Data mining in stabilometry: Application to patient balance study for sports talent mapping

Juan. A. Lara[1], David Lizcano[1], David[1] de la Peña, José M. Barreiro[2]

[1] Open University of Madrid, UDIMA - School of Engineering, Ctra. De la Coruña, km 38.500 – Vía de Servicio, 15 - 28400, Collado Villalba, Madrid, Spain
{juanalfonso.lara, david.lizcano, francis-codavid.delapena}@udima.es

[2] Technical University of Madrid, School of Computer Science, Campus de Montegancedo, s/n - 28660, Boadilla del Monte, Madrid, Spain
jmbarreiro@fi.upm.es

Abstract. Stabilometry is a branch of medicine responsible for the study of balance and postural control in human beings. To do this, it uses devices known as posturographs, which collect data related to people's balance. In this paper we propose the use of data mining techniques in order to build predictive models based on a number of variables related to the balance of the analysed subjects. The resulting models can be applied as classification tools for sports talent mapping by determining the sport or sporting discipline best suited to young sportspeople depending on balance, as balance plays a key role in many sports activities. According to the results for data on 15 professional basketball players and 18 ice-skaters, the predictive power is 90.91% in the best case (Unilateral Stance Test – Left Leg). This suggests that there is a close relationship between balance and the sport practised by professional sportspeople in our experiments.

Keywords: Stabilometry, Data Mining, Classification, Reference model, Sports Talent mapping.

1 Introduction

Stabilometry or **posturography** is a branch of medicine concerned with studying people's postural control [1, 2]. Patients take a series of tests in order to measure postural control [3]. Testing is based on the use of dynamometric platforms, called posturographs.

Stabilometric platforms record a huge amount of interesting data related to people's balance and postural control. Knowledge discovered from stabilometric data has led to quite a few advances in the field of medicine. However, these data are not straightforward to analyse, and specialized data analysis techniques have to be used. Data mining plays a key role in this respect [4].

In this paper, we describe how we applied data mining techniques in order to build classification models based on historical stabilometric data collected from different individuals. In particular, we applied decision trees and logistic regression in order to generate these models.
The analysed data include balance-related information (constituting the independent variables of our study) and other characteristics (each separately considered as a dependent variable in the conducted experiments). These characteristics include age, height and gender, as well as information related to the sports played by the respective individual. The resulting models explain these characteristics in terms of patient balance, as well as acting as a tool for recommending sports or sporting disciplines for young sports talents.

2 Background

Stabilometry was originally conceived merely as a technique for assessing patient postural control and balance. However, it is now considered to be a useful tool for diagnosing [5, 6] and treating [7] balance-related disorders. Some examples of its use are described in [8-13].
Throughout this research we used a modern posturography device called *Balance Master* from NeuroCom® Internacional [14]. Previous research has shown it to be very precise and reliable for analysing postural control [15, 16]. Balance Master consists of a metal platform, which is divided into two lengthwise interconnected plates and placed on the floor. The patient has to stand on the metal platform to perform a number of tests.
There are different types of **tests**. The patient has to perform the tests in a set order following the physician's instructions. Each test is designed to measure a different component of patient balance. Additionally, each test is divided into test **subtypes**, which include slight variations on the test.
As it is of special interest to the experts of this domain, we focused on the **US** (***Unilateral Stance***) test. The aim of this test is to measure sway in patients standing on one foot with eyes open and with eyes closed.
This test lasts 10 seconds, during which the patient has to stand as steadily as possible on only one leg on the platform. Patients have to perform each of the four test subtypes to complete the test: a) Stand on left leg with eyes open; b) Stand on left leg with eyes closed; c) Stand on right leg with eyes open; and d) Stand on right leg with eyes closed.
From the expert's point of viewpoint, the most important aspect of this test is the analysis of losses of balance as the patient performs the test. It is especially important to find out the extent and direction of the loss of balance and whether the imbalance ends in a fall, that is, whether patients are obliged to put down the foot that they are not standing on, which should be raised at all times.
As stabilometry is a relatively modern discipline, there is not much background on the application of data mining techniques to stabilometric data. Beyond the two proposals described in [18, 19] (where the authors apply data mining techniques to detect motor

fluctuations in Parkinson's disease and to analyse postural instability and consequent falls and hip fractures associated with the use of hypnotics in the elderly), we have not found any research in the literature specific to the application of data mining in the stabilometric domain, except for the investigation that we have conducted over the last few years in this domain [20-25]. This is the first paper in which we conduct research in order to use data mining as a tool for gaining a better understanding of the stabilometric domain.

