Modelling the Training Practices of Recreational Marathon Runners to Make Personalised Training Recommendations

Ciara Feely*
ciara.feely@ucdconnect.ie
SFI Centre for Research Training in Machine Learning, University College Dublin
Ireland

Brian Caulfield
Insight Centre for Data Analytics, University College Dublin
Ireland
b.caulfiled@ucd.ie

Aonghus Lawlor
Insight Centre for Data Analytics, University College Dublin
Ireland
aonghus.lawlor@ucd.ie

Barry Smyth
Insight Centre for Data Analytics, University College Dublin
Ireland
barry.smyth@ucd.ie

ABSTRACT

These days we have all become increasingly aware of the role that exercise plays in a healthy lifestyle. Activities such as cycling, triathlons, and running have become popular ways for people to keep fit and test their abilities. For recreational athletes there is no shortage of training advice or programmes to follow, yet most offer only one-size-fits-all, or minimally tailored guidance, which often leaves novices under-supported on their fitness journeys. In this work, we describe a case-based reasoning system to generate personalised training recommendations for marathon runners, based on their training histories and the training histories of similar runners with comparable race goals. The system harnesses the type of activity data that is routinely collected by smartwatches and apps like Strava. It uses *prefactual* explanations to suggest to runners how they may wish to adjust their training as their fitness goals evolve. We evaluate the approach using a large-scale dataset of more than 300,000 real-world runners and we show that it is feasible to generate tailored, personalised recommendations for up to 80% of these runners. Additionally, we show that the recommendations produced are realistic and reasonable for a runner to implement, as part of their training programme. These suggestions typically include a small number (3-5) of incremental training adaptations, such as a change in weekly distance, long-run distance, or mean training pace. We argue that by engaging runners in this type of dialog about their training progress and race goals, we can better support novice runners, as their training unfolds, which may help to keep runners motivated on their long journey to race day.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning algorithms**;
• **Applied computing** → **Health informatics**.

UMAP '23, June 26–29, 2023, Limassol, Cyprus

ACM ISBN 978-1-4503-9932-6/23/06.
https://doi.org/10.1145/3565472.3592952

KEYWORDS

marathon running, case-based reasoning, user-adaptated personalisation, explainable AI, recommender systems

ACM Reference Format:
Ciara Feely, Brian Caulfield, Aonghus Lawlor, and Barry Smyth. 2023. Modelling the Training Practices of Recreational Marathon Runners to Make Personalised Training Recommendations. In *UMAP '23: Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (UMAP '23), June 26–29, 2023, Limassol, Cyprus.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3565472.3592952

1 INTRODUCTION

The marathon [32] is an iconic test of endurance that attracts millions of recreational and elite runners every year. In this paper, we demonstrate how user modeling and personalisation (UMAP) techniques can support recreational athletes as they train for and compete in marathons. This is a noteworthy departure from the more conventional *online* targets of UMAP research – e.g. inferring product preferences from online purchases, or assessing knowledge and expertise by monitoring online learning and lesson engagement – but not an unprecedented one, and there is increasing interest in applying these ideas to the physical world [26, 47, 48]. Indeed the marathon has been proposed as an appealing target for personalisation research [38, 42] because: (a) it attracts highly-motivated participants, many of whom are novices in need of support during their training and preparation; and (b) because there are several distinct sub-tasks within marathon preparation that are well-suited to user modeling and personalisation techniques, from profiling a runner's fitness and personalising their training to recommending running routes, training partners, gear or races. In this work we focus on profiling fitness and personalising training.

Preparing for the marathon requires at least 12-16 weeks of dedicated training with a careful mix of sessions designed and timed to improve speed, fitness, strength, and endurance. Unfortunately, when training for a marathon, recreational runners have limited options. Few have access to the type of one-to-one coaching that club and elite athletes enjoy, and most find themselves following a one-size-fits-all training programme found online. At best such fixed programmes are tuned to a runner's (often naïve) goal-time aspirations and perhaps some limited training preferences (e.g. activity

days per week), but most runners find themselves to be on their own for monitoring their progress or adapting to training disruptions – e.g.injury, busy weeks, or time away. Recently, developers of wrist-worn sensors such as Garmin [1] have begun incorporating personalised training into their devices which provides an ideal way for the techniques described in this work to reach a wider audience of runners. The system described in this paper is designed primarily with novice runners in mind, who are inexperienced marathoners. It uses activity data that runners routinely collect as they exercise with smartwatches and mobile devices. This data is used to recommend training adjustments to a runner, by using ideas from case-based reasoning (CBR) [10] to harness the training practices of similar runners.

Briefly, a runner is *profiled* using their prior training history. Each runner is represented by a set of *weekly training cases* covering every week of training. Each weekly case encodes a number of important distance and pacing features to reflect the current week of training, and the marathon time that the runner went on to achieve. Our approach uses CBR techniques in two important ways: (i) to assess a runner's current fitness, by predicting their likely race-time based on their training so far [21, 23, 40, 41]; and (ii) to generate personalised training suggestions based on a runner's training progress [20]. For (ii) – which is the primary unique contribution of this work – we use ideas from *prefactual* reasoning [6, 13, 18] to engage the runner in a dialog about how they might adjust their training to achieve a more or less ambitious goal. For example, the system might suggest the following: *"Your current training predicts a goal time of 244 minutes. If you wish to target a sub-4 hour marathon, then, based on similar runners to you, you should increase your long-run distance to 25km over the next two weeks and target a fastest 10km pace of 4:45 mins/km"*. If the runner accepts this suggestion then future training sessions will be adjusted accordingly, based on the corresponding sessions of the most similar runners who have informed this prefactual suggestion. In this paper we describe how to achieve this and present the results of a retrospective evaluation of the approach, using real-world data from approximately 300,000 recreational Strava runners, who trained for 500,000 unique marathons between 2014-2017.

2 RELATED WORK

The *fitness data revolution* began with the introduction of the first wireless heartrate monitor by Polar[2] in 1982 [30] but only became mainstream decades later, when companies like Fitbit brought cheap, wearable sensors to the masses. Today millions of people use these devices to track their daily activities [7, 14] with apps like Strava[3] and RunKeeper[4] to share their progress with friends. Recently, companies such as Whoop[5] have begun to integrate data about sleep, recovery, and exercise to help users to train more effectively and to recover more quickly. Services such as Training Peaks [6] provide their users with functional estimates of their fitness to guide their training efforts and monitor their recovery. Strava's users have access to information about their race-readiness and can be recommended recovery activities to help keep them healthy as they increase their training efforts in preparation for a big event. This explosion in the adoption of fitness devices and apps created new market opportunities, but current offerings still only scratch the surface of what may be possible. Attracted by the availability of data and a motivated user-base, machine learning and recommender systems researchers have begun to focus their interest in areas such as fitness assessment, training load estimation, recovery guidance, personalised training recommendations, performance prediction and race planning [36].

2.1 Estimating Fitness Metrics

Sports scientists use a variety of important laboratory metrics to estimate the fitness levels of individuals to analyse how they change under various training conditions and how they impact performance. For example, the well-known $V0_2max$ score is an important determinant of endurance capacity during prolonged exercise, but in the past has had to be measured in a laboratory setting. Recently, however, there have been efforts to use machine learning techniques to predict $V0_2max$ without relying on laboratory data [2]. Similar ideas have been recently applied to other important fitness metrics such as as a runner's *lactate threshold*; the pace at which a runner can no longer clear lactic acid from her muscles, which will quickly impact running performance [8, 19]. These estimation problems can be framed as classical supervised learning tasks and the resulting models may transform the effectiveness of training programmes by providing personalised advice and tailored recommendations about how an athlete should train on a given day in terms of their target pace, duration, and effort. Often training paces are determined with respect to these pacing thresholds – a runner's $V0_2max$ pace for fast sessions, their lactate threshold pace for so-called *tempo* or *threshold* sessions, or their *aerobic threshold* pace for easy runs – but many novice runners may not know how to calculate these paces for themselves and will often end up training in a sub-optimal manner.

