}(\max_{i,j}(\delta(i,j)), 2.1, 2)\,,$$

is a score based on the number of consecutive exercises of one style/goal using the normal distribution, assuming that exercise styles should alternate after 2 consecutive movements from the same style. $\delta = j - i + 1$ if the sequence of actions $a_i \ldots a_j$ are of the same style, else 0.

5. **Session matching Focus Muscles (FM)**, Consider that we have a finite set of muscles, $\mathcal{M}$, and that the user selects from this set a subset of primary muscles, $\mathcal{M}_P$, and secondary muscles, $\mathcal{M}_S$, to train within the session. This reward measures the extent to which the exercises in a session correspond to these selected muscles. For any exercise a, write M_a for its set of associated muscles i.e. the muscles which are exerted by that exercise. Then, for real-valued weights w_0 and w_1,

$$R_{\text{musc}} = \frac{w_0}{N}\sum_{i=1}^{n_s} \mathbb{1}(M_{a_i} \subseteq \mathcal{M}_P \cup \mathcal{M}_S) + w_1 \frac{\sum_{i=1}^{n_s} \mathbb{1}(M_{a_i} \subseteq \mathcal{M}_P)\mathbb{1}(i \not\equiv 0 \bmod 4)}{\sum_{i=1}^{n_s} \mathbb{1}(M_{a_i} \subseteq \mathcal{M}_P)}$$

where $\mathbb{1}(.)$, is the indicator function that evaluates to 1 when its argument is true and 0 otherwise. The use of $\mathrm{mod} 4$ allows for a rest between sets.

6. **Muscle fatigue**, which measures the extent to which muscles become tired through the course of a session, for example by the same group of exercises being performed in close succession. For some weight w,

$$R_{\text{fatigue}} = \max_{m\in\mathcal{M}} \sum_{i,j} w^{|i-j|-1}\mathbb{1}(m \in M_{a_i} \cap M_{a_j})\,.$$

The reward score is then rescaled using the Gaussian Kernel with mean 1.6 and deviation 2 to provide a lower score when sessions have more than 2 consecutive exercises training the same muscle.

7. **Wrists**: This reward penalizes the recommendation of too many wrist dependent exercises (which was extracted as a requirement by users in an older trial). For example, mountain climbers, push-ups and burpees all use wrists, and although they are exercises of different categories training different muscles, they all put pressure on the wrists which can cause injury to users.

$$R_{\text{wrists}} = rescale(\max_{i,j}[\delta(i,j)]\ , 1.6, 2), \tag{2}$$

where $\delta(i,j) = j - i + 1$ if the sequence of actions $a_i, a_{i+1} \ldots a_j$ exercise the wrists, else 0.

Though a reward that discourages the model from recommending previously disliked movements was found to be broadly effective, not knowing the empirical distribution of real user preferences created a danger that the model would learn to extrapolate preferences during simulated training in a manner that wouldn't generalise to the preferences of users in production. As such, it was found to be safer to temporarily mask out negative feedback items during production.

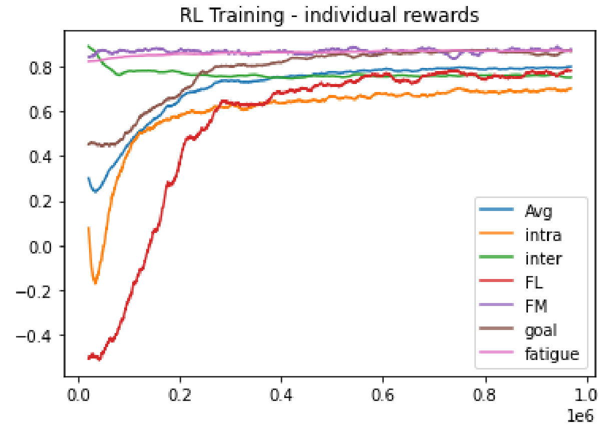

Figure 2: Progress of the individual rewards during training.