3 Data and methods used

For this research we used stabilometric data from a total of 56 individuals. Of 56 subjects under analysis, 15 are professional basketball players, 18 are elite ice skaters and the other 23 are members of a control group of healthy people of different gender who are not professional sportspeople.
The studies focused on the US test. US is one of the tests that provides more interesting information about balance. The test was confined to studying both the *Left* and *Right* subtypes with eyes closed. Eyes open subtypes provided hardly any information of interest because the data in the different classes were almost constant.
The stabilometric data under analysis were acquired using the Balance Master static posturograph manufactured by Neurocom (Figure 1). This posturograph is composed of four sensors. Each sensor records the pressure of the patient's feet at regular 10-millisecond intervals throughout the test. These data generate time series.

Figure 1. Patient performing a test on a stabilometric platform.

This device generates time series that are hard to interpret, for which reason they were pre-processed by the times series knowledge discovery framework that we proposed elsewhere [20, 21, 22]. This framework was applied to acquire the raw data for analysis. These data include indicators related to subject balance, such as sway velocity, number of recorded imbalances, number of recorded falls, sum of the lengths of the recorded falls and maximum intensity of the recorded fall measurements. These are all the attributes generated by the above framework. We used all these attributes as **independent variables** in the later experiments.

The above indicators are measured by a framework that we implemented ad hoc for the stabilometry domain [22-25]. The patient time series constitute the framework input. The framework uses specialized time series analysis techniques to identify and characterize the time series events using a special-purpose event identification language [20]. Note that traditional techniques like Fourier transforms or wavelets are not applicable as they analyse times series as a whole, whereas experts in stabilometric time series focus exclusively on certain regions of interest in the time series that have particular features. These regions are the events based on which the balance indicators used in this paper are calculated (mean values of the different events). Figure 2a shows a snippet of a stabilometric time series, and Figure 2b illustrates an example of a fall event in the US test.

Number of Data Sample: 1000

DP	LF	RR	SH	LR	RF
1	86	20	0	89	22
2	86	20	0	89	22
3	86	20	0	89	22
4	86	21	0	89	22
5	86	21	0	89	22
6	86	21	0	89	22
7	86	21	0	89	23
8	86	21	0	89	23
9	86	21	0	89	23
10	86	21	0	89	23
11	86	21	0	89	23
12	86	21	0	89	23
13	86	21	0	89	23
14	86	21	0	89	23

Figure 2a. Snippet of a stabilometric time series.

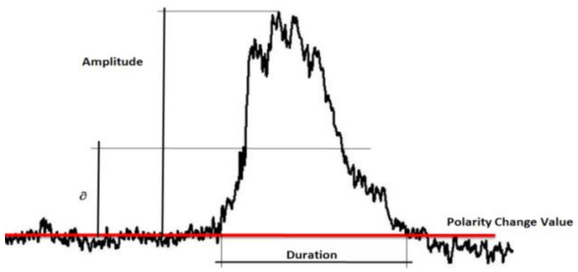

Figure 2b. Example of an identified and characterized fall event.

For each test, information is stored about patient age, height, sport (BASKETBALL, SKATING or CG, control group) and gender. Each attribute will be the **dependent variable** in one of the experiments run.