2.2 Performance Prediction

Predicting an athlete's performance has been a staple of sports science for many years. In running there is a wealth of approaches to predicting finish-times across a range of distances, using a variety of data, and applied to different types of runners [28]. The advent of large quantities of activity data has helped to validate many of these conventional approaches using larger numbers of runners, and has also helped to to generate new models for performance prediction [17, 21, 40]. Indeed, machine learning techniques have now been used to successfully predict performance in sports such as cycling [29], tennis [35], soccer [1], Australian football, [9], and even archery [45].

Understanding how an athlete is likely to perform is not only useful as a way to set expectations, but also as the basis for planning an athlete's training and race-day strategy. For example, as mentioned above, when training for an endurance event such as the marathon, it is important to have a realistic estimate of an athlete's goal time (or their *marathon pace* because many training sessions will be calibrated with respect to this time or pace; e.g. *"run a 5km*

[1] http:\www.garmin.com
[2] http:\www.polar.com
[3] http://www.strava.com
[4] http://www.runkeeper.com
[5] http://www.whoop.com
[6] http://www.trainingpeaks.com

session 10-15 seconds faster than marathon pace." Moreover, when it comes to race day, it will be important for a marathon runner to have a goal time in mind so that they can plan their pacing over the long distance of the race; in the absence of a sensible race-day plan, they risk starting too fast [11, 12, 37] and ruining their race if they hit the wall [3–5, 39].

2.3 Towards Personalised Training

In this work, we want to generate personalised training programmes that are capable of adapting to the progress of an athlete as they train and compete. A marathon programme, for example, will typically consist of 12-16 weeks of training, organised in 3-4 week blocks of activities. As training progresses, runners will complete sessions that are designed to promote physiological adaptations, to improve endurance, strength, speed, running economy, fuel consumption etc., while trying to ensure that runners balance their training and rest. It can be challenging for an inexperienced marathoner to find a training programme that suits their personal circumstances and goals, and many are left struggling to follow a *one-size-fits-all* programme that is illsuited to their needs. This is why the idea of a more personalized training programme – or an intelligent virtual coach – is so appealing.

To date, several virtual coaching solutions surfaced to help guide runners and other athletes. For example, the work of [31] presents a running assistant to help runners during *heartrate zone* training, where they need to maintain a certain heartrate during each session. The work of [44] describes a technique for generating personalised interval training sessions, which may be useful as part of an overall training programme. Other researchers have used techniques such as *reinforcement learning* and *particle swarm optimisation* to generate and adapt personalised training plans [24, 25].

In this work we propose a case-based reasoning approach to generating personalised training recommendations. We adopt this approach because the *local*, *lazy*, *interpretable* nature of CBR is well suited to the task at hand:

(1) CBR methods have proven to be useful when it comes to predicting marathon performance [21, 23, 40]. One reason for this is the local reasoning used by CBR techniques for prediction. A prediction for a target runner is based on maximally similar runners (local neighbours) in the feature space which is appropriate given the noisy nature of running data; runners don't always have a good training/race day, raw activity data can include GPS errors leading to outlier paces, runners may engage in sessions that don't reflect their ability etc.
(2) CBR methods are lazy: they retain and rely upon the raw instances used for reasoning. Here, a case is a representation of a runner for a given week of training in terms of various distance and pacing features. Advantageously, this allows us to add new cases to the system, without the need to re-train a complex model.
(3) Another advantage is that CBR approaches are more interpretable than the "black-box" style reasoning used by some other ML approaches. This is especially important in the context of making personalised training recommendations to runners. Recommendations need to be understandable if runners are to trust them [34], and CBR methods have proven to be well suited to producing different styles of explanations as recommendations, including *factual* [16], *counterfactual* [27], and *prefactual* [13] explanations.

In this work we focus on generating *prefactual* explanations to encourage the runner to consider what might be achievable in the future if certain adjustments are made to their current training, based on their goals and progress to date. It is motivated by recent research on how prefactual thinking may motivate marathoners to train harder [43].

3 A CASE-BASED REASONING SYSTEM TO RECOMMEND MARATHON TRAINING PLANS

In this section we describe the technical details of our case-based approach to personalised training recommendation. To understand one important use-case, imagine a runner, Lucy, who began training for her first marathon using a training programme based on a goal-time of between 3.75 and 4 hours as she tried to break the 4-hour barrier. Lucy is now 6 weeks from race-day and her training has gone reasonably well, but she has missed a couple of recent sessions. The system predicts that Lucy's marathon time is likely to be about 4 hours and 5 minutes, just shy of her desired goal time, and recommends that if Lucy wants to get back on track for a sub-4 hour finish then she should increase her weekly training volume by about 10km and aim to run her fastest 10km during training at approximately 4:45 mins/km.

There are a few points to make about this approach:

(1) First, the system predicts Lucy's current marathon time, based on her training to date.
(2) Second, since the predicted time does not agree with Lucy's goal time, the system suggests a training adaptation.
(3) Third, the system does not prescribe a specific training session or set of sessions per se, but rather recommends certain goals for the weeks ahead (e.g. distance and pacing goals). This allows runners to adapt their training within the context of their existing training programme whatever that may be.

A related use-case allows Lucy to specify how she wants to adjust her goal to receive a suitable training adjustment. For example, in the above scenario, perhaps Lucy realises that her goal-time aspirations are unrealistic – perhaps she is feeling exhausted from her training so far – and wants to consider the implications of targeting a 4 hour and 15 minute goal-time (a 5% slowdown from her current predicted marathon time). In this case, the CBR system generates a prefactual explanation based on this slower finish-time; for example, Lucy might be recommended to maintain her weekly volume, but to ease off on her effort, perhaps by slowing her fastest 10km pace 5:30 mins/km.

In the sections that follow we describe how to transform activity data into cases that the CBR system can then use to predict marathon times for runners based on their training so far. We further describe how to suggest training adjustments to runners

who wish to modify their predicted time, as per the pseudo-code presented in Algorithm 1

3.1 From Raw Activity Data to Training Cases

The raw data provided by Strava is made up of individual activities for users. Each activity is associated with a number of time series capturing information about distance, time, and elevation. For the purpose of this work we convert the distance (m) and time (secs) information into pace (mins/km) and use the elevation data (m) to produce a grade-adjusted pacing value [33] (GAP). Grade-adjusted pace is a pacing value (mins/km) that has been adjusted for elevation/gradient. Uphill paces are adjusted to be faster, while downhill paces are adjusted to be slower, so that it is easier to compare paces independently of gradient.

Since runners use different devices to track their activities, the raw Strava data contains time, distance, and elevation data samples at different rates. We transform this raw data into smoothed 100m averages. Thus, each GAP value represents the grade-adjusted pace for a 100m segment so that an activity with a time series comprising 100 GAP values corresponds to a 10km activity.

To produce a training representation that is suitable for CBR, for each runner r and week w of training we extract a number of features for the set of activities for week w, including:

(1) $TotDist$ - Total weekly distance (km).
(2) $MaxDist$ - Distance of the longest activity (km).
(3) $NumRest$ - Total number of rest days.
(4) $MaxBreak$ - The longest consecutive number of rest days.
(5) $MeanPace$ - Mean weekly pace (mins/km) overall all activities.
(6) $Fastest10kmPace$ - Fastest 10km pace (mins/km) in a single activity in week w.