Reward design: The overall principle of the reward design is to apply boosting and penalisation mechanisms to each individual reward to help the RL agent to learn to distinguish good session from bad ones. All the individual rewards have been constructed so as to provide a real-valued number in $[0, 1]$. For each individual reward, a threshold th_{rp} has been defined, so that sessions with rewards below that threshold are considered "bad" and are penalised, while sessions with rewards above a threshold th_{rb} are boosted. Additionally, if at least one of the rewards is penalised, the overall reward returned for the episode becomes negative. If no individual reward is penalised, then the overall reward is the *weighted average* of the individual rewards $R = \sum_n w_i R_i$ where R_i and w_i are the individual rewards and their corresponding weights, with $\sum_n(w_i) = 1$. The weights are determined by domain knowledge and represent the importance that each reward has on the overall quality of a session, i.e. the intra-session diversity and the fitness level rewards have a higher weight compared to the inter-session diversity or session style rewards. The progress of the individual rewards during training can be seen in Figure 2, which shows that despite the fact that the agent only sees the weighted sum of the rewards, with the exception of inter-session diversity, all rewards increase towards a plateau. The slight decrease in the inter-session diversity is due to the high diversity of the largely random sessions generated at the beginning of training, but it also plateaus in the end.

User Simulator

Due to a limited amount of real world data, we developed a simple but efficient User Simulator as part of the fitness RL framework. Briefly, the goal of the User Simulator is to generate synthetic data in the form of users that can provide feedback that "emulates" the actions of real users. The User Simulator samples users from probability distributions, giving them random descriptions (i.e. age, gender, initial fitness level), workout goals (cardio/strength, workout duration, target muscles, available equipment), and preferences per exercise. To generate "realistic" user feedback and behaviour, we conducted an analysis of the data gathered during an initial exploratory user trial executed in 2021. This analysis helped to understand when users skip exercises during a workout and when users change exercises to easier or more difficult ones. This analysis resulted in probability equations to estimate at each specific state if a simulated user will skip or change an exercise, providing such feedback to the environment and the agent. The analysis showed a higher probability of skipping an exercise if: (i) it was encountered in the later stages of the session, (ii) the user was likely to have an intrinsic dislike for the exercise, and (iii) multiple movements within the same exercise category had already been recommended in the session. The analysis also showed that the probability that a user changed an exercise to easier/more difficult was directly related with the difference between the exercise difficulty and the user's difficulty preference for the movement's exercise category. However, these probabilities were rescaled accordingly, considering that the overall percentage of users skipping or changing exercises was low.

3.3 Reinforcement Learning Model

Given that the user simulator and gym encapsulate our assumptions about the real world environment, it is possible to approach the problem of choosing a sequence of actions with an off the shelf RL algorithm. In this instance, we select the Policy Proximal Optimization (PPO) [Azizzadenesheli *et al.*, 2018] algorithm due to it's high parallelism, and it's known efficacy on practical problems. The input of the policy network is an observation of size 1287, and the output of the network is a probability distribution over the 161 candidate movements. The middle of the network is composed of a couple of densely connected layers. The value network, which is used to calculate the advantage in PPO, shares all layers with the policy network, except that the output layer is a single node. We use the PPO implementation in RLlib [Liang *et al.*, 2017], adding an action mask that blocks candidate movements in certain situations, for instance when movements require special equipment that a user does not possess. We train the model for a total of 97000 episodes.

4 Evaluation

4.1 Baseline

The baseline model that was built into the mobile phone application by the funders of this work utilises Particle Swarm Optimization (PSO) [Poli *et al.*, 2007] to generate sessions. The PSO algorithm optimizes for a number of session attributes such as diversity and fitness level, although it doesn't personalise the sessions in accordance with user feedback.