A series of data pre-processing tasks were performed on the original data [17]: **T1**. *Age* attribute discretization (transforming a quantitative attribute into an ordinal qualitative attribute) and numeration (associating a numerical value with each of the qualitative values taken by the original variable): 0 if Age > 20 and 1 if Age ≤ 20; **T2**. *Height* attribute discretization and numeration: 0 if Height > 170 and 1 if Height ≤ 170; **T3**. *Sport* attribute numeration: 0 Basketball, 1 Skating. The control group individuals are omitted for this attribute; **T4**. *Gender* attribute numeration: 0 Male, 1 Female; **T5**. *Sportsperson* attributization (creating a new attribute from the values of other existing attribute(s)): 0 Sportsperson, 1 not Sportsperson; **T6**. *Skater* attributization: 1 Skater, 0 not Skater. Basketball players were omitted for this attribute, as the aim is to distinguish skaters from the control group; **T7**. *Basketball Player* attributization: 1 Basketball Player, 0 not Basketball player. Skaters were omitted for this attribute, as the aim is to distinguish basketball players from the control group. Finally, we divided the data into two subsets of records, one for each of the considered subtypes (*Left* and *Right*).

In this research we have employed decision trees and logistic regression techniques. Decision trees are tree-shaped structures that are used as predictive models in many different areas [26]. To do this, the value of the known attributes of the object is used to move down through the tree (each tree node contains a condition on known attribute values which determines the branch to be taken) to a leaf node. The algorithm that we have used in this research is CART [27]. Logistic regression is a technique used in data mining to predict the unknown value of a categorical, particularly a binary (two-valued), variable based on the known values of other numerical variables [28].

The *Age* and *Height* attributes were discretized following the instructions of domain experiments in stabilometry. Experiments without attribute discretization will be conducted as part of future research in order to check whether there is any difference in the results.

4. Data analysis and results

We have five independent variables for the experiments: *Sway_Vel, Falls, Imbalances, Fall_Length* and *Max_Fall_Int.* The dataset also includes another seven variables that will be used as dependent variables in as many experiments. Also, each of these seven experiments will be conducted twice (once for each of the two *Left* and *Right* test subtypes). Therefore a total of 14 experiments, denoted Exp1-Exp14, will be conducted.

Table 1 summarizes these experiments, specifying the respective subtype covered (*Left* or *Right*), dependent variable, and number of samples of each of the classes established by the dependent variable in percentage terms. With regard to the topic addressed in this paper, the experiments considering the status of sportsperson or the practised sporting discipline (Exp 7- Exp 14) are of most interest.

Table 1. Summary of the experiments

			Instances for each class		
Experiment	**Subtype**	**Dependent variable**	**0 (%)**	**1 (%)**	**Total**
Exp1	Left	Age	62.50	37.50	56
Exp2	Right	Age	62.50	37.50	56
Exp3	Left	Gender	67.86	32.14	56
Exp4	Right	Gender	67.86	32.14	56
Exp5	Left	Height	60.71	39.29	56
Exp6	Right	Height	60.71	39.29	56
Exp7	Left	Sport	45.45	54.55	33
Exp8	Right	Sport	45.45	54.55	33
Exp9	Left	Sportsperson	58.93	41.07	56
Exp10	Right	Sportsperson	58.93	41.07	56
Exp11	Left	Skater	56.10	43.90	41
Exp12	Right	Skater	56.10	43.90	41
Exp13	Left	Basketball player	60.53	39.47	38
Exp14	Right	Basketball player	60.53	39.47	38

The resulting models were validated using 10-fold cross validation and the results are shown in Table 2.

Table 2. Results of the experiments in terms of classification accuracy

Experiment	Decision Trees	Logistic Regression
Exp1	55.36%	58.93%
Exp2	69.64%	69.64%
Exp3	69.64%	80.36%
Exp4	60.71%	64.29%
Exp5	73.21%	85.71%
Exp6	66.07%	75.00%
Exp7	60.61%	**90.91%**
Exp8	81.82%	84.85%
Exp9	60.71%	48.21%
Exp10	39.29%	55.36%
Exp11	56.10%	65.85%
Exp12	56.10%	63.41%
Exp13	63.16%	76.32%
Exp14	47.37%	71.05%
MEAN	**61.41%**	**70.71%**

5. Discussion of Results

Looking at the results we find that logistic regression yields better results in 12 out of the 14 conducted experiments (Exp1, Exp3-8 and Exp10-14), the CART algorithm is better in one (Exp9) and the results of both techniques are similar in one (Exp2). Generally speaking, logistic regression therefore provides better results, outperforming CART by on average 9.3% in terms of accuracy.