These features were chosen because they are meaningful training indicators for marathon runners and are often used in training programmes. Other than the $Fastest10kmPace$, these features have been included in a meta-regression [15]. Additionally, we also calculate the cumulative mean values for the above features for all training weeks up to and including week w and also the cumulative "best" (longest/fastest) values for the above features up to and including week w. Thus, $CumMeanTotDist(r, w)$ refers to the average total weekly distance for r for all training weeks up to and including week w and $CumBestFastest10kmPace(r, w)$ corresponds to the fastest 10km pace so far seen for r during training weeks up to an including week w. Thus, these sets of weekly and cumulative features correspond to a feature-based representation for a given week of training ($F(r, w)$). This representation will be used to find similar runners to r given the training complete by week w.

It is worth noting that this list of features is far from complete. While we do further split the case-base by sex, few runners track their age and weight which were excluded (although previous experimentation of the authors found these to be insignificant features in the model). Additionally, heartrate can provide a useful indicator of training intensity, however since the data accessed in this work was tracked between 2014-2017, many runners do not have heartrate data. Finally, since we don't have any access to the runners themselves, we don't have any input from them about the types of sessions they were completing, their physiological fitness metrics, or whether breaks in their training were due to injury, a lack of activity, or simply not wearing their tracker. To mitigate this last point, we restrict the data to those who have tracked at least 8 weeks of training in the 16 weeks before race-day.

To recommend training adjustments in the weeks that follow week w, we will need the case representation to also contain a pointer to the training features for $w + 1$, as well as the cumulative training features that occur in all of the subsequent training weeks up to race-day, denoted $F'(r, w + 1)$.

Then, the case representation for a runner r in week w is made up the *current week's* training features and cumulative features to date ($F(r, w)$), plus the *next week's* training features and the cumulative training features in the subsequent period from week $w + 1$ to race-day ($F'(r, w + 1)$), and the runner's marathon time (MT) as shown in Equation 1.

$$C(r, w) = (F(r, w), F'(r, w + 1), MT) \tag{1}$$

In this way, for every runner in our dataset we produce a set of weekly training cases which represent their training up to this point, and after this point in training and the marathon finish-time that this produced. In what follows we describe how we use these cases to: (i) predict marathon times for new runners; and (ii) to suggest training adaptations for runners, based on their predicted or target times.

3.2 Race-Time Prediction

To generate a predicted marathon finish-time ($P(r_q, w)$) for a (novice) runner r_q in week w, we adopt the same approach as demonstrated and evaluated in [23]. Specifically, we use $F(r_q, w)$, r_q's current training week, to identify k other cases with week w features that are maximally similar to $F(r_q, w)$ and compute the average of their marathon finish-times; see line 3 in Algorithm 1. For this we use a standard Euclidean distance metric, to compare the features of the query and candidate cases; all feature values are min-max scaled to facilitate matching. We also, filter cases, prior to similarity, based on the sex of the runner because the physiology of male and females runners is materially different.

These predicted finish-times are useful for runners in several ways. Firstly, as they approach race-day the predicted finish-times can help runners to plan for a realistic goal-time and to determine their pacing accordingly. Secondly, during training, these predictions provide a useful evaluation of how training is progressing. If the weekly finish-time predictions remain stable, or are improving, then training is going well. If the predicted finish-times are disimproving, then training is not going well. Finally, these predictions offer runners a way to compare their training progress against their planned goal-time, which we use as a stepping stone for training adaptations as discussed next.

3.3 Generating Prefactual Training Recommendations

If a runner's predicted marathon time is slower than their goal-time then the runner has a choice. They can continue as they have been, accepting the slower finish-time and adjusting their race-day expectations accordingly. Or they can explore what training

adjustments might be required to improve their predicted time and get closer to their goal-time. This is a relatively novel concept in marathon training, especially for most runners who are left following a one-size-fits-all training programme. Often they have no real way to evaluate their training progress or to translate this into a predicted finish-time – although some smart-watches will at least do this – and unless they are an experienced runner they will have little ability to adjust their training in such as way as to improve their prospects.

The novel contribution of this work is to provide runners with a solution to this problem. For example, consider a runner r_q who wishes to improve her predicted time $P(r_q, w)$ by a factor of δ. To do this we generate two sets of cases that are similar to $F(r_q, w)$, as follows:

(1) *Factual Cases* (C_f): the top k most similar cases to $F(r_q, w)$ with a marathon time that is $P(r_q, w)$ or *slower*; line 5 in Algorithm 1.
(2) *Prefactual Cases* (C_p): the top k most similar cases to $F(r_q, w)$ with a marathon time equal to $P(r_q, w) * (1 + \delta)$ or *faster*; line 6 in Algorithm 1.

Thus, C_f corresponds to runners who are similar to r_q and with a marathon time that is no faster that r_q's current predicted time, whereas C_p correspond to runners who are similar to r_q, but with a marathon time that is a factor of δ *faster* than r_q's predicted time. Because of their faster finish-times, we should expect the training features associated with the C_p to be different from the training features associated with the C_f. Moreover, these differences, if significant, may serve as a useful way to explain to r_q how the faster runners are training compared to runners like r_q.

With this in mind, we use a two-sided t-test to compare the mean values of each of the features in the next week of training and the cumulative training features that follow in the weeks leading up to race-day ($F'(r_i, w + 1)$) for the cases in C_f and C_p; see lines 11-14 in Algorithm 1. Then, any features associated with statistically different feature values ($p < 0.1$) form the basis for a training adjustment recommendation for r_q, such as:

> *Runners who achieved a marathon time that is at least 3% faster than your projected marathon time (245 mins) completed a greater total distance (55km) and ran a quicker fastest 10k pace (5:10 mins/km) for next week's training compared with runners who complete a marathon with a similar time to your projected time (total distance 50km, fastest 10k pace of 5:20mins/km).*

Later, in Section 5, we will further discuss how this type of information can be used to suggest training adjustments to runners. Additionally, it should be noted that an analogous approach can be applied in the case where the runner decides to target a *slower* finish-time, for example if she feels that the training programme she is following is too challenging, but for now we will focus on the use-case where a runner is optimistic and wants to improve their time.

3.4 Discussion

In this section we have discussed how raw activity data can be transformed into weekly marathon training cases for use in a CBR system to predict marathon finish-times during training, and to recommend training adjustments based on these predictions. While predicting marathon finish-times is not novel per se – similar ideas have been presented by others [23] – recommending training adjustments is novel and valuable. Moreover, the approach, as described, does not rely on any detailed representation of a target runner's specific marathon programme, only a feature based summary of their training so far. And the recommendations produced are not for specific sessions, which may or may not be compatible with a runner's recent training or programme, but rather a prescription about how a runner may want to adjust their training week (more/less distance, slower/faster pace).

Algorithm 1 Generating Race-Time Predictions and Prefactual Training Recommendations

Input: q, the query case for runner **r**; **CB**, the case base;
Parameters: **w**, the week in training; δ the factor difference between a runner's predicted and goal finish-times; **k**, the number of factual and prefactual cases to retrieve; **p**, the level of significance.
Output: **P** the predicted marathon-time C_f, a set of Factual cases; C_p, a set of Prefactual cases; **sig**, features that differ significantly between the factual and prefactual cases.