4.2 Simulation Comparison Against the Baseline on Simulated Users

The first step of the evaluation was to compare the RL model with PSO on a number of simulated users, modelled within the gym environment. For this evaluation, we employed session quality metrics, which are modified versions of the rewards upon which the RL rewards were based without boosting or penalisation ("avg" is the weighted average of the rewards). Figure 3 shows a comparison of the rewards averaged across 100 simulated users completing 10 sessions each, with workout recommendations provided either via the baseline or

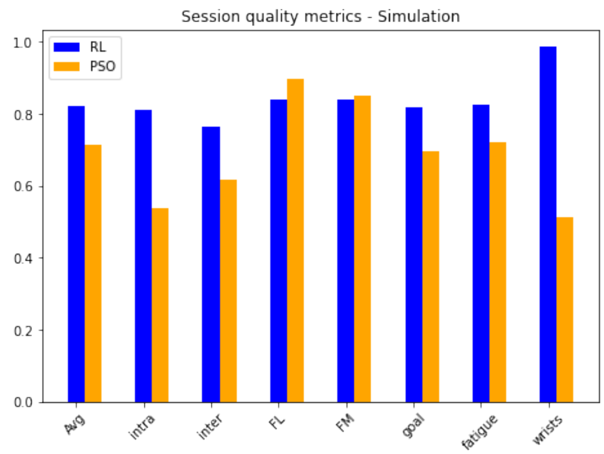

Figure 3: Comparing the session quality metrics for the RL model and the PSO on 100 simulated users for 10 sessions per user

RL model. The figure demonstrates that the RL model recommendations are superior in terms of these metrics. It is interesting to see that PSO performs well on the fitness level metric, which means that it provides sessions with exercises that are closer to the user's fitness level compared to RL. However, as optimizing this metric limits the number of exercises available, it directly conflicts with diversity (both intra-session and inter-session). As such PSO provides users with recommendations that are largely similar between sessions, with multiple similar exercises within the same session. This effect can also be seen in the very low score PSO achieves for the wrists reward - using many exercises that put strain on the wrists. Overall, as seen in the figure, RL's average score per session is approximately 20% higher than that of PSO.

4.3 User Trial Evaluation

The fitness recommendation framework was further evaluated in a real world trial that ran for 15 weeks. We collect a dataset of 69 users with 559 workout sessions in total. There were 27 males and 42 females participating in the study. 268 sessions were completed featuring RL recommendations, and 291 sessions using the baseline. The trial was performed using a randomised crossover protocol, similar to [Bonafiglia *et al.*, 2016] with an interval of one week. At the start of the trial, participants were randomly assigned to either RL or PSO recommendations, and switched at the end of each week.

At the end of each session, users were presented with a Paces-8 questionnaire [Mullen *et al.*, 2011] rating their session experience according to eight criteria. Generalized estimating equations (GEE) [Hardin and Hilbe, 2012] were used to assess differences in user responses, using an exchangeable correlation structure. The model was corrected for dependent observations by including participants' id as a subject effect. The a priori p-value for this analysis was set at $p < 0.05$.

The GEE estimated a main mean effect for condition (p = 0.015). The mean PACES-8 response differed significantly between sessions generated by the RL model and the baseline (mean response for RL condition = 4.0 $[95\%CI$: 6 to 6.78]; mean response for the baseline condition $= 3.73[95\%CI$: 4.9

	Paces8								Overall
	Pleasurable	Fun	Pleasant	Invigorat.	Gratifying	Exhilar.	Stimulat.	Refreshing	Overall
PSO	3.66	3.53	3.53	3.75	3.84	3.74	3.85	3.84	3.73
RL	**3.88***	**3.72**	**3.71**	**4.15***	**4.17***	**4.04***	**4.13***	**4.06***	**4***

Table 1: User Trial Results comparing the average PACES-8 responses for PSO and RL. Results with an $*$ indicate significance at $p < 0.05$.