On the other hand, looking at logistic regression, we find that the *Left* subtype yields better results in five out of the seven cases (*Gender, Height, Sport, Skater* and *Basketball player* attributes), whereas the *Right* subtype is better in two out of the seven cases (*Age* and *Sportsperson* attributes). This result is probably due to the fact that most of the analysed basketball players are right handed (11 vs. 4), whereas the skater population is less skewed (8 right-handed vs. 10 left-handed skaters).

Considering the global results, the variables that appear to be most related to balance are *Height* and *Sport*. This suggests, for example, that balance is related to the sport played. However, the poor results for the *Sportsperson* variable suggest that there appears to be no relationship between a person being or not being a professional sportsperson and their degree of balance.

The models output in Experiments 7 and 8 (*Sport* variable) could be said to be especially applicable for mapping out the career of young sports talents. These models could be used to propose, depending on balance, a sports discipline at an early age for talented young sportspeople enrolling in high-performance training programmes designed to forge future sports talents. However, the results suggest that height affects balance characteristics, which is an issue that is worth analysing. In this respect, note that the mean height of the samples is 200.3 centimetres for basketball players and

162.3 centimetres for skaters. This is a big difference, and it could be behind the good classification results in Experiments 7 and 8. Therefore, further experiments should be run considering other sports in which there is not such a pronounced mean height difference between the sportspeople from the two groups.

With regard to the above, other authors have published research on talent detection and development in sport [29, 30]. Our research, however, focuses on more practical issues related to talent management. Several lines of research have been opened in this respect [31]. Vaeyens et al. claim that traditional cross-sectional talent identification models are likely to exclude many, especially late maturing, promising children from development programmes due to the dynamic and multidimensional nature of sport talent [31]. Other practical research has focused above all on the soccer field [32, 33]. Other research has addressed other sports like water-polo [34].

Considering its relationship to the work presented here, we should mention research by Mohammed et al., studying which specific morphological and performance measures describe differences between elite and non-elite young handball players [35]. They found that elite players were heavier and had greater muscle circumferences that their non-elite peers. Elite players scored significantly better on strength, speed and agility, and cardiorespiratory endurance, but not on balance.

In this respect, this paper appears to fit in with our results in the sense that balance does not appear to differ substantially when comparing elite and non-elite sportspeople (Experiments 9 and 10). According to our results, however, balance is applicable in the field of talent management, as it is related to the practised sporting discipline (Experiments 7 and 8). These results are applicable for matching sports talents to the most suited discipline depending on their balance control.

In this regard, the research presented here is, to the best of our knowledge, the first to address the topic of talent mapping and is one of very few to date that has established balance as the discriminating factor. As far as we are aware, it is also the first paper to use data mining techniques (classification in this case) for sports talent management based on balance indicators.

6. Conclusions and Future Work

The stabilometric data acquired after examining a particular person can provide an enormous amount of information concerning their balance and postural control. In this article, we described the experiments conducted based on the stabilometric data of a series of individuals, some of whom were elite sportspeople whereas others were members of a control group.

The results of our experiments can be used to discover interesting knowledge about existing relationships between people's balance and other characteristics. A prominent finding of this research is the close relationship between people's balance and the sport that sportspeople play. This is applicable in practice, where the models can be applied in order to determine the best sport or sports discipline for each subject depending on their balance. This is very useful for state programmes for capturing young sports talents. Interestingly, the experiments did not reveal any significant relationship between balance and the fact that a person does or does not practise sport professionally. Some possible future research lines are:

- Considering that the experiments were conducted on a very small sample of sportspeople and significance is limited, it would be worth broadening the sample in order to conduct a richer and more comprehensive analysis.
- Another future line of work is to consider other stabilometric tests to check whether they confirm the results of the US test. It would also be interesting to add other dependent variables to the analysis, such as right- or left-handedness.
- In view of the results, it would also be worthwhile gathering stabilometric data on elite professional sportspeople who play other sports apart from basketball and skating. It would be very useful to consider sports that are more alike (for example, compare basketball players with handball or volleyball players). Classification accuracy can be expected to drop in this case, although further experiments should be run in order to confirm that there is a drop and by how much.
- It would be a good idea to use other techniques in order to analyse the correlations between the analysed data: clustering, plots, SVMs.
- Although not a central issue to this paper, there appears, according to the experiments, to be a close relationship between balance and gender. One line of research would be to study this relationship and check whether it holds for other datasets, tests, sports, etc. Also it would be interesting to further analyse the relationship between balance and age or height because postural control varies with age or is dependent on height.
- Again it would be worth analysing the causality between balance and sports discipline in order to confirm whether balance determines the sport for which a player is best suited or balance is the result of training for each sport. In this paper, elite skaters clearly had much better balance.

References

1. Barigant, P., Merlet, P., Orfait, J., Tetar, C., New design of E.L.A. Statokinesemeter. Agressol, 13(C): pp. 69-74, 1972.
2. Boniver, R., Posture et posturographie. Rev Med Liege. May 1, 49(5): pp. 285-290, 1994.
3. Sanz, R., Test vestibular de autorrotación y posturografía dinámica. Verteré, 25: pp. 5-15, 2000.
4. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., From Data Mining To Knowledge Discovery: An Overview. In Advances In Knowledge Discovery And Data Mining, eds. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, AAAI Press/The MIT Press, Menlo Park, CA., pp. 1-34, 1996.
5. Ronda, J.M., Galvañ, B., Monerris, E., Ballester, F., Asociación entre Síntomas Clínicos y Resultados de la Posturografía Computerizada Dinámica. Acta Otorrinolaringología Española, 53: pp. 252-255, 2002.

6. Rama, J., Pérez, N., Artículos de Revisión: “Pruebas vestibulares y Posturografía”. Revista Médica de la Universidad de Navarra, vol. 47, nº 4, pp. 21-28, 2003.
7. Barona, R., Interés clínico del sistema NedSVE/IBV en el diagnóstico y valoración de las alteraciones del equilibrio. Revista de Biomecánica del Instituto de Biomecánica de Valencia (IBV), Ed. February, 2003.
8. Lázaro, M., Cuesta, F., León, A., Sánchez, C., Feijoo, R., Montiel, M., Valor de la posturografía en ancianos con caídas de repetición. Med. Clin., Barcelona, pp. 124:207-10, 2005.
9. Nguyen, D., Pongchaiyakul, C., Center, J. R., Eisman, J. A., Nguyen, T. V., Identification of High-Risk Individuals for Hip Fracture: A 14-Year Prospective Study. Journal of Bone and Mineral Research, 20(11), 2005.
10. Raiva, V., Wannasetta, W., Gulsatitporn, S., Postural stability and dynamic balance in Thai community dwelling adults. Chula Med J 2005 Mar, 49(3): pp. 129 – 141, 2005.
11. Sinaki, M., Brey, R. H., Hughes, C. A., Larson, D. R., Kaufman, K. R., Significant Reduction in Risk of Falls and Back Pain in Osteoporotic-Kyphotic Women Through a Spinal Proprioceptive Extension Exercise Dynamic (SPEED) Program. Mayo Clin., 80(7): pp. 849-855, 2005.
12. Song, D., Chung, F., Wong, J., Yogendran, S., The Assessment of Postural Stability After Ambulatory Anesthesia: A Comparison of Desflurane with Propofol. Anesth. Analg., 2002.
13. Martín, E., Barona, R., Vértigo paroxístico benigno infantil: categorización y comparación con el vértigo posicional paroxístico benigno del adulto. Acta Otorrinolaringología Española, 58(7): pp. 296-301, 2007.
14. BM Neurocom® International, Balance Master Operator’s Manual v8.2. www.onbalance.com (accessed in October 2010), 2004.
15. Liston, R. A., Brouwer, B. J., Reliability and validity of measures obtained from stroke patients using the Balance Master. Arch. Phys. Med. Rehabil., 77: pp. 425–430, 1996.
16. Brouwer, B., Culbam, E. G., Liston, R. A., Grant, T., Normal variability of postural measures: implications for the reliability of relative balance performance outcomes. Scand J Rehabil. Med., 30: pp. 131–137, 1998.
17. Lara, J. A., Manual de Minería de Datos, Ed. Udima, 2013.
18. Bonato, P., Sherrill, D. M., Standaert, D. G., Salles, S. S., Akay, M., Data Mining Techniques to Detect Motor Fluctuations in Parkinson's Disease, Proc. of the 26th Annual International Conference on the IEEE Engineering in Medicine and Biology Society, pp. 4766 – 4769, 2004.
19. Allain, H., Bentué-Ferrer, D., Polard, E., Akwa, Y., Patat, A., Postural Instability and Consequent Falls and Hip Fractures Associated with Use of Hypnotics in the Elderly, Drugs & Aging 22(9), pp. 749-756, 2005.
20. Anguera, A., Lara, J. A., Lizcano, D., Martínez, M.A., Pazos, J., Sensor-generated Time Series Events: A definition language, Sensors 12(9), pp. 11811-52, 2012.