1: $C \leftarrow filter(CB, week = q.week, sex = q.sex)$
2: $C' \leftarrow sort(C, sim(q, c))$
3: $P \leftarrow mean(C'.MT.head(k))$
4: **if** $\delta \leq 0$ **then**
5: $\quad C_f \leftarrow C'[C'.MT \geq P].head(k)$
6: $\quad C_p \leftarrow C'[C' \leq P * (1 + \delta)].head(k)$
7: **else**
8: $\quad C_f \leftarrow C'[C'.MT \leq P].head(k)$
9: $\quad C_p \leftarrow C'[C' \geq P * (1 + \delta)].head(k)$
10: **end if**
11: $sig \leftarrow []$
12: **for** f in $C'.F'$ **do**
13: $\quad sig.append(f)$ **if** $ttest(C_f.f, C_p.f) < p$
14: **end for**
15: **return** P, C_f, C_p, sig

4 EVALUATION

As an initial evaluation of our approach we will focus on an (offline) evaluation of the training adjustment recommendations; the effectiveness of finish-time prediction has been demonstrated previously [23], and in that work the authors reported error rates of 6-7% with $k = 15$. The present work implements an equivalent finish-time prediction model with similar error rates, but this is not reported in detail here for reasons of space.

4.1 Dataset and Methodology

For this evaluation we use a dataset of Strava activities logged between 2014-2017. The dataset was made available to the authors as part of a data sharing agreement with Strava Inc. The dataset included training activities up to 16 weeks before race-day for almost 60,000 female runners and over 233,000 male runners; summary statistics for this dataset, including the average age, marathon

Table 1: A summary of the dataset used in this work divided between male and female runners and showing: the number of unique runners; their mean age and finish-times; the mean number of marathons per runner completed in the period; the mean weekly training distance per week.

Sex	Unique Runners	Age	MT	No. Races	Dist/Wk
F	59118	38.51	264.23	1.57	40.56
M	233205	40.22	239.85	1.79	41.8

finish-time, and number of marathons completed is shown in Table 1.

To evaluate the recommended training adjustments we will focus on the following important metrics:

(1) *Coverage* - The percentage of query cases that can be associated with a training adjustment, for a given δ.
(2) *Number of Adjustments* - The number of training features that require modification for r_q.
(3) *Degree of Adjustments* - The average difference in training feature values between factual and prefactual cases.
(4) *Adjustment Consistency* - The extent to which the same features are repeatedly suggested for adjustment.

Ideally, our approach should deliver high coverage – it should be able to offer most runners a suggested training adjustment for a range of desired δ's – while at the same time limiting the number of training parameters that are changed. Additionally, runners will likely expect a certain degree of consistency in the adjustments that are recommended; for example, it would be strange to find very different features being recommended for adjustment from week to week.

For this evaluation we split the data into 90% training and 10% testing to perform a standard hold-out set validation. Each test case was used as a query against the training cases for different values of $-.2 \leq \delta \leq .2$; acknowledging that this is a wide range of δ values and it is unlikely that many runners will want to adjust their finish-time by as much as ± 20%. Nevertheless it is useful to understand how coverage and adjustments vary across this range. For these test queries and δ values we calculated the coverage (number of queries where a suggested training adjustment can be made), the mean number of feature adjustments and the scale of the adjustment. We report these results in the following sections.

4.2 Explanation Coverage

Figure 1(a) shows the coverage results for varying values of δ, comparing male and female runners, early (12 weeks from race-day) and late (6 weeks from race day) in their training. For realistic values of δ ($abs(\delta) \leq 0.10$) coverage is reasonably high (70-80%) but coverage declines for more extreme δ values, due to insufficient prefactual cases.

Interestingly, females usually enjoy greater levels of coverage than males when we control for δ and training stage (early or late). This may be due to differences in the proportion of slower and faster males and females, or it may indicate greater variation in the training patterns of females. It makes for an interesting avenue for further exploration.

Coverage is also marginally higher when runners seek training adjustments close to race day (6 weeks to race day) compared with when they seek training adjustments earlier in their training (12 weeks to race day). This difference is statistically significant based on a two proportion z-test with $p = 0.05$ and is consistent with the work of [21, 23] showing that marathon finish-time prediction tends to be more accurate closer to race-day, which likely improves the system's ability to find similar factual and prefactual cases.

4.3 A Feature Analysis of Explanations

Being able to make personalised suggestions for a majority of runners who wish to modify their training bodes well for this approach, but whether runners will respond well to these suggestions will likely depend on their ability to interpret them, and whether they view them as practical to implement. For example, it is reasonable to expect that runners will not want to have to implement large changes to many aspects of their training, and thus it is preferable for training adjustments to involve small numbers of features that need to change by modest amounts.

To explore this, in Figure 1(b) we present the average number of feature adjustments per recommendation, again by varying δ and based on runner sex and training stage. Once again the results are encouraging. For reasonable values of δ we see that, when recommendations can be made, they only involve 3-5 training feature adjustments (recall that there are 18 possible features that could be called upon, including the average and cumulative best and max versions of each feature). Females tend to be associated with more training adjustments than males, and there are marginally more adjustments recommended later in training compared with earlier in training. The difference between males and females may once again indicate that there there is greater training variety among females than males. While the difference between the number of training adjustments close to race-day (more) versus earlier in training (fewer) is modest, it is statistically significant (based one sided t-test with $p = 0.01$) for males and females. This difference could be due to the fact that early in training there is still plenty of time to make up training gradually, whereas closer to race-day, changes will need to be more extreme to have the desired effect.

It is also instructive to examine the degree of change that is associated with the different features that are recommended for adjustment. For example, a large change might be viewed as unrealistic by a runner: suggesting a runner should increase their weekly distance by more than 10% per week is ill-advised, for example. In Figure 2(a & b) we show how often different types of training features are recommended as part of an adjustment for males (a) and females (b) 6 weeks before race day and for a 5% improvement in marathon time ($\delta = -0.05$). Separately, Figure 2(c & d) show the average change for the different features for males (c) and females (d) 6 weeks before race day and for δ=-0.05. We can see that for males and females the cumulative maximum fastest 10km is the most common training adaptation, appearing in almost 40% of male and 60% of female adjustments for a 5% improvement in marathon time. The cumulative average fastest 10km also appears high in the recommendations. Recall that in the recommendations, the cumulative features represent the average and best training completed in the following weeks until race-day, in this case weeks

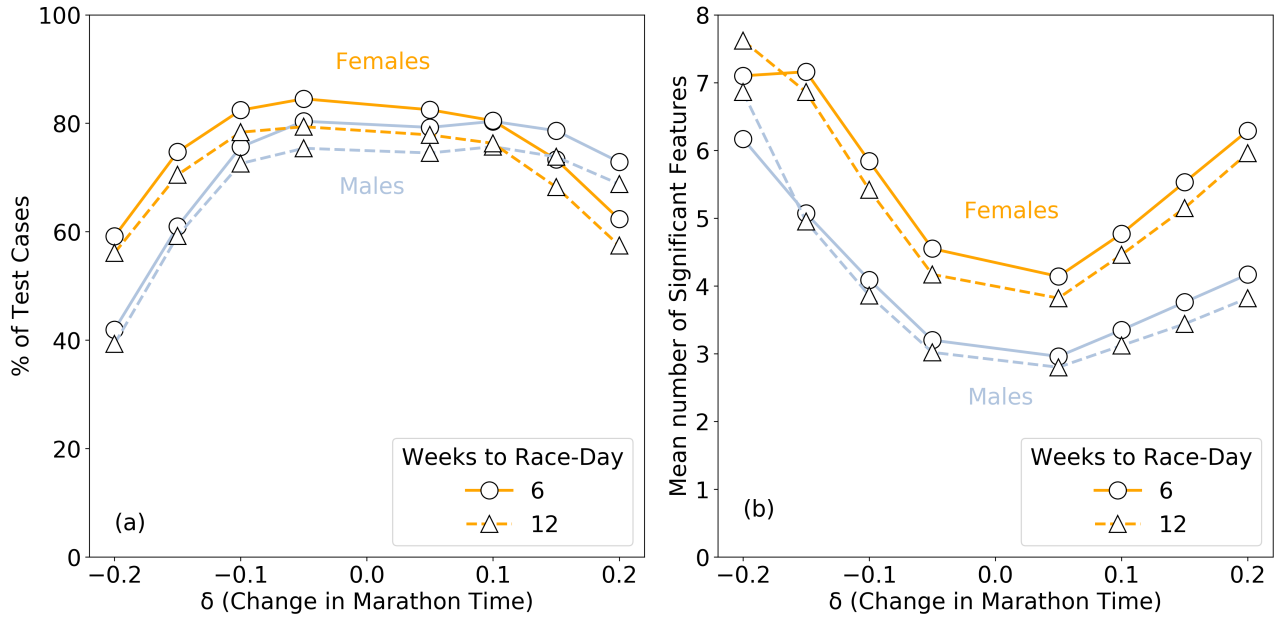

Figure 1: (a) The proportion of test cases for which we can make a training recommendation versus δ, by sex and training stage. (b) The average number of features that are significantly different ($p < 0.1$) among factual and prefactual cases for different values of δ and showing differences between sex and point in training.