Sentiment	Corr. with Satisfaction	P-Value	Mean RL	Mean PSO
Diversity	0.594	10^{-7}	0.103	-0.722
Intensity	0.102	0.51	0	-0.167
Exercises	0.266	0.06	-0.565	-0.852
App Issues	-0.13	0.437		

Table 2: Qualitative analysis of user feedback comments

to 5.55], P = 0.015). An analysis of the parameter estimates associated with these main effects revealed that participants were likely to be more satisfied with an exercise session that was generated by the RL model. Summary statistics of the PACES-8 responses are presented in Table 1, showing that RL produced a significant effect over the baseline for six of the eight PACES-8 responses. The table shows clearly an improvement of 7% in the mean response of users when they received RL recommendations compared to the baseline.

One motivation for building this personalised fitness recommendation framework was to keep users engaged over time. As shown in Figure 4 (a), an analysis of the mean response over time showed that satisfaction with PSO broadly decreased, whilst for RL it in fact tended upwards. This positive result indicates that RL kept users engaged for a longer time, motivating them to continue with their workouts in the long run, as opposed to the baseline method where satisfaction degraded from an initial naive peak response.

Increased satisfaction with the RL framework can also be seen using a sentiment analysis on text feedback provided by the users of the app. When submitting feedback at the end of a session, users were asked to fill in a textbox with short comments on the workout recommendations. A sentiment analysis using NLTK's Vader package [Hutto and Gilbert, 2014] in Python was performed and the results are shown in Figure 4(b). The results show that the comments given to RL sessions had a much higher positive score (23.13%) than those for the baseline model, while the overall compound score of the RL comments was 48.88% higher than that of the baseline model. This indicates that the users were significantly more satisfied and gave more positive comments when they received recommendations from the RL model, as opposed to recommendations from the baseline.

Furthermore, shown in Table 2, a qualitative analysis of the user comments was performed, manually labelling each comment as responding either positively (+1), negatively (-1), or neutrally (0) to a number of key criteria (diversity, intensity, and choice of exercises) and measuring the correlation of each metric with the user satisfaction. We note here that users tend to comment negatively more than positively. The diversity of recommended exercises was seen to correlate strongly with the PACES-8 responses, with a significant difference in response between RL and PSO. A small but not significant effect was seen relating to exercise choice. General issues with the application (e.g loading times, quality of the exercise videos, crashes) were seen to only weakly correlate with the PACES-8 response, indicating that users adhered to evaluating the recommended workouts, rather than the overall app experience.

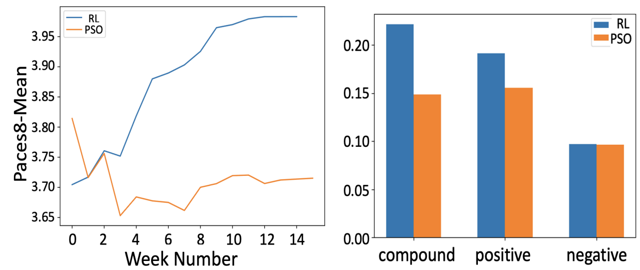

Figure 4: (a) Weekly averages for RL and PSO. (b) Sentiment analysis on the user trial showing positive, negative and compound scores.

5 Conclusions

An RL framework for health and fitness has been developed, providing personalised recommendations for promoting healthy and active living at home. Compared to existing work, the proposed framework uses RL to provide a complete workout session to users and learns from both their actions during the workout and feedback after the workout so that it builds an accurate user profile and provides more tailored recommendations in future sessions. A comparison of the proposed RL model with a PSO baseline model was performed using both simulations and a real-world user trial. In both cases, the proposed RL model significantly outperformed the baseline, with higher user satisfaction and sentiment scores.

