Title	A Novel Recommender System for helping Marathoners to Achieve a new Personal-Best
Authors(s)	Smyth, Barry; Cunningham, Pádraig
Publication date	2017-08-31
Publication information	: 116-120
Conference details	RecSys '17 the Eleventh ACM Conference on Recommender Systems, Como, Italy, 27-31 August 2017
Publisher	ACM
Item record/more information	http://hdl.handle.net/10197/9053
Publisher's statement	© ACM, 2017. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in RecSys '17 Proceedings of the Eleventh ACM Conference on Recommender Systems, (2017) doi:10.1145/3109859.3109874
Publisher's version (DOI)	10.1145/3109859.3109874

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.

1. INTRODUCTION

Marathon running is a popular mass-participation sport that routinely attracts millions of runners, from all walks of life, to our city streets. Running a marathon is hard and completing the 42.2km course on race-day is just the final stage after months of long, hard 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 [2, 10, 13]. 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 [1,3,9], and considerable research has been devoted to understanding pacing in the marathon [6, 7, 12]. In this work we ar-

gue 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 pace effectively, throughout the race, segment by segment, hill by hill, but particularly during the crucial early stages when many go out too fast, and help 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 we 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. We convert these split-times into *average paces* (mins/km) for each of the 8x5km segments plus the final race segment as per Equation 1; e.g., the pace for runner r for the 3rd (10-15km segment), $pace(r, d)$ is based on the difference in time between this 3rd segment and the 2nd segment, divided by the distance between them, 5km in this case; it would be 2.2km when $d = 9$, for the final 2.2km segment. For convenience we also calculate the *relative pace* of a runner for a given segment, as a fraction of their average pace for the race; see Equation 2

$$pace(r, d) = \frac{time(seg(d)) - time(seg(d-1))}{dist(d) - dist(d-1)} \quad (1)$$

$$rel\_pace(r, d) = \frac{pace(r, d)}{mean_pace(r)} \quad (2)$$

To predict PB times we need training and test data which pairs *regular*, *non personal-best* (nPB) races with corresponding *personal-best* (PB) races, for a given runner.
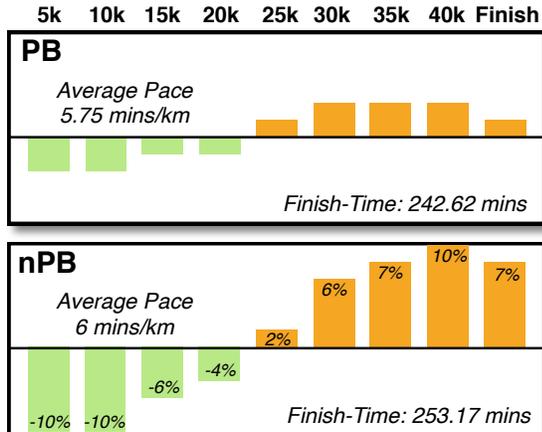

Figure 1: An example race-case showing nPB and PB finish-time and pacing components.

Given a *repeat* runner, r, with race-records, $m_1, ... m_n$, we identify the personal-best as the record with the fastest finish-time; of course this may not be their true personal-best but it represents their best race within the window of data available. We generate $n-1$ race cases by pairing each nPB race record with this PB record; see Equation 3.

$$c_{ij}(r, m_i, m_j) = \left\langle nPB_i(r, m_i), PB(r, m_j) \right\rangle \quad (3)$$

Figure 1 shows a sample case with a 253-minute nPB paired with a 242-minute PB. The nPB race is characterised by a much more varied pacing profile; the runner started out fast and finished slow. In contrast, their PB race is represented by a much more disciplined pacing profile, 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 classical prediction problem. The intuition is that the features of nPB races are predictive of future PB finish-times. Thus, we use the nPB parts of race-cases as training data and the PB finish-times as the target prediction feature. In the present work we evaluate a number of standard machine learning algrotihms for this task, as per Section 3.

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

	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 prediction error and race-plan similarity results for Reg, kNN, and EN.

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 provides 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, and elastic nets (EN) — each of these will typically generate different PB predictions, which in turn will lead to different race-plan recommendations.

3.1 Prediction Error & Plan Similarity

Table 1 shows the mean prediction error and race-plan similarity for these algorithms, for men and women. On average 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 [13] 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 a lot 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, be-

fore 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 perhaps hints that their race-cases are 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 [7, 12] 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 cases with closer nPB and PB times easier to predict? Figure 2(c) & (d) show how *PB Difference* impacts prediction 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 who are enjoying modest incremental improvements; or (b) slower, infrequent runners who register only marginal improvements, 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

To make this more tangible, Figure 2(e) shows the actual PB improvements predicted for different finish-times. For example, a runner with a 240-minute nPB is predicted to achieve a PB some 20 minutes faster ($+/-$ 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, as in Figure 2(h), we see a marked effect on average prediction error. For, race-cases with high quality PBs (pace variation threshold $< 3\%$) prediction error is low, and it disimproves steadily as this threshold.increases, and lower quality cases are included. For example, when we admit cases with PB pace

variations of up to 0.1 the prediction error is 4.7% compared to just under 4% when we only consider 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 [4, 5, 8, 11].

