



ТЕОРИЯ И МЕТОДИКА СПОРТА/THEORY AND METHODOLOGY OF SPORTS

DOI: <https://doi.org/10.60797/IRJ.2026.165.74> EDN: UVYHOK

APPLYING THE K-SHAPE CLUSTERING ALGORITHM TO SPORTS TACTICS ANALYSIS: A USE-CASE IN FINSWIMMING

Research article

Bogdashkin A.Y.^{1,*}, Morozov S.N.², Saurov Y.A.³^{1,2,3} Russian University of Sports (SCOLIPE), Moscow, Russian Federation

* Corresponding author (swim.alex[at]mail.ru)

Abstract

This article provides the effectiveness of the application of the k-Shape clustering algorithm to the tactics analysis in endurance sports, such as swimming, skiing, cycling, etc. To popularize the sport of finswimming, this study is conducted on one of the events of this sport. The most and the least efficient tactical schemes for 400 meters distances in finswimming for men and women were found. To achieve that, we apply an effective and modern clustering algorithm named k-Shape. The cross-correlation measure is used as a distance measure between event splits organized as time series. Our study shows that the most productive tactics at 400 meters disciplines in finswimming represent even pacing strategy with fast 1st lap, gradual slowdown from 2nd to 3rd lap, constant speed from 4th to 7th and speed-up at the 8th lap.

Keywords: finswimming, k-shape, clustering, sport results, tactics, competitive activity.**ПРИМЕНЕНИЕ АЛГОРИТМА КЛАСТЕРИЗАЦИИ K-SHAPE В АНАЛИЗЕ СПОРТИВНЫХ ТАКТИК: ПРИМЕР ИСПОЛЬЗОВАНИЯ НА ПЛАВАНИИ В ЛАСТАХ**

Научная статья

Богдашкин А.Е.^{1,*}, Морозов С.Н.², Сауров Е.А.³^{1,2,3} Российский университет спорта «ГЦОЛИФК», Москва, Российская Федерация

* Корреспондирующий автор (swim.alex[at]mail.ru)

Аннотация

В данной статье представлены результаты эффективности применения алгоритма кластеризации k-Shape для анализа тактики в силовых видах спорта, таких как плавание, лыжи, велоспорт и т. д. С целью популяризации плавания в ластах данное исследование проводится на одном из соревнований по этому виду спорта. Были выявлены наиболее и наименее эффективные тактические схемы для дистанций 400 метров в плавании в ластах для мужчин и женщин. Для этого мы применяем эффективный и современный алгоритм кластеризации под названием k-Shape. В качестве меры расстояния между отрезками соревнований, упорядоченными в виде временных рядов, используется мера кросс-корреляции. Наше исследование показывает, что наиболее продуктивная тактика на дистанции 400 метров в плавании в ластах представляет собой стратегию равномерного темпа с быстрым первым кругом, постепенным замедлением со второго по третий круг, постоянной скоростью с четвертого по седьмой круг и ускорением на восьмом круге.

Ключевые слова: плавание в ластах, k-shape, кластеризация, спортивные результаты, тактика, соревновательная деятельность.**Introduction**

Tactics is an essential part of the training process and competitions at any level. Different groups of sport arts have different tactical characteristics [1], [4]. Endurance sport arts have time-trial [9] competitions and represent repetitions of similar actions during a race. Competition rules restrict the variety of technical and tactical actions. Therefore, it may misrepresent the absence of necessity of the integral preparation for competitions. Unfortunately, the sports research community don't investigate this topic enough.

Tactics in endurance sport arts (skating, running, finswimming, swimming, cycling and others) is mostly considered as pacing [11]. Athletes aim to choose the most suitable race pace based on their individual skills and theoretical knowledge. Analysis of tactics in this group of sports might be seen as the analysis of time series [8]: athletes repetitively execute standardized laps several times during a competition race which makes it convenient to find similar patterns in the data.

The most appropriate way to analyze tactics in cyclic sport is to use diverse methods of applied statistics. The discriminant method is the most popular way in research of this kind [7]. This method assumes distribution of data into formerly created classes. This method has many disadvantages: results largely rely on researcher's opinion and on the rules that they create the classes with; it is very time consuming; it is not scalable [6].

