A Novel Recommender System for Helping Marathoners to Achieve a New Personal-Best

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

Pádraig Cunningham
Insight Centre for Data Analytics
School of Computer Science
University College Dublin, Ireland
padraig.cunningham@ucd.ie

ABSTRACT

We describe a novel application for recommender systems – helping marathon runners to run a new personal-best race-time – by predicting a challenging, but achievable target-time, and by recommending a tailored race-plan to achieve this time. A comprehensive evaluation of prediction accuracy and race-plan quality is provided using a large-scale dataset with almost 400,000 runners from the last 12 years of the Chicago marathon.

CCS CONCEPTS

• **Information systems → Personalization**; *Data analytics*;

KEYWORDS

Recommender systems, sports analytics

1 INTRODUCTION

Marathon running is a popular mass-participation sport that attracts millions of runners, from all walks of life, to our city streets. Running a marathon is hard and completing the 42.2 kms on race-day is the final stage, after months of training. But running a marathon is also incredibly rewarding, and for many it is an opportunity to push themselves to a new *personal-best* finish-time.

Running a marathon personal-best needs careful planning. It starts with a target-time to aim for; a time that is not so easy that you will feel untested, but also not so hard that you run the risk of ruining your race because you hit the wall. Various race predictors exist to help runners predict their likely finish-times, based on factors including gender, age, experience, past races, even training and strategy; see [3, 12, 17]. However, they usually base their predictions on shorter races and are not specifically tuned for helping runners to predict the *stretch-goal* that a personal-best represents. And none help the runner when it comes to planning how to achieve this time.

A race-plan — how a runner paces their race — is critically important, especially in an endurance event (see [2, 4, 11]), and considerable research has been devoted to understanding pacing in the marathon [7, 9, 16]. In this work we argue that a target finish-time alone is not enough to ensure marathon success: runners need a race-plan or *pacing plan* to achieve this time, a segment by segment plan for how fast or slow they should run, given the characteristics of the course, so that they will meet the target-time. For example, some runners may plan to run *even-splits*, rarely varying their pace throughout the race. Others will aim for *positive-splits*, running a slower second-half compared to the first, while others will aim for *negative-splits*, running a faster second-half. We argue that such coarse-grained strategies do not go far enough. A good pacing plan will help a runner to manage their effort throughout the race, segment by segment, hill by hill This is especially important during the crucial early stages of the marathon, when many go out too fast, and helps to reduce the risk of hitting the wall later in the race.

The main contribution of this work is to introduce a novel recommender system for helping marathon runners to identify, and plan for, new personal-best (*PB*) finish-times. We describe how to construct suitable *training cases* from conventional race-records, and how to use these cases to *predict* a *PB* time and *recommend* a tailored pacing plan. We evaluate the results using data from the last 12 years of the Chicago marathon.

2 RECOMMENDING A PERSONAL-BEST

In this section we describe how to transform marathon race data into suitable training cases for generating *PB* predictions and their corresponding race-plans.

2.1 From Races to Cases

The starting point for this work is a marathon *race-record*, a set of *split-times* at regular intervals. Most big-city marathons provide 5km split-times, which we use here.

Our solution adopts a Case-Based Reasoning (CBR) approach: CBR seeks to solve a new problem by reusing and adapting the solutions to similar, past problems; see [1, 8, 13]. Here the 'problem' we want to solve is to predict a *PB* marathon (that is, a finish-time and a race plan) for a runner based on a suboptimal (*non personal-best* or *nPB*) race. The *nPB* race acts as the *problem description* while the runner's *PB* race is the *problem solution* [14]. A runner with data for n past races provides us with $n-1$ problem-solution cases, corresponding to the $n-1$ non *PB*s, each paired with the runner's identified *PB*. Thus, each case is a $\langle nPB, PB\rangle$ pair and the *PB* can be used as the basis of a *PB* time and race plan for other runners who have run races similar to the *nPB*. An example case is shown in Figure 1. The *nPB* and *PB* races have the same representation. The most fundamental feature is the average pace with the race profile represented as deviations from this baseline.

RecSys '17, August 27-31, 2017, Como, Italy

ACM ISBN 978-1-4503-4652-8/17/08…$15.00
https://doi.org/10.1145/3109859.3109874

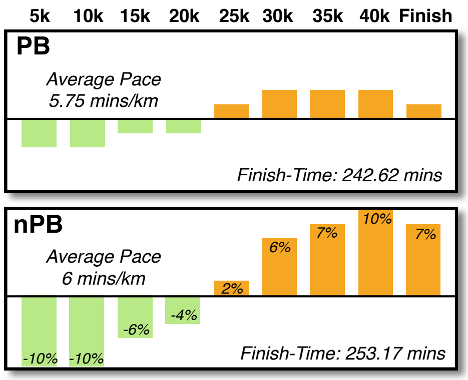

Figure 1: Sample case with *nPB* and *PB* times and pacing.

