Klara Schlüter
Machine Learning
NTNU - Department of Computer Science

October 3, 2019

ASSIGNMENT 2 — Essay

Running with cases: A CBR Approach to Running Your Best Marathon

In their paper "Running with cases: A CBR Approach to Running Your Best Marathon", Barry Smyth and Pádraig Cunningham from the University College Dublin publish their research in the use of case-based reasoning (CBR) for suggesting race plans for marathon runners [1]. This essay starts with a short summary of the papers content before inspecting the researcher's goals and methodology as well as their evaluation and discussion of the results. To conclude, a personal opinion on the paper is given.

Most marathon runners follow a certain racing strategy. To achieve the best possible result without overestimating the own abilities, they are aiming for a certain, precalculated finish time and try to adapt their pace (measured in time per distance) in each section of the race. CBR can be used to suggest to a runner both an achievable target time and what is called a pacing plan, containing the goal pace for every section of the race. The proposed system accepts the finish time and pacing data of a previous race and retrieves similar other runners and data for their personal best races. From the retrieved records, it computes a race plan for the next race according to three different prediction strategies. Comparing suggestions obtained for runners of the London Marathon to their best times and paces, the error rates and profile similarity indicates which one of the three prediction strategies works best on the test data.

Smyth's and Cunningham's goal is to help runners to improve their performance at their next marathon. More precise, they argue that the race plans occational runners are using are often too general and do not make use of the runners personal abilities and strengths. The paper states that "more personalized and customized approaches to pacing" [1] and target finish time are the better solution. Thus they are aiming at using CBR on data of the London Marathon to produce such tailormade race plans.

To adress this goal, they developed a CBR prototype on marathon running. The problem is formalized by defining a case as a pair of two races of the same runner, where the second entry represents the runners best performance. One race contains the finish-time and a pacing profile (a set of paces for the different sections of the race). To calculate a suggestion for a given runner, a query race is needed, i.e. the race data of a race he or she already completed. With this race, the most similar case is searched. Before actually retrieving similar cases, the known cases are filtered by two criterions: Firstly, only cases with a finish time that differs only in a predefined range from the finish time of the query race are taken into account. Secondly, cases that represent runners of an other gender are excluded. Now, the query race data is compared to the first entry (the slower race) of each known case. The *k-Nearest-Neighbour* algorithm is used to retrieve the cases, that are most similar to the query case with respect to the euclidian distance.

The race plan is computed from the second entries of the retrieved cases, the personal best records of the certain runner. The paper suggests three different strategies: *Best* selects the case with the best finish time and returns its pace plan and finish time, weighted by the ratio of the query time and the selected finish time. In the *Mean* strategy, the suggestion is simply calculated as the mean finish time and the mean pace plan of all retrieved cases records. The last strategy, *Even*, preferes pace plans with the least possible variation in racing speed, and again weights the best racing plan before suggesting it.

To evaluate their CBR system, Smyth and Cunningham use the data of the London Marathon between 2011 and 2016. They perform 10-fold cross-validation: in ten iterations, a test set containing 10% of the cases is drawn randomly. The rest of the cases is used as known cases. For every case in the test set, the first entry is used as input for the prototype. The output is then compared to the second entry. The comparison is based on the error regarding the finish time and the similarity of the pacing profile. Assessing it for different values for the size of the set of retrieved cases indicates *Mean* as the overall best prediction strategy. The authors explain this by the other two strategies in general prefering the cases with better personal best times. Furthermore, the evaluation of the suggestion is plotted against the overall finish-time and the magnitude of improvement of the cases. This shows that the computed suggestions are more accurate for faster runners and bigger improvements between the two entries of a case. Overall, the prediction is better for female runners than for male runners. This is justified by female runners training more disciplinated and therefore more predictable.

In the discussion of this results, the paper firstly points out the improvement the suggested race plan represents in average as a strength of the system. Secondly, two weaknesses are identified: There is no data about runners trying to follow the suggested plans, and the race records in the data basis are chosen only by finish time and therefore may not represent the optimum. For both weaknesses, the authors plan on doing further research, and describe how the specific problem could be adressed.

In my opinion, the authors gave a good introduction to the relevant parts of marathon running. They explained the problem of the finish time shortly, but understandable and therefore picked up readers with no idea of marathon running to still be able to follow their motivation and results.

Furthermore, a comprehensible and reasonable system to address the explained problem is introduced and well explained. Only in the evaluation of the system, I think that the paper lost track of their research goal of providing "personalized and customized" [1] race plans. Smyth's and Pádraig's CBR approach gives tailored race plans, but they do not test if runner's following their race plans actually improve. There are no results of runner's competing according to the suggested race plan. Instead, they compare their race plans to the runner's performances in races where they followed other plans or no plans at all. With this testing approach, they do not investigate if the suggested race plans help runner's (if their system gives reasonable output), they can only make conclusions about how well their system predicts race performances according to earlier races. In my opinion, that is a totally different task. The authors justification of their results is thus based on tests which are not suitable for the research goal.

To judge the impact on future work in the field, from my point of view, the field has to be specified, since the paper might have a different impact to marathon running than to computer science. The authors speak of the former as the "practical reason for the proposed system".

There are two reasons why I do not believe that the paper has an impact on further reseach in marathon running. Firstly, the paper clearly seems to be made for computer scientists instead of sport scientists. While the explanation of the marathon running part of the work is very understandable, there is no introduction of case-based reasoning, so for someone with no knowledge in that area or not used to this way of thinking, the paper is hard to retrace. This may lead to less researchers in marathon running taking into account the provided CBR approach. The second reason is explained in the previous paragraph: since Smyth and Pádraig never had an athlete running a marathon according to a race plan suggested by their sytem, the paper does not provide any output data on if the race plans actually help marathon runners. It does not assess if the system gives race plans helping athletes, only, if the system predicts performances from previous races.

From the computer science point of view, the paper applied CBR to a real world example. The application was successfull, even if the results were not properly evaluated. It therefore represents a good example for the application of CBR. Still, since there are no new insights on CBR itself, the impact on future work on CBR is limited.

References

[1] B. Smyth and P. Cunningham, "Running with Cases: A CBR Approach to Running Your Best Marathon," in *Lecture Notes in Computer Science*, pp. 360–374, June 2017.

Submitted by Klara Schlüter on October 3, 2019.