1-6 before race-day. Overall the current week of training features rank lower than the cumulative training features, indicating that gradual training changes must be made to achieve faster times, rather than simply trying to up the training load or intensity for the next week alone. We can see that out of the next week of training features, the feature that is most commonly recommended is again the fastest 10km pace. This aligns with previous work [23] which showed that the fastest 10km pace was the most important feature for predicting marathon performance. For each of the mean pace, and fastest 10km pace features, the recommendation is to reduce the pace (run faster) by about 4-5% which is reasonable considering the goal to run the marathon 5% faster. For the distance-based features (total weekly distance, and longest distance per week), the advice is to increase the next week of training and average training distances by about 10% for males and by 10 and 12% respectively for females and to increase the best (longest) distances by about 13% for males and 15% for females. The recovery based features were recommended least commonly which aligns with previous work demonstrating that these features were not important indicators in a model for marathon performance, likely due to the fact that they correlate with total weekly distance [23]. The recommendation for these features is a slight increase or decrease for example reducing the number of rest days per week or max break by 5%, which is not a meaningful difference, compared to reducing a pace by 5% or increasing distance by 10-15%. More generally, we can see that the degree of change in training features in Figure 2(c & d) all fall within a reasonable range from -0.1 to +0.15, which means that runners need not change their training too drastically. The standard deviations vary for these values, for the current training week features they are more drastic indicating that there is larger variation in the current training week than in the overall training in subsequent weeks. This makes sense since runners will be varied in their approaches to training week to week but often these variations even out over a longer period.

Finally, in Figure 3, we show how often a given feature type is the most significant feature to adjust in a recommendation for various points in training (12, 9, 6, and 3 weeks before race day). Pacing features tend to be the most common feature adjustments followed by distance and recovery features. Moreover, there is good consistency across the training weeks meaning that runners will not be surprised to see certain features being recommended.

4.4 Discussion

We have described a preliminary evaluation of our approach to recommending training adjustments as runners train for a marathon and as their race expectations evolve. The evaluation is preliminary because it is a retrospective analysis rather than a live-user trial. This facilitates the use of a very large dataset of ($> 300,000$) runners to evaluate various aspects of the training recommendations, including coverage, number/degree of adjustments, and consistency. However, a live-user trial should be completed to examine how real runners respond to the recommendations, whether they find them useful, and whether they have the desired effect on performance. Such a user trial is planned in the future.

Notwithstanding the above, we have learned a lot about the type of training adjustments that can be made from this preliminary evaluation. Coverage is typically high, meaning that we can make suggestions for most runners in most situations. The recommendations are not overly complex – they typically involve minor adjustments to small numbers of features – and they appear to be

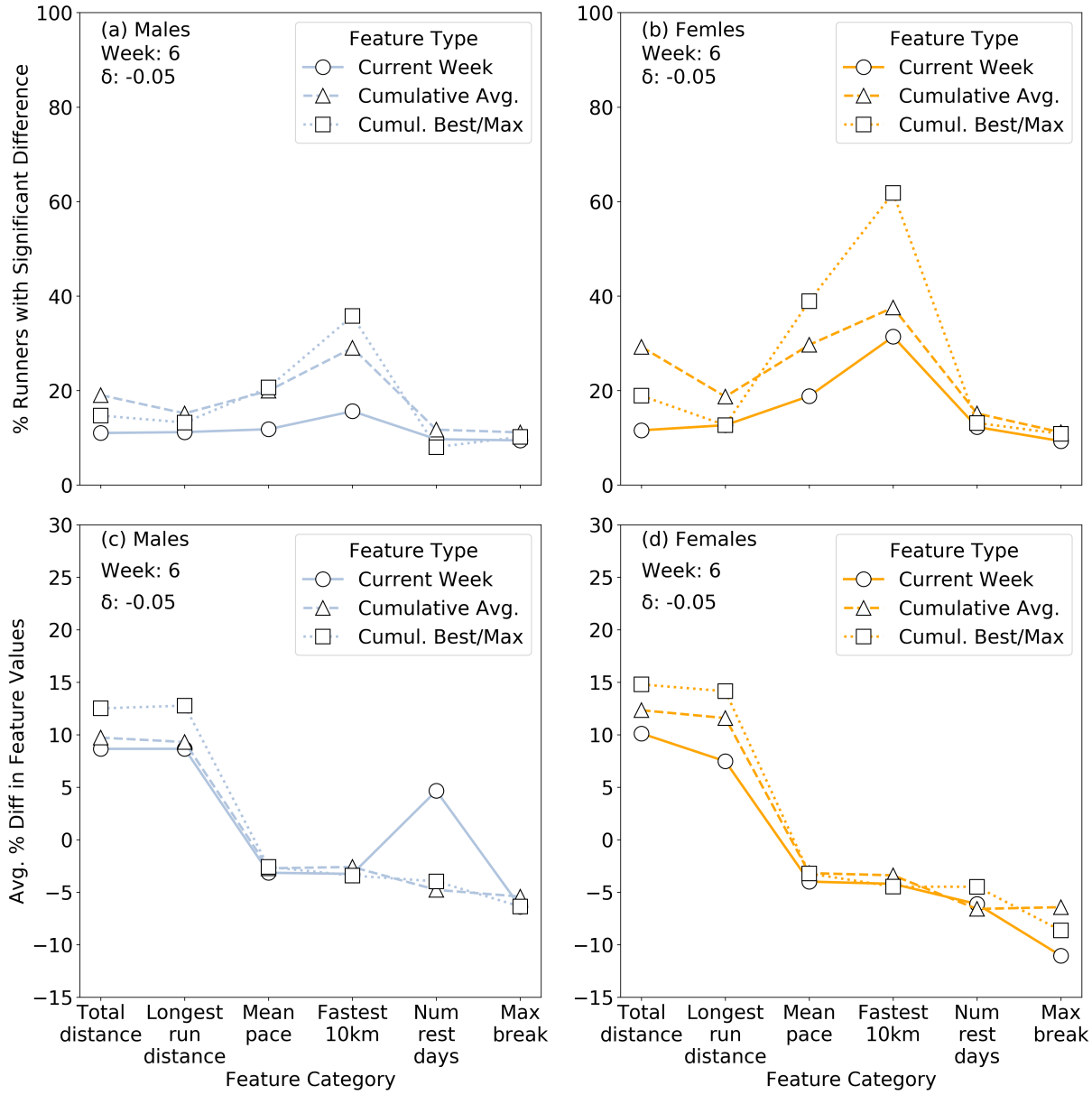

Figure 2: The proportion of test cases for which each feature is in the recommended training adjustments for (a) males and (b) females. The average percentage difference change recommended when comparing prefactual to factual cases for (c) males and (d) females.

.

reasonably straightforward to implement given that the degree of change required is modest.