More formally, given a *repeat* runner r with race-records, $m_1, \ldots m_n$, we identify the personal-best as the race with the fastest finish-time; this may not be their true personal-best, but it represents their best race within the available data. We generate $n - 1$ race cases by pairing each *nPB* race with this *PB* race; see Equation 1.

$$c_{ij}(r, m_i, m_j) = \left\langle nPB_i(r, m_i), PB(r, m_j) \right\rangle \tag{1}$$

Figure 1 shows a sample case with a 253-minute *nPB* and a 242-minute *PB*. The *nPB* is characterised by a much more varied pacing profile; the runner started fast and finished slow. In contrast, their *PB* race is much more evenly paced, neither starting out too fast, nor finishing too slow, and completing their race with a modest positive-split and an 11-minute *PB*.

2.2 Predicting a Best Achievable PB Time

We treat the task of determining a challenging but achievable *PB* time as a prediction problem. The intuition is that the features of *nPB* races are predictive of future *PB* times. Thus, we use the *nPB* parts of race-cases as training data and the *PB* times as the target prediction feature. In Section 3 we evaluate a number of standard machine learning algorithms for this.

2.3 Recommending Suitable Race Plans

Next we need to recommend a suitable race-plan for achieving this *PB* time. For the purpose of this work, a race-plan is a sequence of paces during each of the (5km) race segments; rather than using *actual* paces we focus on *relative* paces for the purpose of race-plan recommendation. To generate a plan we identify the k cases whose *PB* times are closest to the predicted *PB* time. These cases correspond to runners who managed to achieve a similar *PB*, to the one predicted for the current runner. The assumption is that the *PB* pacing profiles for these k runners provide a basis for the new race-plan. For now we generate a plan based on the mean relative segment paces for the k cases; obviously this is just one of a range of strategies that will be considered as part of future work.

3 EVALUATION

We use an evaluation dataset of marathon records from the last 12 years of the Chicago marathon. There are 387,077 individual race records for 287,906 unique runners (45% female). From these we produce 99,171 race cases involving pairs of *nPB* and *PB* races. We use 10-fold cross-validation to evaluate the PB predictions and race-plan recommendations. For the former we calculate the percentage difference between the predicted PB and actual PB of each test case. And to evaluate race-plans we compute the similarity between the recommended plan and the actual race-plan as the mean percentage difference between race segment paces. We test 3 standard machine learning algorithms for prediction — linear regression (*Reg*), *kNN* (with $k = 10$), and elastic nets (*EN*) — each of these will typically generate different PB predictions, which in turn will lead to different race-plan recommendations.

	Reg		*kNN*		*EN*	
	Err	*Sim*	*Err*	*Sim*	*Err*	*Sim*
M	5.20	92.75	5.25	92.61	5.38	92.58
F	4.78	93.49	4.87	93.33	4.96	93.31

Table 1: Summary results for *Reg*, *kNN*, and *EN*.

3.1 Prediction Error & Plan Similarity

Table 1 shows the mean prediction error and race-plan similarity for these algorithms, for men and women. We can see error rates of about 5%, lower for women than men, and race-plans that are more than 90% similar to the actual *PB* plans, again slightly better for women than for men. These prediction errors are competitive with those reported by [17], albeit for the different problem of *personal-best* prediction, rather than regular race-time prediction, and without the benefit of training and injury data, but with more race data. In what follows we will present more detailed results for *Reg* (the other algorithms behave similarly) leaving further algorithmic tuning and evaluation as a matter for future work.

3.2 Ability as Finish-Time

Figure 2(a) & (b) shows the prediction error and plan similarity for runners with different (*nPB*) finish-times. Error and similarity tend to deteriorate for increasing finish-times. For example, we can generate PB predictions for fast 180-minute marathoners within about 2.5% of actual PB times, but the error grows to over 4% for 240-minute finishers, before plateauing around 5% after the 300-minute mark. Similarly, recommended race-plans become less and less like the actual race-plans, as finish-times increase.

Thus, faster runners may be more predictable, and this hints that their race cases may be of higher quality than those of slower runners, a point that we will return to presently. Female runners enjoy better predictions and more similar plan recommendations than males, which is consistent with research [9, 16] on the better pacing discipline of female runners, making their races more predictable.