Swimming tactics pieces of research are also usually done with the discriminant method. For example, a group of British scientists [10] studied tactics in 400 meters freestyle, comparing tactical schemes with six standard options described in another study [1]: parabolic, fast-start-even, positive, negative, even, variable. The comparison showed that elite swimmers used only three out of six schemes: parabolic, fast-start-even and positive, while the athletes preferred tactical schemes with a quick start and a parabolic change in speed, but none of the tactical schemes had a significant impact on the final result [10]. Disadvantages of this study are the same: it is highly subjective, time consuming, and carries high risk of producing incorrect data.

A more appropriate and scientific way of analyzing time series is clustering. Studies of this kind in finswimming were not found.

Research methods and principles

2.1. Clustering

Clustering aims to find homogeneous groups in unlabeled data [3]. Over the past decades, a lot of different approaches for measuring similarities were introduced. Among those, we can name shape-based, feature-based and model-based approaches [8]. In the shape-based approach, shapes of two time-series traces are matched as well as possible by a non-linear stretching, e. g. utilizing the DTW distance. In the feature-based method, raw data are transformed into a lower-dimensional representation. Some examples of this transformation might be value-based principal component analysis [13] or singular value decomposition [6]. In model-based approaches, raw data are used to train a model. Then, conventional clustering algorithms are applied to the residuals of the fitted model [14].

In this study, we leverage one of the shape-based clustering algorithms, and namely k-Shape [12]. In comparison to other approaches, this algorithm offers a scalable and interpretable solution to the time series clustering problem. Indeed, k-Shape has time complexity $O(\max(nm^2, m^3))$, but a significantly better accuracy in comparison to k-Means [8].

2.2. Frameworks

All parts of this software were developed in Python 3.7.8. For data preprocessing and scaling (MinMaxScaler), pandas 1.4.2, numpy 1.21.0 and scikit-learn 1.1 were used. For clustering, a built-int implementation of k-Shape of tslearn 0.5.2 was utilized. All experiments were run on a laptop with a "Intel Core i7 CPU @2.8 GHz" processor and 16 GB RAM.

Main results

We analyzed 594 results at 400 meters surface, bifins and immersion disciplines for men and women at top world competitions from 2010 to 2021 years during this research. Time for 50 meters laps were calculated for every performance. We performed k-Shape clustering for $k = 2, 3, 4$, and 5. We found that $k = 2$ and 3 gave us the most distinct shapes for different disciplines. Each performance was assigned a number of points using the ranking system developed earlier [2]. All the performances were ranged into four groups divided by three quantiles. Productivity of every tactical scheme was evaluated by the percentage of its results that are in the range higher than 0,75 quantile (table 1).

Table 1 - Frequency of use of different tactical schemes at 400 meters disciplines in finswimming

DOI: <https://doi.org/10.60797/IRJ.2026.165.74.1>

№	Number of points								Overall	
	0,75 quantile		0,75-0,5 quantile		0,5-0,25 quantile		0,25 quantile			
	n	%	n	%	n	%	n	%	n	%
400 meters surface men										
1	27	32,93	24	29,27	19	23,17	12	14,63	82	57,75
2	8	14,81	9	16,67	14	25,93	23	42,59	54	38,03
3	0	0	3	50	2	33,33	1	16,67	6	4,23
400 meters surface women										
1	30	25	29	24,17	31	25,83	30	25	120	88,24
2	4	25	5	31,25	3	18,75	4	25	16	11,76
400 meters bifins men										
1	11	27,5	10	25	10	25	9	22,5	40	64,52
2	4	18,18	6	27,27	5	22,73	5	22,73	22	34,48
400 meters bifins women										
1	12	27,27	10	22,73	11	25	11	25	44	73,33
2	3	18,75	5	31,25	4	25	4	25	16	26,67
400 immersion men										
1	20	27,4	21	28,77	16	21,92	16	21,92	73	72,28
2	6	21,43	4	14,29	9	31,14	9	31,14	28	27,72
400 meters immersion women										
1	9	36	3	12	4	16	9	36	25	26,88
2	3	17,65	6	35,29	4	23,53	4	23,53	17	18,28
3	12	23,53	14	27,45	15	29,41	10	19,61	51	54,84