5 AN EXAMPLE USER INTERFACE

The road to deployment for this system involves developing a Strava companion app using the Strava developer API. This would pair with Strava and pull in a user's uploaded activity data and require the runner to input a goal marathon time as well as a proposed date for their future marathon. From there, the CBR system would be capable of providing a weekly update to the runner about their progress in terms of their predicted marathon time, as well as recommend training adjustments to keep the runner on track for their goal. Developing a companion app was beyond the scope of this current work and we leave it for future work but for now, see Figure 4 for the current prototype. It is worth noting that the user-interface of this system could take many formats, and will involve extensive designing and testing with real runners. In this work we give one example of what the interface *could* look like.

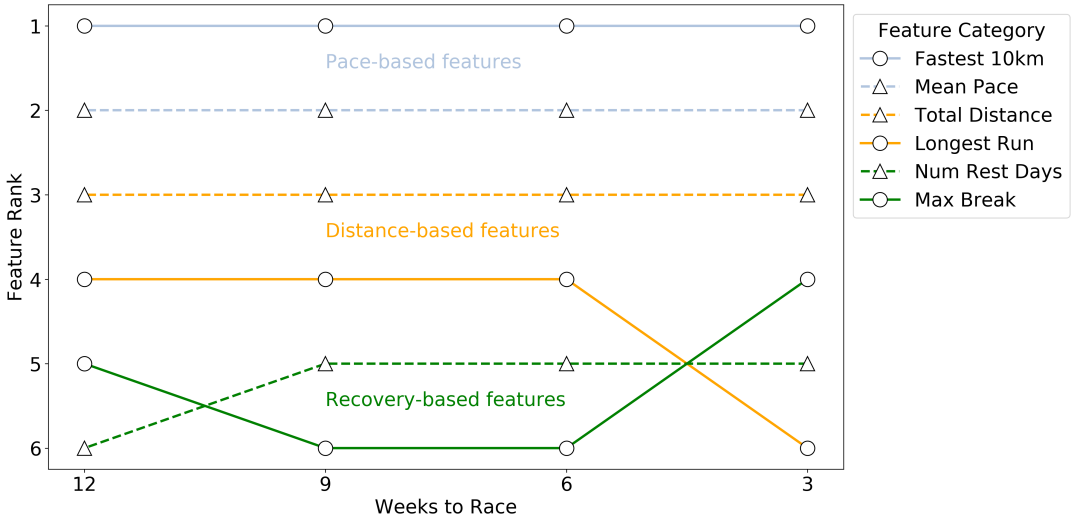

Figure 3: Each feature category ranked 3, 6, 9, and 12 weeks from race-day by how frequently the top most significant feature for a runner is in the feature category.

In the left-hand side image we can see a weekly progress report which provides a factual explanation of why the query runner Lucy is not currently meeting their goal – "You missed two sessions this week which reduced your overall mileage and slowed your predicted time". Lucy is then offered two training adjustments – to increase her planned distance by 10km, and to try to run a faster 10km at a pace of 5:10 mins/km. These recommended training adjustments are a form of prefactual explanation of Lucy's training since if she can do what is recommended, she will be predicted to achieve her goal marathon time, whereas if she does not make the changes she will stay at her current predicted time.

The right-hand side image contains a "peer analysis" to give the runners a sense of how their training fits into the the most similar training plans and indeed how this contributes to their predicted marathon time. The peer-set can offer somewhat of a confidence interval for the runner's future marathon time as well as a suitable range for the training features tracked by the runners who achieve a finish-time similar to their goal. This factual explanation may improve the user's trust of the system's ability to predict their time since they can see exactly how similar they are to their nearest neighbours driving their prediction: if the feature values provided are drastically different from their own then they can make an informed decision about how accurate the predicted race-time is likely to be for them.

What we would like to provide users with is a fully interactive interface whereby the sliders can be dragged to alter one of the training features of the next week of training and a new plan would be provided by finding the most similar cases. Similarly if our query runner Lucy wanted to learn about how she could alter her training to achieve a more challenging race-time then she could simply drag the lowest slider towards the right and a new set of features above would be provided. With a case-based reasoning approach, this level of user-adaptation is achievable by filtering the case-base according to a set of user-defined feature values.

Another key element to this system would be providing runners with a sense of the risk imposed by the training recommendations, since taking on more strenuous training could lead them to become injured. This could build upon previous work on using counterfactual explanations to understand injury risk [22], and domain knowledge of how acute-chronic workload ratio (training load measure) changes, indicating injury risk with the training adjustments [46].

6 CONCLUSIONS AND FUTURE WORK

This paper describes a novel case-based reasoning system for recommending training adjustments to marathon runners. The system predicts a runner's marathon time based on their current training to date, using the marathon times of similar previous marathoners. By leveraging prefactual reasoning, the system provides runners with a set of recommended training adjustments to help them meet their goal-time by comparing groups of runners who did and did not attain their goal-time.

The system was evaluated offline in terms of its ability to recommend sensible training adjustments using a real-world dataset of over 500,000 marathon training programmes, from nearly 300,000 unique marathoners, who tracked their activities in Strava between 2014-2017. We found that it was possible to make training recommendations for to up to 80% of runners when they target a conservative ambitious goal-time (5-10% faster than predicted time). The results indicate that if runners want to meet their goals they need to make changes by increasing their weekly distance and overall activity, and improve their pace, and keep these changes consistent to race-day.

As discussed, in order to fully test this recommender system and determine the efficacy of the recommendations, a live-user study is planned for future work. Several opportunities also exist for extending the system as described, for example by considering further training features such as average or maximum heart-rate.