3.3 Personal-Best Difference

How does performance vary with differences between *nPB* and *PB* finish-times? Are runners with closer *nPB* and *PB* times easier to predict? Figure 2(c) & (d) show how *PB Difference* impacts prediction

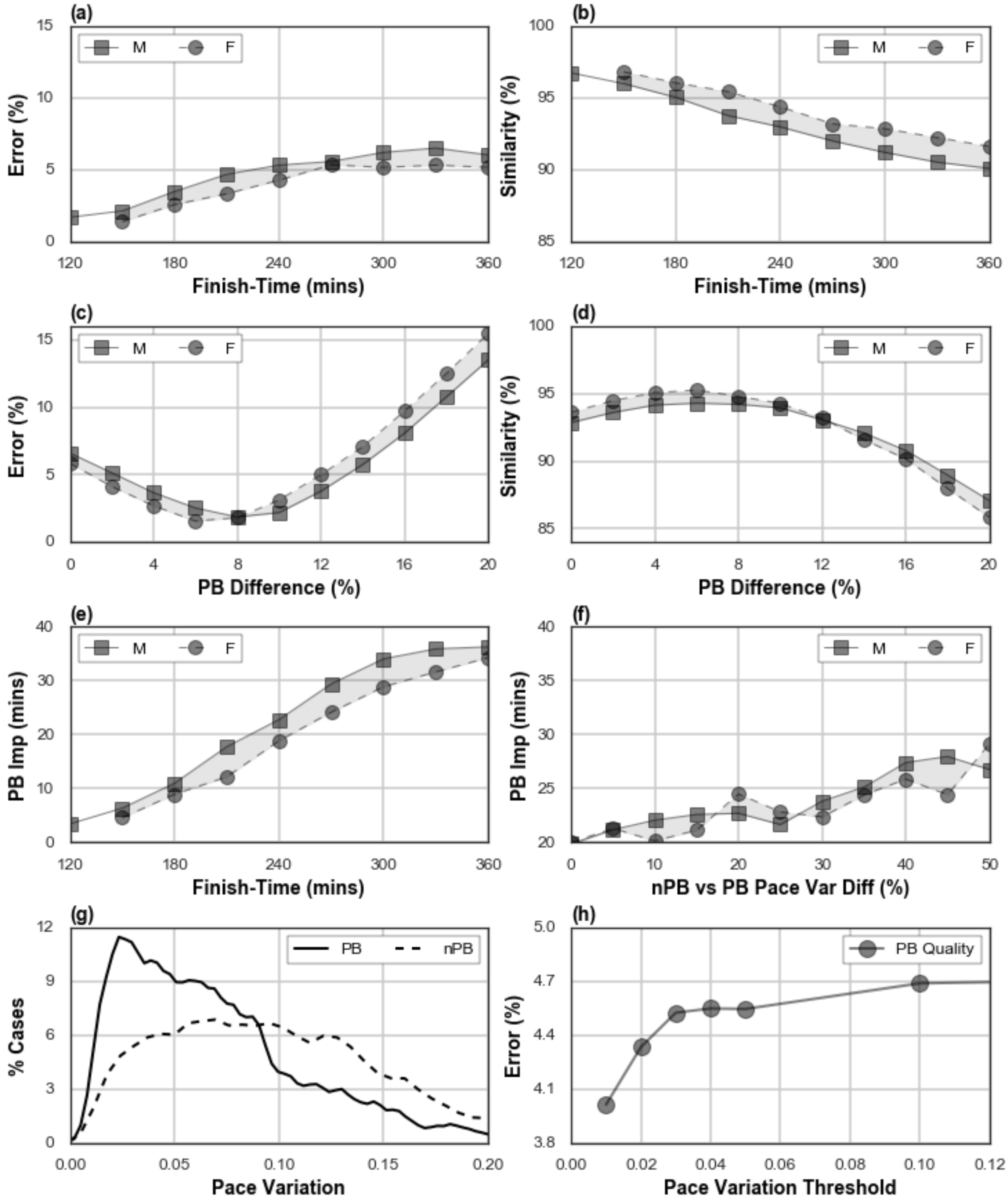

Figure 2: Summary evaluation results.

error and plan similarity; note, a *PB Difference* of 10% means the *PB* time is 10% *faster* than the *nPB* time. Runners with very similar (<4%), or very different (>12%), *nPB* and *PB* races are more difficult to predict, and recommend for, with the best performance seen for *PB Difference* values of about 7-8%.