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-time. 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 central contribution of this work is a novel application for recommender systems: helping marathon runners to achieve a new personal-best finish-time by predicting a challenging, yet achievable, target-time for their next race, and a tailored race-plan for achieving it. Evaluation results suggest 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 his-
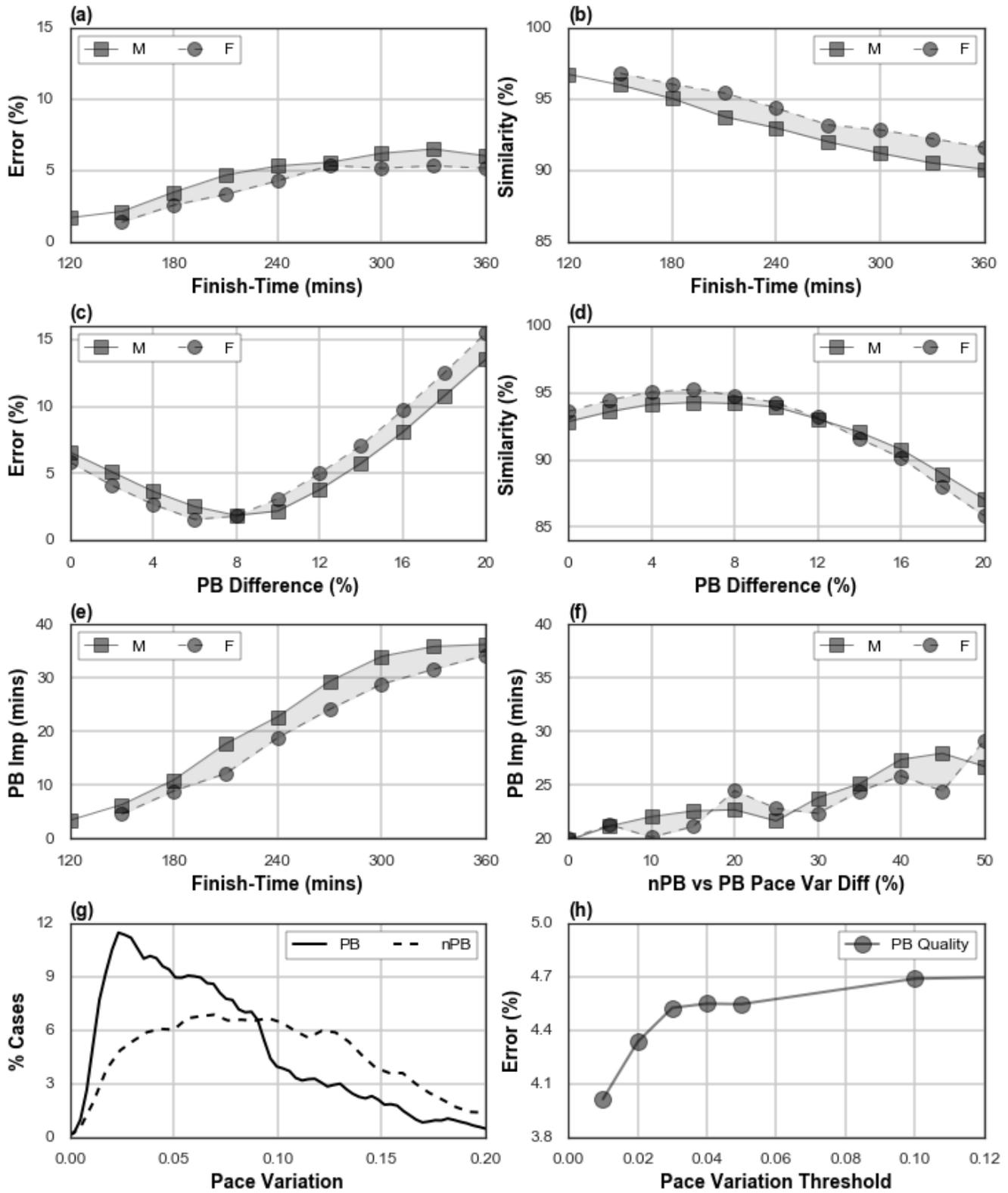
Figure 2: Summary evaluation results.

tories. We will also apply these methods for 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 de-

tailed race-plans to athletes.

7. REFERENCES

[1] B. Baron, F. Moullan, F. Deruelle, and T. D. Noakes. The role of emotions on pacing strategies and performance in middle and long duration sport events. *British Journal of Sports Medicine*, 2009.

[2] F. Bartolucci and T. B. Murphy. A finite mixture latent trajectory model for modeling ultrarunners' behavior in a 24-hour race. *Journal of Quantitative Analysis in Sports*, 11(4):193–203, 2015.

[3] D. Bath, L. A. Turner, A. N. Bosch, R. Tucker, E. V. Lambert, K. G. Thompson, and A. S. C. Gibson. 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):26–32, 2012.

[4] S. Berkovsky and J. Freyne. 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.*, page 42, 2013.

[5] S. Berkovsky, J. Freyne, and M. Coombe. Physical activity motivating games: Be active and get your own reward. *ACM Trans. Comput.-Hum. Interact.*, 19(4):32:1–32:41, 2012.

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

[7] R. O. Deaner, R. E. Carter, M. J. Joyner, and S. K. Hunter. Men are more likely than women to slow in the marathon. *Med Sci Sports Exerc*, 4(3):607–616, 2014.

[8] E. Dunford, H. Trevena, C. Goodsell, K. H. Ng, J. Webster, A. Millis, S. Goldstein, O. Hugueniot, and B. Neal. 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):e37, 2014.

[9] C. Foster, M. Schrager, A. C. Snyder, and N. N. Thompson. Pacing strategy and athletic performance. *Sports Medicine*, 17(2):77–85, 1994.

[10] P. S. Riegel. 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):285–290, 1981.

[11] M. A. L. Thé, R. C. Bell, K. G. Camargo, R. Weber, A. Martins, and R. M. Barcia. *Case-Based Reasoning for Nutrition Consulting*, pages 180–190. Springer London, London, 2000.

[12] N. W. Trubee. *The Effects of Age, Sex, Heat Stress, and Finish Time on Pacing in the Marathon*. PhD thesis, University of Dayton, 2011.

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