Note: $n=594$

It was found that the most productive tactical schemes represent parabolic pacing strategy (picture 1). All the athletes swam the 1st lap faster than others because of the speed-up after jump off the blocks, the big amount of phosphocreatine and



glycogen in the muscles and high level of adrenaline in the blood due to the pre-competition stress. At the 2nd to 3rd lap the phosphocreatine level is very low while glycogen level is decreasing which leads to a gradual slowdown. Starting from the 4th lap the less powerful aerobic pathways take a bigger role in the ATP resynthesis which causes a slower but even speed. At the last 50 meters athletes go "all-in" and try to spend the rest of their energy. *The least productive schemes represent constant slowdown from start to finish or significant slowdown at 7th lap and speed-up at 8th lap (picture 2). For the better visualization comparison, all the results were scaled from 0 to 1.*

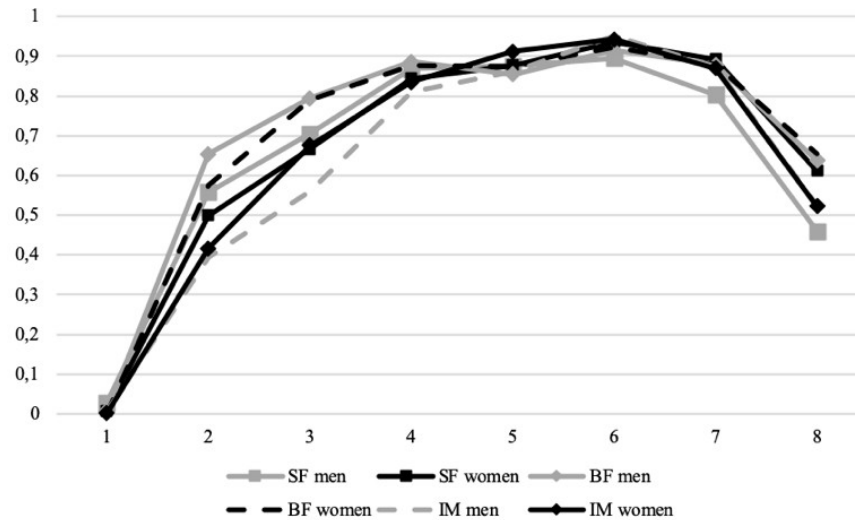


Figure 1 - The most productive tactical schemes at 400 meters distances for men and women
DOI: <https://doi.org/10.60797/IRJ.2026.165.74.2>

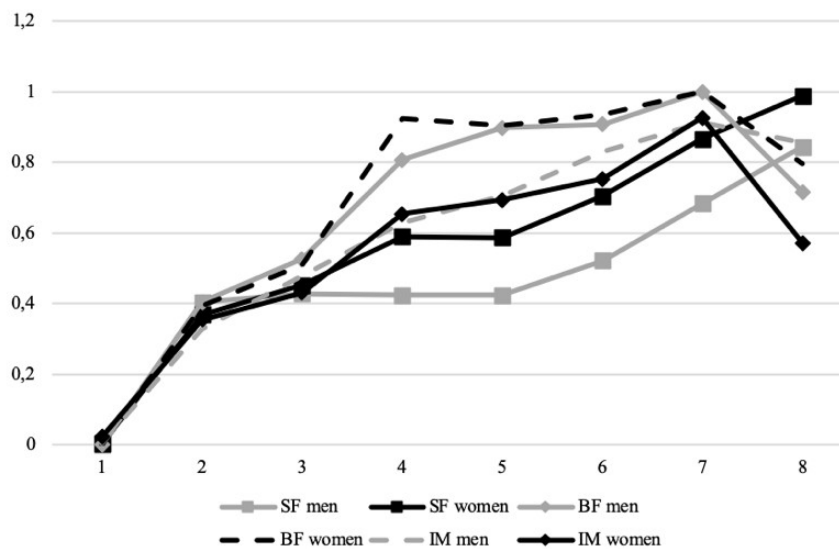


Figure 2 - The least productive tactical schemes at 400 meters distances for men and women
DOI: <https://doi.org/10.60797/IRJ.2026.165.74.3>

Discussion

In this study, different pacing strategies were extracted by the applying a state-of-the-art data-driven approach. The similar shape-based groups were found by the utilization of k-Shape algorithm. As shown in the literature, this approach is used for natural signals in finance, medicine, biology and many other fields [13]. However, to the best of our knowledge, this approach has never been applied to the sports tactics analysis. It should be admitted that the new approach proved to be superior to the previously applied methods such as the discriminant method, which might be characterized as subjective and strongly affected by the initialization. These drawbacks have been overcome by the state-of-the-art data-driven algorithm of k-Shape clustering. However, this algorithm was only applied to middle-distance and, certainly, can be extended and used for long distances.