The use of a crossover protocol in the experiments hindered a direct assessment of long term engagement. However, when taken as proxy for user engagement, the responses indicate that the level of personalisation of RL sessions RL served as increased motivation for participants to exercise. Thus, further development of such methods can reduce the motivation gap identified in [Smyth, 2019], and nudge users into better daily routines contributing to longer term adoption of a healthier lifestyle. A limitation was the lack of extensive data, which required the RL model to be trained entirely on a user simulator. However, the trial results show that the user simulator was aligned with the real world environment. It is expected that with more data, methods such as RL from human feedback [Ziegler *et al.*, 2019; Ouyang *et al.*, 2022] might be employed for further finetuning, replacing the handcrafted rewards with empirically learned value functions.

Ethical Statement

The aim of this study was to provide motivational recommendations towards healthy exercise. To this end, the safety of the workouts generated by the platform was appraised by a team of sports scientists. Approval for the live user trial included in this study was obtained from the ethics office at University College Dublin. Participants' data was anonymized and stored securely.

Acknowledgments

This work was supported by the Science Foundation Ireland through the Insight Centre for Data Analytics under grant number SFI/12/RC/2289_P2 and by Samsung Electronics co. Ltd., Seoul, Republic of Korea. We would like to thank Rory Lambe, Ben O' Grady, Megan Burns and Jack Sargent for assisting in running the user trial.

References

[Abdulaziz *et al.*, 2021] Manal Abdulaziz, Bodor Almotairy, Mona Al-ghamdi, and Norah Al-qahtani. Building a personalized fitness recommendation application based on sequential information. *International Journal of Advanced Computer Science and Applications*, 12(1), 2021.

[Afsar *et al.*, 2022] M Mehdi Afsar, Trafford Crump, and Behrouz Far. Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys*, 55(7):1–38, 2022.

[Azizzadenesheli *et al.*, 2018] Kamyar Azizzadenesheli, Yisong Yue, and Animashree Anandkumar. Policy gradient in partially observable environments: Approximation and convergence. *arXiv preprint arXiv:1810.07900*, 2018.

[Barthélémy *et al.*, 2016] Jean-Claude Barthélémy, Sophie Berthouze-Aranda, Xavier Bigard, Cloes Marc, Damien Davenne, Paul Delamarche, Pascale Duche, Martine M Duclos, Guinhougouya Benjamin, Marie Martine Lefèvre-Colau, et al. *Actualisation des repères du PNNS–Révisions des repères relatifs à l'activité physique et à la sédentarité*. PhD thesis, Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement ..., 2016.

[Bassuk and Manson, 2005] Shari S Bassuk and JoAnn E Manson. Epidemiological evidence for the role of physical activity in reducing risk of type 2 diabetes and cardiovascular disease. *Journal of applied physiology*, 2005.

[Berndsen *et al.*, 2019] Jakim Berndsen, Barry Smyth, and Aonghus Lawlor. Pace my race: recommendations for marathon running. In *Proceedings of the 13th ACM conference on recommender systems*, pages 246–250, 2019.

[Berndsen *et al.*, 2020] Jakim Berndsen, Barry Smyth, and Aonghus Lawlor. A collaborative filtering approach to successfully completing the marathon. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 653–658. IEEE, 2020.

[Bonafiglia *et al.*, 2016] Jacob T Bonafiglia, Mario P Rotundo, Jonathan P Whittall, Trisha D Scribbans, Ryan B Graham, and Brendon J Gurd. Inter-individual variability in the adaptive responses to endurance and sprint interval training: a randomized crossover study. *PloS one*, 11(12):e0167790, 2016.

[Cf, 2015] ODDS Cf. Transforming our world: the 2030 agenda for sustainable development. *United Nations: New York, NY, USA*, 2015.

[CoEU, 2013] CoEU. Council recommendation of 26 november 2013 on promoting health-enhancing physical activity across sectors. *Council of the European Union, Off. J. Eur. Union*, 56:1–5, 2013.

[Dharia *et al.*, 2018] Saumil Dharia, Magdalini Eirinaki, Vijesh Jain, Jvalant Patel, Iraklis Varlamis, Jainikkumar Vora, and Rizen Yamauchi. Social recommendations for personalized fitness assistance. *Personal and Ubiquitous Computing*, 22:245–257, 2018.