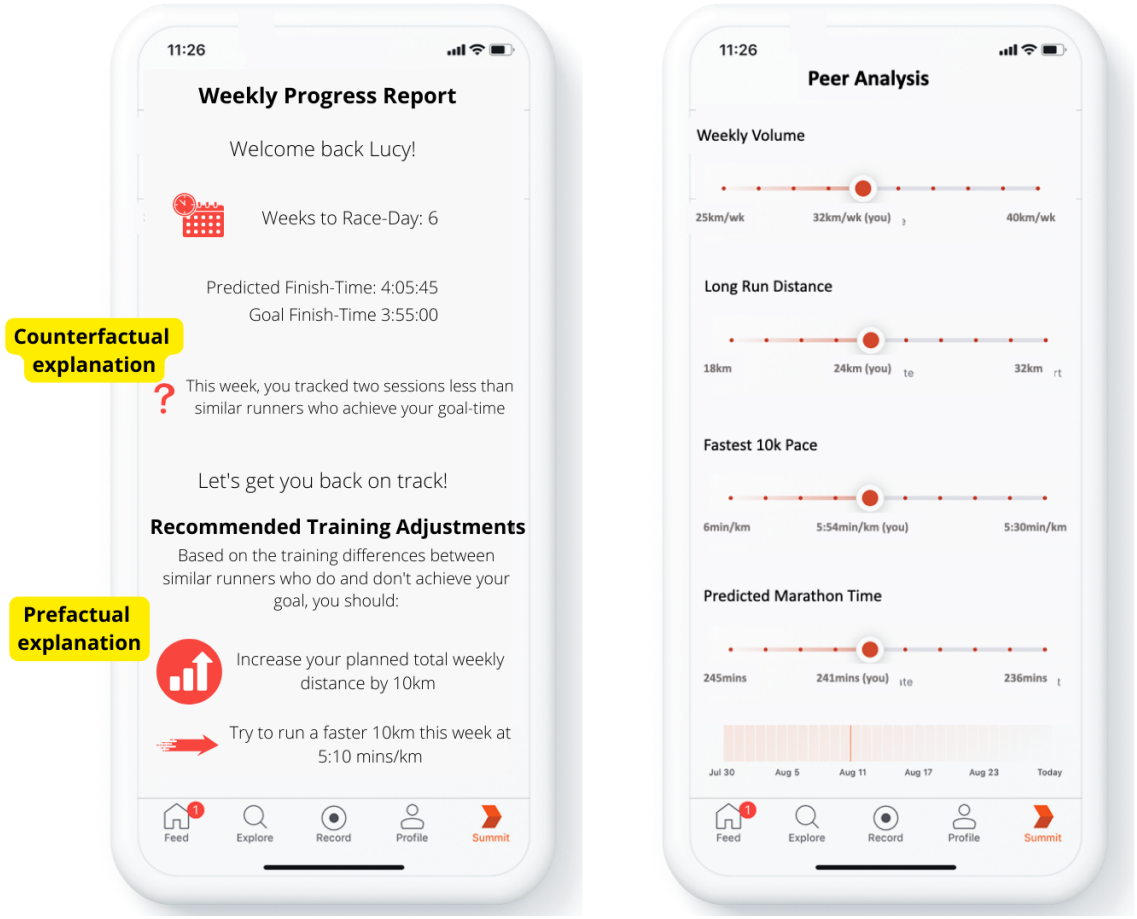

Figure 4: The current prototype implementing the proposed system in the form of the weekly training advice that could be provided to runners via a Strava companion app.

Additionally, ideas from previous work on understanding injury implications of marathon training could be incorporated by expressing how the recommended changes might alter the runner's injury likelihood [22, 46]. The proposed user interface needs to be fully developed and tested rigorously before deployment as a Strava companion app. A similar system could be applicable to races of other distances for example 10km races or ultramarathons, as well as other endurance sports.

6.1 Acknowledgements

Supported by Science Foundation Ireland through the Insight Centre for Data Analytics (12/RC/2289_P2) and the SFI Centre for Research Training in Machine Learning (18/CRT/6183).

REFERENCES

[1] Grant Abt and Ric Lovell. 2009. The use of individualized speed and intensity thresholds for determining the distance run at high-intensity in professional soccer. *Journal of sports sciences* 27, 9 (2009), 893–898.

[2] Atiqa Ashfaq, Neil Cronin, and Philipp Müller. 2022. Recent advances in machine learning for maximal oxygen uptake (VO2 max) prediction: A review. *Informatics in Medicine Unlocked* 28 (2022), 100863.

[3] Jakim Berndsen, Aonghus Lawlor, and Barry Smyth. 2020. Exploring the wall in marathon running. *Journal of Sports Analytics* 6 (06 2020), 1–14. https://doi.org/10.3233/JSA-200354

[4] Jakim Berndsen, Barry Smyth, and Aonghus Lawlor. 2019. Pace my race: recommendations for marathon running. In *Proceedings of the 13th ACM Conference on Recommender Systems*. ACM, ACM, New York, New York, USA, 246–250.

[5] Matthew P Buman, Britton W Brewer, Allen E Cornelius, Judy L Van Raalte, and Albert J Petitpas. 2008. Hitting the wall in the marathon: Phenomenological characteristics and associations with expectancy, gender, and running history. *Psychology of sport and exercise* 9, 2 (2008), 177–190.

[6] Ruth MJ Byrne and Suzanne M Egan. 2004. Counterfactual and prefactual conditionals. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 58, 2 (2004), 113.

[7] Ranganathan Chandrasekaran, Vipanchi Katthula, Evangelos Moustakas, et al. 2020. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. *Journal of medical Internet research* 22, 10 (2020), e22443.

[8] Alexander Chikov, Nikolay Egorov, Dmitry Medvedev, Svetlana Chikova, Evgeniy Pavlov, Pavel Drobintsev, Alexander Krasichkov, and Dmitry Kaplun. 2022. Determination of the athletes' anaerobic threshold using machine learning methods. *Biomedical Signal Processing and Control* 73 (2022), 103414.

[9] David Cornforth, Piers Campbell, Keith Nesbitt, Dean Robinson, and Herbert F Jelinek. 2015. Prediction of game performance in Australian football using heart rate variability measures. *International Journal of Signal and Imaging Systems Engineering* 8, 1-2 (2015), 80–88.

[10] Ramon López de Mántaras, David McSherry, Derek G. Bridge, David B. Leake, Barry Smyth, Susan Craw, Boi Faltings, Mary Lou Maher, Michael T. Cox, Kenneth D. Forbus, Mark T. Keane, Agnar Aamodt, and Ian D. Watson. 2005. Retrieval,

reuse, revision and retention in case-based reasoning. *Knowledge Eng. Review* 20, 3 (2005), 215–240. https://doi.org/10.1017/S0269888906000646

[11] Robert O Deaner, Vittorio Addona, Rickey E Carter, Michael J Joyner, and Sandra K Hunter. 2016. Fast men slow more than fast women in a 10 kilometer road race. *PeerJ* 4 (2016), e2235.

[12] Robert O Deaner, Rickey E Carter, Michael J Joyner, and Sandra K Hunter. 2014. Men are More Likely than Women to Slow in the Marathon. *Med Sci Sports Exerc* 4, 3 (2014), 607–616.

[13] Eoin Delaney, Derek Greene, Laurence Shalloo, Michael Lynch, and Mark T Keane. 2022. Forecasting for Sustainable Dairy Produce: Enhanced Long-Term, Milk-Supply Forecasting Using k-NN for Data Augmentation, with Prefactual Explanations for XAI. In *Case-Based Reasoning Research and Development: 30th International Conference, ICCBR 2022, Nancy, France, September 12–15, 2022, Proceedings*. Springer, Springer, Berlin, Heidelberg, 365–379.

[14] Keith M Diaz, David J Krupka, Melinda J Chang, James Peacock, Yao Ma, Jeff Goldsmith, Joseph E Schwartz, and Karina W Davidson. 2015. Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International journal of cardiology* 185 (2015), 138–140.

[15] Cailbhe Doherty, Alison Keogh, James Davenport, Aonghus Lawlor, Barry Smyth, and Brian Caulfield. 2020. An evaluation of the training determinants of marathon performance: A meta-analysis with meta-regression. *Journal of Science and Medicine in Sport* 23, 2 (2020), 182–188.

[16] Dónal Doyle, Alexey Tsymbal, and Pádraig Cunningham. 2003. A review of explanation and explanation in case-based reasoning. (2003).

[17] Thorsten Emig and Jussi Peltonen. 2020. Human running performance from real-world big data. *Nature communications* 11, 1 (2020), 4936.

[18] Kai Epstude, Annika Scholl, and Neal J Roese. 2016. Prefactual thoughts: Mental simulations about what might happen. *Review of General Psychology* 20, 1 (2016), 48–56.

[19] Urtats Etxegarai, Eva Portillo, Jon Irazusta, Ander Arriandiaga, and Itziar Cabanes. 2018. Estimation of lactate threshold with machine learning techniques in recreational runners. *Applied Soft Computing* 63 (2018), 181–196.

[20] Ciara Feely, Brian Caulfield, Aonghus Lawlor, and Barry Smyth. 2020. Providing Explainable Race-Time Predictions and Training Plan Recommendations to Marathon Runners. In *Fourteenth ACM Conference on Recommender Systems* (Virtual Event, Brazil) *(RecSys '20)*. Association for Computing Machinery, New York, NY, USA, 539–544. https://doi.org/10.1145/3383313.3412220

[21] Ciara Feely, Brian Caulfield, Aonghus Lawlor, and Barry Smyth. 2020. Using Case-Based Reasoning to Predict Marathon Performance and Recommend Tailored Training Plans. In *Case-Based Reasoning Research and Development*. Springer, Berlin, Heidelberg, 67–81.

