

Talent Identification for Top Athletes

Christophe Billiet

Supervisor(s): Guy De Tré

I. INTRODUCTION

‘Talent Identification’ (TID) in sports is of crucial importance to the search for excellence in top-class sports and the efficient coaching of sports programmes. However, current TID-models struggle with low predictive power caused by, among other things, the absence of clearly notable relationships between the properties of young athletes and their eventual success on the highest levels of competition.

This project aims to research and develop new ‘soft computing’ (SC) techniques which could lead to an efficient decision system with a sufficiently high predictive value, intended to support TID.

II. RESEARCH AND PROPOSED METHODS

Within the Department of Movement and Sports Sciences of the UGent, a unique database was created, containing the chronological results of sports and medical tests administered to athletes from young age. Some of these athletes have evolved into elite athletes currently at the top of their discipline of sport.

Recent developments in ‘data mining’ (DM) have led to new techniques for data analysis, among which facilities for knowledge extraction from data collections [1]. More particularly, classification techniques based on decision trees allow the discovery of hidden knowledge rules in datasets. These knowledge rules are often hidden to experts on the matter (in this case: sports experts). These techniques could theoretically be used to construct the intended decision system.

However, these techniques can’t usually handle an imperfect or uncertain nature of data, which is typical for data originating from mea-

surements and/or human sources. Therefore, new SC techniques will be developed to be combined with the DM techniques [2], [3], [4]. Additionally, this will lead to

- a higher level of predictive value, since less information will be lost.
- better interpretable, semantically richer results, since human (sports experts) reasoning is better modeled with SC techniques.

Currently, the research is in a very early stage, but activities of the supporting ‘Database, Document and Content Management’ (DDCM) research group are centered around the use of SC techniques in information management.

III. CONCLUSIONS

It is too early to discuss results concerning content. Expectation is, however, that the improvement of the applicability of the DM techniques by the use of SC, will result in a greatly raised level of prediction performance of the decision system.

REFERENCES

- [1] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rüdiger Wirth, “CRISP-DM 1.0 Step-by-step data mining guide,” 2000.
- [2] Eyke Hüllermeier and Stijn Vanderlooy, “Why Fuzzy Decision Trees are Good Rankers,” *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 6, pp. 1233–1244, 2009.
- [3] Smith Tsang, Ben Kao, Kevin Y. Yip, Wai-Shing Ho, and Sau Dan Lee, “Decision Trees for Uncertain Data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 99, no. PrePrint, pp. 15, 2009.
- [4] Xiaomeng Wang, Detlef D. Nauck, Martin Spott, and Rudolf Kruse, “Intelligent data analysis with fuzzy decision trees,” *Soft Computing*, vol. 11, no. 5, pp. 439–457, 2007.