Low *PB Difference* runners tend to be either: (a) faster, regular marathoners enjoying modest incremental improvements; or (b) slower, infrequent runners registering marginal gains, and who are less motivated by a new personal-best. In combination this makes these runners more difficult to model. On the other hand the high *PB Difference* runners are more difficult to predict for, because they are registering unusually large *PB* improvements.

3.4 Personal-Best Improvements

More concretely, Figure 2(e) shows the actual PB improvements predicted for various finish-times. For example, 240-minute *nPB* runner is predicted a 20-minute faster *PB* (+/- 5 or 6 minutes) under the right conditions.

Another factor that impacts *PB* improvement is the difference in *pace variation* between a runner's races; we can measure pace variation as the coefficient of variation of the segment paces of a race. More even pacing is usually associated with better quality races and larger differences between the pace variation of *nPB* and *PB* races usually means that the *nPB* race is a poor one (lots of pace variation, perhaps indicating the runner hit the wall) relative to a higher quality *PB* race, with a lot less variation. Such a case should exhibit more scope for improvement, which is what we see in Figure 2(f); cases with similar pace variations predict 20 minute *PB* improvements where as cases with greater pace variation differences predict 25-30 minute improvements.

3.5 On Case Quality

This suggests not all cases are created equally. In Figure 2(g) — pace variation histograms for *nPB* and *PB* races — we see, not surprisingly, that *PB*'s exhibit less pace variation than *nPBs*. Thus, using pace variation as a measure of race quality, we can filter cases, for quality, by excluding those whose *PB* pace variation exceeds a minimum threshold.

When we do this for different thresholds, in Figure 2(h), we see a marked effect on prediction error. For race-cases with high quality *PBs* (pace variation threshold < 3%) the prediction error is low, and it disimproves steadily as this threshold.increases, because lower quality cases are included. For example, when we admit cases with *PB* pace variations of up to 0.1 the error is 4.7% compared to just under 4% when we only admit more evenly paced *PB* races (pace variation = 0.01), a relative increase in error of almost 20%

4 DISCUSSION

In this paper we have described a novel use-case for recommender systems: helping marathon runners to achieve a personal-best in a future race by providing them with a challenging but achievable goal-time and an actionable race-plan to achieve it. Our results show that accuracte predictions can be made and that high-quality race-plans can be recommended, at least in the sense that these predictions and recommendations are close matches for the properties of the PB's that test runners have completed. This work is related to a growing interest in the application of recommender systems and similar technolgies to areas such as personal health and wellbeing; see for example [5, 6, 10, 15].

As always there is room for improvement. While prediction error rates are low, they increase for slower runners; for those finishing after the 4-hour mark, predictions, which come with an error rate of 5+%, are likely to be 12+ minutes off relative to the 'true' PB time. These runners stand to benefit most, from this system and, therefore, they stand to suffer most from growing error rates. This speaks to the need for more effective prediction methods that can provide for more stable, lower error rates across all finish-times. To do this we will explore further algorithms and feature-sets in the future, paying particular attention to the benefit of including enriched race histories as part of our training cases.

Another important matter to bear in mind concerns the nature of the evaluation itself. By design our measure of prediction success, and recommendation quality, is the personal-best race eventually run by a test runner. Since this race was completed without the benefit of this recommender system it raises the question of whether these runners may have achieved even faster PBs had they received our predictions and race-plans? This is certainly a valid question and, not doubt there was room for improvement for many of these runners, even during their PB races. Whether our approach can drive even further improvement remains to be seen, and this can only be tested by evaluating the outcomes of races where runners have had the benefit of these recommendations. Another opportunity for future research.

5 CONCLUSIONS

The main contribution of this work is a novel application for recommender systems: helping marathoners to achieve personal-best times by predicting a challenging, yet achievable, target-time for their next race, and by recommending a tailored race-plan for achieving it. Evaluation results show strong prediction performance when tested against historical *PB* times.

This short paper is less about the sophistication of the prediction/recommendation algorithms used – we use straightforward techniques, which offer considerable room for tuning and improvement – and more about the novel domain, race representation, and the dual tasks of prediction and recommendation. Going forward, as mentioned above, we plan to explore the potential for further algorithmic improvements, including, for instance, using multiple *nPB* races within *PB* cases to evaluate the benefits more comprehensive race histories, or by harnessing auxialliary data such as training or injury data. We will also apply these methods to other endurance sports, such as cycling, and look to leverage additional performance data such as heart-rate and power-meter readings as further indicators of effort, and also to provide more detailed race-plans to athletes.