21. Alonso, F., Lara, J. A., Martínez, L., Pérez, A., Valente, J. P., Generating Reference Models for Structurally Complex Data: Application to the Stabilometry Medical Domain, Methods of Information in Medicine, 52, pp. 441-453. 2013.
22. Lara, J. A., Moreno, G., Pérez, A., Valente, J. P. , López-Illescas, A., Comparing posturographic time series through events detection, 21st IEEE International Symposium on Computer-Based Medical Systems, CBMS '08, pp. 293-295, 2008.
23. Lara, J. A., Pérez, A., Valente, J. P., López-Illescas, A., Generating time series reference models based on event analysis, 19th European Conference on Artificial Intelligence - ECAI 2010, pp. 1115-116, 2010.
24. Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., Riera, T., A UML Profile for the Conceptual Modelling of Structurally Complex Data: Easing Human Effort in the KDD Process, Information and Software Technology 56, pp. 335-351, 2014a.
25. Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., Data preparation for KDD through automatic reasoning based on description logic, Information Systems 44, pp.54-72, 2014b.
26. Huo, X., Kim, S. B., Tsui, K.-L., & Wang, S. A frontier-based tree pruning algorithm (FBP). INFORMS Journal on Computing, 18, 494–505, 2006.
27. Hastie, T., Tibshirani, R., Friedman, J. The element of statistical learning. New York, NY: Springer, 2001.
28. James, G., Witten, D., Hastie, T., Tibshirani, R., An Introduction to Statistical Learning, Springer, 2013.
29. Morris, T., Psychological characteristics and talent identification in soccer. Journal of Sport Sciences, 18:9, 715-726, 2000.
30. Abbott, A., Collins, D., Eliminating the dichotomy between theory and practice in talent identification and development: considering the role of psychology. Journal of Sport Sciences, 22:5, 395-408, 2004.
31. Vaeyens, R., Lenoir, M., Williams, A. M., Philippaerts, R. M., Talent identification and development programmes in sport. Sports Med, 38:9, 703-714, 2008.
32. Reilly, T., Williams, A. M., Nevill, A., Franks, A., A multidisciplinary approach to talent identification in soccer. Journal of Sport Sciences, 18:9, 695-702, 2000.
33. Williams, A. M., Perceptual skill in soccer: Implications for talent identification and development. Journal of Sport Sciences, 18:9, 737-750, 2000.
34. Falk, B., Lidor, R., Lander, Y., Lang, B., Talent identification and early development of elite water-polo players: a 2-year follow-up study, Journal of Sport Sciences, 22:4, 347-355, 2004.
35. Mohamed, H., Vaeyens, R., Matthys, S., Multael, M., Lefevre, J., Lenoir, M., Philippaerts, R., Anthropometric and performance measures for the development of a talent detection and identification model in youth handball, Journal of Sport Sciences, 27:3, 257-266, 2009.