[DoH, 2016] DoH. *Healthy Ireland: Get Ireland Active. National Physical Activity Plan for Ireland*. Department of Health, Government of Ireland Dublin, Ireland, 2016.

[Fernández-Martínez *et al.*, 2020] Antonio Fernández-Martínez, Víctor Murillo-Lorente, Alberto Sarmiento, Javier Álvarez-Medina, and Alberto Nuviala. Exercise addiction and satisfaction of fitness center users as precursors to the intention of continuing to engage in physical activity. *Sustainability*, 13(1):129, 2020.

[Fitness, 2023] Apple Fitness. Apple fitness plus. https://www.apple.com/ie/apple-fitness-plus/, 2023. Accessed: 2023-05-16.

[FORM, 2008] IN FORM. German national initiative to promote healthy diets and physical activity. *The National Action Plan for the Prevention of Poor Dietary Habits, Lack of Physical Activity, Overweight and Related Diseases. Federal Ministry of Food, Agriculture and Consumer Protection & Federal Ministry of Health, Berlin*, 2008.

[Freelitics, 2023] Freelitics. Freelitics. https://www.freelitics.com/, 2023. Accessed: 2023-05-16.

[Guo *et al.*, 2017] Xiaonan Guo, Jian Liu, and Yingying Chen. Fitcoach: Virtual fitness coach empowered by wearable mobile devices. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pages 1–9. IEEE, 2017.

[Gymfitty, 2023] Gymfitty. Gymfitty. https://gymfitty.com/, 2023. Accessed: 2023-05-16.

[Hardin and Hilbe, 2012] James W Hardin and Joseph M Hilbe. *Generalized estimating equations*. CRC press, 2012.

[Hoffman and others, 2011] Jay Hoffman et al. *NSCA's guide to program design*. Human Kinetics, 2011.

[Hutto and Gilbert, 2014] Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225, 2014.

[Ie *et al.*, 2019a] Eugene Ie, Chih-wei Hsu, Martin Mladenov, Vihan Jain, Sanmit Narvekar, Jing Wang, Rui Wu, and Craig Boutilier. Recsim: A configurable simulation platform for recommender systems. *arXiv preprint arXiv:1909.04847*, 2019.

[Ie *et al.*, 2019b] Eugene Ie, Vihan Jain, Jing Wang, Sanmit Narvekar, Ritesh Agarwal, Rui Wu, Heng-Tze Cheng, Morgane Lustman, Vince Gatto, Paul Covington, et al. Reinforcement learning for slate-based recommender systems: A tractable decomposition and practical methodology. *arXiv preprint arXiv:1905.12767*, 2019.

[Ifivolve, 2023] Ifivolve. Ifivolve. https://www.ifivolve.com/, 2023. Accessed: 2023-05-16.

[Levy *et al.*, 2021] Jacob J Levy, Terrance L Tarver, and Hannah R Douglas. Examining the impact of gym closures due to the covid-19 pandemic on combat sport athletes' mental health. *Journal of Clinical Sport Psychology*, 15(4):289–305, 2021.

[Liang *et al.*, 2017] Eric Liang, Richard Liaw, Robert Nishihara, Philipp Moritz, Roy Fox, Joseph Gonzalez, Ken Goldberg, and Ion Stoica. Ray rllib: A composable and scalable reinforcement learning library. *arXiv preprint arXiv:1712.09381*, 85, 2017.

[Liebman *et al.*, 2014] Elad Liebman, Maytal Saar-Tsechansky, and Peter Stone. Dj-mc: A reinforcement-learning agent for music playlist recommendation. *arXiv preprint arXiv:1401.1880*, 2014.

[Lister *et al.*, 2014] Cameron Lister, Joshua H West, Ben Cannon, Tyler Sax, David Brodegard, et al. Just a fad? gamification in health and fitness apps. *JMIR serious games*, 2(2):e3413, 2014.