ACKNOWLEDGMENTS

Supported by Science Foundation Ireland through the Insight Centre for Data Analytics, under grant number SFI/12/RC/2289, and by Accenture Labs, Dublin.

REFERENCES

[1] Agnar Aamodt and Enric Plaza. 1994. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Commun.* 7, 1 (1994), 39–59. https://doi.org/10.3233/AIC-1994-7104

[2] Bertrand Baron, Farouck Moullan, Fabien Deruelle, and Timothy D Noakes. 2009. The role of emotions on pacing strategies and performance in middle and long duration sport events. *British Journal of Sports Medicine* (2009).

[3] Francesco Bartolucci and Thomas Brendan Murphy. 2015. A finite mixture latent trajectory model for modeling ultrarunners' behavior in a 24-hour race. *Journal of Quantitative Analysis in Sports* 11, 4 (2015), 193–203.

[4] Deryn Bath, Louise A Turner, Andrew N Bosch, Ross Tucker, Estelle V Lambert, Kevin G Thompson, and Alan St Clair Gibson. 2012. The effect of a second runner on pacing strategy and RPE during a running time trial. *International journal of sports physiology and performance* 7, 1 (2012), 26–32.

[5] Shlomo Berkovsky and Jill Freyne. 2013. Food Recommendations: Biases that Underpin Ratings. In *Proceedings of the 3rd Workshop on Human Decision Making in Recommender Systems in conjunction with the 7th ACM Conference on Recommender Systems (RecSys 2013), Hong Kong, China, October 12, 2013.* 42. http://ceur-ws.org/Vol-1050/paper8.pdf

[6] Shlomo Berkovsky, Jill Freyne, and Mac Coombe. 2012. Physical Activity Motivating Games: Be Active and Get Your Own Reward. *ACM Trans. Comput.-Hum. Interact.* 19, 4 (2012), 32:1–32:41. https://doi.org/10.1145/2395131.2395139

[7] Matthew P Buman, Britton W Brewer, and Allen E Cornelius. 2009. A discrete-time hazard model of hitting the wall in recreational marathon runners. *Psychology of Sport and Exercise* 10, 6 (2009), 662–666.

[8] Pádraig Cunningham, Barry Smyth, and Neil J. Hurley. 1995. On the use of CBR in optimisation problems such as the TSP. In *Case-Based Reasoning Research and Development, First International Conference, ICCBR-95, Sesimbra, Portugal, October 23-26, 1995, Proceedings.* 401–410. https://doi.org/10.1007/3-540-60598-3_36

[9] 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.

[10] Elizabeth Dunford, Helen Trevena, Chester Goodsell, Ka Hung Ng, Jacqui Webster, Audra Millis, Stan Goldstein, Orla Hugueniot, and Bruce Neal. 2014. FoodSwitch: a mobile phone app to enable consumers to make healthier food choices and crowdsourcing of national food composition data. *JMIR mHealth and uHealth* 2, 3 (2014), e37.

[11] Carl Foster, Matthew Schrager, Ann C Snyder, and Nancy N Thompson. 1994. Pacing strategy and athletic performance. *Sports Medicine* 17, 2 (1994), 77–85.

[12] Peter S Riegel. 1981. Athletic Records and Human Endurance: A time-vs.-distance equation describing world-record performances may be used to compare the relative endurance capabilities of various groups of people. *American Scientist* 69, 3 (1981), 285–290.

[13] Barry Smyth. 2007. Case-Based Recommendation. In *The Adaptive Web, Methods and Strategies of Web Personalization.* 342–376. https://doi.org/10.1007/978-3-540-72079-9_11

[14] Barry Smyth and Pádraig Cunningham. 2017. Running with Cases: A CBR Approach to Running Your Best Marathon. In *Proceedings of the 25th International Conference on Case-Based Reasoning ICCBR 2017)*, David W. Aha and Jean Lieber (Eds.). Springer International Publishing, 360–374".

[15] Maria Alice L. Thé, Ronda C. Bell, Katia G. Camargo, Rosina Weber, Alejandro Martins, and Ricardo M. Barcia. 2000. *Case-Based Reasoning for Nutrition Consulting.* Springer London, London, 180–190. https://doi.org/10.1007/978-1-4471-0465-0_11

[16] Nicholas William Trubee. 2011. *The Effects of Age, Sex, Heat Stress, and Finish Time on Pacing in the Marathon.* Ph.D. Dissertation. University of Dayton.

[17] Andrew J Vickers and Emily A Vertosick. 2016. An empirical study of race times in recreational endurance runners. *BMC Sports Science, Medicine and Rehabilitation* 8, 1 (2016), 26.