Unfortunately, this method is not suitable for sprint distances. The future pieces of research should consider expansion of clustering algorithms for short distances. This will be the objective of our future studies.

Conclusion

K-Shape clustering algorithm is proven to be an appropriate and modern method to analyze tactic patterns in cyclic sports. This method allows to avoid subjectivity and provides more scientific and valid data than others. The most productive tactics at 400 meters disciplines in finswimming represent even pacing strategy *with* fast 1st lap, gradual slowdown from 2nd to 3rd lap, constant speed from 4th to 7th and speed-up at the 8th lap.

Конфликт интересов

Не указан.

Рецензия

Сообщество рецензентов Международного научно-исследовательского журнала
DOI: <https://doi.org/10.60797/IRJ.2026.165.74.4>

Conflict of Interest

None declared.

Review

International Research Journal Reviewers Community
DOI: <https://doi.org/10.60797/IRJ.2026.165.74.4>

Список литературы на английском языке / References in English

1. Abbiss C. Describing and understanding pacing strategies during athletic competition / C. Abbiss, P. Laursen // Sports Medicine. — 2008. — № 38. — P. 239–252.
2. Bogdashkin A. Development of a ranking system in finswimming as a pedagogical tool for the comparison of athletic performance / A. Bogdashkin, S. Morozov, E. Saurov // International Research Journal. — 2022. — № 119. — P. 27–30.
3. Chung F.L. Flexible time series pattern matching based on perceptually important points / F.L. Chung // Conference on Artificial Intelligence Workshop. — 2001. — № 1. — P. 1–7.
4. De Koning J.J. Using modeling to understand how athletes in different disciplines solve the same problem: swimming versus running versus speed skating / J.J. De Koning, C. Foster, A. Lucia // International Journal of Sports Physiology and Performance. — 2011. — № 6. — P. 276–280.
5. Faloutsos C. Fast subsequence matching in time-series databases / C. Faloutsos, M. Ranganathan // ACM SIGMOD Rec. — 1994. — № 1. — P. 419–429.
6. Filippone M.F. A survey of kernel and spectral methods for clustering / M.F. Filippone, F. Camastra, F. Masulli // Pattern Recognition. — 2008. — № 41 (1). — P. 176–190.
7. Kaufman L. Finding groups in data: An introduction to cluster analysis / L. Kaufman, P.J. Rousseeuw. — New York: John Wiley & Sons, 2009. — 342 p.
8. Meesrikamolku W. . Shape-based clustering for time series data / W. Meesrikamolku, V. Niennattrakul // Advances in Knowledge Discovery and Data Mining. — 2012. — № 3. — P. 530–541.
9. Konings M.J. Pacing Decision Making in Sport and the Effects of Interpersonal Competition: A Critical Review / M.J. Konings, F.J. Hettinga // Sports Medicine. — 2018. — № 48 (11). — P. 1829–1843.
10. Mauger A.R. Of pacing strategy selection in elite 400-m freestyle swimming / A.R. Mauger, J. Neuloh, P.C. Castle // Medicine and Science in Sports and Exercise. — 2012. — № 44 (11). — P. 2205–2212.
11. McGibbon K.E. Pacing in swimming : A systematic review / K.E. McGibbon, D.P. Pyne, M.E. Shephard et al. // Sports Medicine. — 2018. — № 48 (7). — P. 1621–1633.
12. Paparrizos J. K-shape: Efficient and accurate clustering of time series / J. Paparrizos, L. Gravano // Proceedings of the 2015 ACM SIGMOD international conference on management of data. — 2015. — № 1. — P. 1855–1870.
13. Shumway R.H.R. Time-frequency clustering and discriminant analysis / R.H.R. Shumway // Stat. Prob. Lett.. — 2003. — № 3. — P. 307–314.
14. Li X. Linear time complexity time series clustering with symbolic pattern forest / X. Li // Proceedings of the Twenty-eighth International Joint Conference on Artificial Intelligence (IJCAI-19). — 2019. — № 4. — P. 190–205.