[Mahyari and Pirolli, 2021] Arash Mahyari and Peter Pirolli. Physical exercise recommendation and success prediction using interconnected recurrent neural networks. In *2021 IEEE International Conference on Digital Health (ICDH)*, pages 148–153. IEEE, 2021.

[Maugeri *et al.*, 2020] Grazia Maugeri, Paola Castrogiovanni, Giuseppe Battaglia, Roberto Pippi, Velia D'Agata, Antonio Palma, Michelino Di Rosa, and Giuseppe Musumeci. The impact of physical activity on psychological health during covid-19 pandemic in italy. *Heliyon*, 6(6):e04315, 2020.

[Mehrotra *et al.*, 2019] Rishabh Mehrotra, Mounia Lalmas, Doug Kenney, Thomas Lim-Meng, and Golli Hashemian. Jointly leveraging intent and interaction signals to predict user satisfaction with slate recommendations. In *The World Wide Web Conference*, pages 1256–1267, 2019.

[Mikkelsen *et al.*, 2017] Kathleen Mikkelsen, Lily Stojanovska, Momir Polenakovic, Marijan Bosevski, and Vasso Apostolopoulos. Exercise and mental health. *Maturitas*, 106:48–56, 2017.

[Mullen *et al.*, 2011] Sean P Mullen, Erin A Olson, Siobhan M Phillips, Amanda N Szabo, Thomas R Wójcicki, Emily L Mailey, Neha P Gothe, Jason T Fanning, Arthur F Kramer, and Edward McAuley. Measuring enjoyment of physical activity in older adults: invariance of the physical activity enjoyment scale (paces) across groups and time. *International Journal of Behavioral Nutrition and Physical Activity*, 8:1–9, 2011.

[Nyenhuis *et al.*, 2020] Sharmilee M Nyenhuis, Justin Greiwe, Joanna S Zeiger, Anil Nanda, and Andrew Cooke. Exercise and fitness in the age of social distancing during the covid-19 pandemic. *The Journal of Allergy and Clinical Immunology: In Practice*, 8(7):2152–2155, 2020.

[Ouyang *et al.*, 2022] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

[Poli *et al.*, 2007] Riccardo Poli, James Kennedy, and Tim Blackwell. Particle swarm optimization: An overview. *Swarm intelligence*, 1:33–57, 2007.

[Sami *et al.*, 2008] Ashkan Sami, Ryoichi Nagatomi, Masahiro Terabe, and Kazuo Hashimoto. Design of physical activity recommendation system. In *IADIS European Conf. Data Mining*, pages 148–152, 2008.

[Schulman *et al.*, 2017] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[Shi *et al.*, 2019] Bichen Shi, Makbule Gulcin Ozsoy, Neil Hurley, Barry Smyth, Elias Z Tragos, James Geraci, and Aonghus Lawlor. Pyrecgym: A reinforcement learning gym for recommender systems. In *Proceedings of the 13th ACM Conference on Recommender Systems*, pages 491–495, 2019.

[Smyth, 2019] Barry Smyth. Recommender systems: A healthy obsession. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 9790–9794, 2019.

[Tran *et al.*, 2018] Tin Trung Tran, Jae Won Choi, Chien Van Dang, Geon SuPark, Jun Young Baek, and Jong Wook Kim. Recommender system with artificial intelligence for fitness assistance system. In *2018 15th International Conference on Ubiquitous Robots (UR)*, pages 489–492. IEEE, 2018.

[Venkatachalam and Ray, 2022] Parvathy Venkatachalam and Sanjog Ray. How do context-aware artificial intelligence algorithms used in fitness recommender systems? a literature review and research agenda. *International Journal of Information Management Data Insights*, 2(2):100139, 2022.

[WHO, 2022] WHO. *Global status report on physical inactivity 2022*. World Health Organisation, 2022.

[Ziegler *et al.*, 2019] Daniel M Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*, 2019.