[22] Ciara Feely, Brian Caulfield, Aonghus Lawlor, and Barry Smyth. 2021. A Case-Based Reasoning Approach to Predicting and Explaining Running Related Injuries. In *International Conference on Case-Based Reasoning*. Springer, Springer, Berlin, Heidelberg, 79–93.

[23] Ciara Feely, Brian Caulfield, Aonghus Lawlor, and Barry Smyth. 2022. An Extended Case-Based Approach to Race-Time Prediction for Recreational Marathon Runners. In *Case-Based Reasoning Research and Development: 30th International Conference, ICCBR 2022, Nancy, France, September 12–15, 2022, Proceedings*. Springer, Springer, Berlin, Heidelberg, 335–349.

[24] Iztok Fister, Samo Rauter, Xin-She Yang, Karin Ljubič, and Iztok Fister. 2015. Planning the Sports Training Sessions with the Bat Algorithm. *Neurocomput.* 149, PB (Feb. 2015), 993–1002. https://doi.org/10.1016/j.neucom.2014.07.034

[25] Iztok Fister Jr and Iztok Fister. 2017. Generating the training plans based on existing sports activities using swarm intelligence. In *Nature-Inspired Computing and Optimization*. Springer, Berlin, Heidelberg, 79–94.

[26] Ander Garcia, Maria Teresa Linaza, Olatz Arbelaitz, M FUCHS, F RICCI, and L CANTONI. 2012. Evaluation of intelligent routes for personalised electronic tourist guides. In *Information and Communication Technologies in Tourism*. Springer, Berlin, Heidelberg, 284–295.

[27] Mark T Keane and Barry Smyth. 2020. Good counterfactuals and where to find them: A case-based technique for generating counterfactuals for explainable AI (XAI). In *International Conference on Case-Based Reasoning*. Springer, Springer, Berlin, Heidelberg, 163–178.

[28] Alison Keogh, Barry Smyth, Brian Caulfield, Aonghus Lawlor, Jakim Berndsen, and Cailbhe Doherty. 2019. Prediction Equations for Marathon Performance: A Systematic Review. *International Journal of Sports Physiology and Performance* 14, 9 (2019), 1159–1169.

[29] Leonid Kholkine, Tom De Schepper, Tim Verdonck, and Steven Latré. 2020. A machine learning approach for road cycling race performance prediction. In *International workshop on machine learning and data mining for sports analytics*. Springer, Springer, Berlin, Heidelberg, 103–112.

[30] J Kite-Powell. 2016. Polar: The original fitness tracker and heart rate monitor. Forbes.

[31] P López-Matencio, J Vales Alonso, FJ González-Castano, JL Sieiro, and JJ Alcaraz. 2010. Ambient intelligence assistant for running sports based on k-NN classifiers. In *3rd International Conference on Human System Interaction*. IEEE, IEEE, New York, New York, 605–611.

[32] MB Maron and SM Horvath. 1978. The marathon: a history and review of the literature. *Medicine and science in sports* 10, 2 (1978), 137–150.

[33] Alberto E. Minetti, Christian Moia, Giulio S. Roi, Davide Susta, and Guido Ferretti. 2002. Energy Cost of Walking and Running at Extreme Uphill and Downhill Slopes. *Journal of Applied Physiology (Bethesda, Md.: 1985)* 93, 3 (Sept. 2002), 1039–1046. https://doi.org/10.1152/japplphysiol.01177.2001

[34] John O'Donovan and Barry Smyth. 2005. Trust in recommender systems. In *Proceedings of the 10th international conference on Intelligent user interfaces*. ACM Press, San Diego, California, 167–174.

[35] Andrej Panjan, Nejc Sarabon, and Aleš Filipčič. 2010. Prediction of the successfulness of tennis players with machine learning methods. *Kinesiology* 42, 1 (2010), 98–106.

[36] Alen Rajšp and Iztok Fister. 2020. A Systematic Literature Review of Intelligent Data Analysis Methods for Smart Sport Training. *Applied Sciences* 10, 9 (2020), 3013.

[37] Barry Smyth. 2018. Fast Starters and Slow Finishers: A Large-Scale Data Analysis of Pacing at the Beginning and End of the Marathon for Recreational Runners. *Journal of Sports Analytics* 4, 3 (Jan. 2018), 229–242. https://doi.org/10.3233/JSA-170205

[38] Barry Smyth. 2019. Recommender Systems: A Healthy Obsession. *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (July 2019), 9790–9794. https://doi.org/10.1609/aaai.v33i01.33019790

[39] Barry Smyth. 2021. How recreational marathon runners hit the wall: A large-scale data analysis of late-race pacing collapse in the marathon. *PloS one* 16, 5 (2021), e0251513.

[40] Barry Smyth and Pádraig Cunningham. 2017. Running with Cases: A CBR Approach to Running Your Best Marathon. In *Case-Based Reasoning Research and Development - 25th International Conference, ICCBR 2017, Trondheim, Norway, June 26-28, 2017, Proceedings*. Springer, Berlin, Heidelberg, 360–374. https://doi.org/10.1007/978-3-319-61030-6_25

[41] Barry Smyth and Pádraig Cunningham. 2018. An Analysis of Case Representations for Marathon Race Prediction and Planning. In *Case-Based Reasoning Research and Development - 26th International Conference, ICCBR 2018, Stockholm, Sweden, July 9-12, 2018, Proceedings*. Springer, Berlin, Heidelberg, 369–384. https://doi.org/10.1007/978-3-030-01081-2_25

[42] Barry Smyth, Aonghus Lawlor, Jakim Berndsen, and Ciara Feely. 2022. Recommendations for marathon runners: on the application of recommender systems and machine learning to support recreational marathon runners. *User Modeling and User-Adapted Interaction* 32, 5 (2022), 787–838.

[43] Marta Stragà and Donatella Ferrante. 2014. Will You Train Harder for the Next Marathon? The Effect of Counterfactual and Prefactual Thinking on Marathon Runners' Intentions. In *Proceedings of the Trieste Symposium on Perception and Cognition*. EUT Edizioni Università di Trieste, Trieste, EUT Edizioni Università di Trieste, 144–146.

[44] Myung-Kyung Suh, Ani Nahapetian, Jonathan Woodbridge, Mahsan Rofouei, and Majid Sarrafzadeh. 2012. Machine Learning-Based Adaptive Wireless Interval Training Guidance System. *Mob. Netw. Appl.* 17, 2 (April 2012), 163–177. https://doi.org/10.1007/s11036-011-0331-5

[45] Zahari Taha, Rabiu Muazu Musa, Anwar PP Abdul Majeed, Muhammad Muaz Alim, and Mohamad Razali Abdullah. 2018. The identification of high potential archers based on fitness and motor ability variables: A Support Vector Machine approach. *Human movement science* 57 (2018), 184–193.

[46] Brett G Toresdahl, Jordan D Metzl, James Kinderknecht, Kathryn McElheny, Polly de Mille, Brianna Quijano, and Mark A Fontana. 2023. Training patterns associated with injury in New York City Marathon runners. *British Journal of Sports Medicine* 57, 3 (2023), 146–152.

[47] Gytis Tumas and Francesco Ricci. 2009. Personalized mobile city transport advisory system. *Information and communication technologies in tourism 2009* (2009), 173–183.

[48] Bereket Abera Yilma, Yannick Naudet, and Hervé Panetto. 2019. Introduction to personalisation in cyber-physical-social systems. In *On the Move to Meaningful Internet Systems: OTM 2018 Workshops: Confederated International Workshops: EI2N, FBM, ICSP, and Meta4eS 2018, Valletta, Malta, October 22–26, 2018, Revised Selected Papers*. Springer, Springer, Berlin, Heidelberg, 25–35.