

Représentation symbolique de séries temporelles cycliques basée sur les propriétés des cycles : application à la biomécanique

Vanel Steve Siyou Fotso¹

Engelbert Mephu-Nguifo¹

Philippe Vaslin¹

¹ Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes (LIMOS)

Clermont Université, Université Blaise Pascal, LIMOS,
BP 10448, F-63000 CLERMONT-FERRAND CNRS, UMR 6158, LIMOS, F-63173 AUBIERE
siyou@isima.fr, mephu@isima.fr, vaslin@isima.fr

Résumé

L'analyse des séries temporelles cycliques issues de la biomécanique repose sur la comparaison des propriétés de leurs cycles. Les algorithmes usuels de classification de séries temporelles ignorant cette particularité, nous proposons une représentation symbolique de séries temporelles cycliques basée sur les propriétés de leurs cycles et nommée SAX-P. Les chaînes de caractères ainsi obtenues peuvent ensuite être comparées à l'aide de la distance Dynamic Time Warping. L'application de SAX-P aux moments propulsifs de trois sujets (S1, S2, S3) se déplaçant en Fauteuil Roulant Manuel a permis de mettre en évidence la dissymétrie de leur propulsion en ligne droite. La représentation symbolique SAX-P facilite la lecture des séries temporelles cycliques et l'interprétation des résultats de leur classification par les cliniciens.

Mots Clef

Séries temporelles cycliques, Représentation symbolique, SAX, Biomécanique, Fauteuil Roulant Manuel.

Abstract

The analysis of cyclic time series from biomechanics is based on the comparison of the properties of their cycles. As usual algorithms of time series classification ignore this particularity, we propose a symbolic representation of cyclic time series based on the properties of cycles, named SAX-P. The resulting character strings can be compared using the Dynamic Time Warping distance. The application of SAX-P to propulsive moments of three subjects (S1, S2, S3) moving in Manual Wheelchair highlight the asymmetry of their propulsion. The symbolic representation SAX-P facilitates the reading of the cyclic time series and the clinical interpretation of the classification results.

Keywords

cyclic time series, symbolic representation, SAX, Biomechanics, Manual Wheelchair

1 Introduction

D'une manière générale, pour se déplacer, l'être humain effectue des mouvements cycliques (ex : marche, course, natation, cyclisme). L'analyse biomécanique de ces mouvements est réalisée avec différents instruments de mesure (ex : capteurs de force et d'accélération, systèmes d'analyse cinématique) qui permettent l'enregistrement en continu, sur de longues périodes, de nombreuses grandeurs mécaniques (cinématiques et dynamiques). Ces enregistrements produisent de longues séries temporelles composées de nombreux cycles, ou patrons, représentatifs des mouvements réalisés et des efforts produits par l'individu au cours de ses déplacements (Figure 2).

Ces cycles constituent les unités d'analyse des séries temporelles et possèdent plusieurs propriétés caractéristiques telles que : la valeur minimale, la surface sous le cycle, la durée du cycle [1]. Plusieurs études antérieures ont comparé des séries temporelles en les décomposant en petits segments puis en comparant les propriétés de ces segments. Un segment d'une série temporelle est une suite de valeurs consécutives appartenant à celle-ci [2].

Keogh et Chakrabarti [3] ont proposé de remplacer chaque segment $x_{\frac{n}{N}(i-1)+1}, x_{\frac{n}{N}(i-1)+2}, \dots, x_{\frac{n}{N}i}$ d'une série temporelle $X = x_1, x_2, \dots, x_n$ par la moyenne \bar{x}_i de ses valeurs ($\bar{x}_i = \frac{N}{n} \sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} x_j$) transformant ainsi la série temporelle, qui est une suite de valeurs, en la suite des moyennes de ses N segments $\bar{X} = \bar{x}_1, \bar{x}_2, \dots, \bar{x}_N$. Cette méthode, connue sous le nom Piecewise Aggregate Approximation (PAA), avait pour objectif premier de réduire la taille des séries temporelles. Cependant, elle permet également de comparer deux séries temporelles C et Q en calculant la distance DR entre les suites \bar{C} et \bar{Q} des moyennes leurs segments (équation 1).

$$DR(\bar{C}, \bar{Q}) = \sqrt{\frac{n}{N} \sum_{i=1}^N (\bar{c}_i - \bar{q}_i)^2} \quad (1)$$

Lin et al. [4] se sont basés sur la méthode PAA pour pro-

poser une représentation symbolique des séries temporelles baptisée Symbolic Aggregate approXimation (SAX). L'objectif de SAX est d'assigner de manière équiprobable une lettre à chaque segment. Pour ce faire, le domaine des valeurs de la série temporelle est divisé en intervalles de sorte que chaque point de la série temporelle ait approximativement la même probabilité d'appartenir à un intervalle et une lettre est associée à chacun de ces intervalles. Ensuite, chaque segment de la série temporelle est associé à la lettre de l'intervalle auquel appartient sa moyenne. Avec SAX, la différence entre deux chaînes de caractères est calculée par la distance entre les bornes des intervalles que représente chaque caractère de la chaîne (équation 2).

$$MINDIST(\hat{Q}, \hat{C}) = \sqrt{\frac{n}{N} \sum_{i=1}^N (dist(\hat{q}_i, \hat{c}_i))^2} \quad (2)$$

\hat{q}_i et \hat{c}_i sont des caractères et $dist()$ est la distance entre les bornes des intervalles que représentent ces caractères [4]. Cependant, la moyenne seule n'est pas suffisante pour associer un segment à une lettre. En effet, des segments peuvent avoir la même moyenne et être associés à la même lettre alors qu'ils ont des allures très différentes.

Pour y remédier, Lkhagva et Kawagoe [5] ont proposé le modèle ESAX qui prend en compte trois propriétés de chaque segment : sa moyenne, son minimum et son maximum. Par la suite, Sun et al. [6] ont proposé le modèle SAX-TD qui prend en compte deux propriétés pour chaque segment : la moyenne et la tendance. Ils adaptent ensuite la distance utilisée par la méthode SAX pour qu'elle prenne en compte la tendance.

Ces deux dernières méthodes fournissent de meilleurs résultats que la méthode SAX [6]. Elles ont toutefois l'inconvénient d'augmenter le nombre de symboles nécessaires à la représentation des séries temporelles. En effet, la méthode ESAX triple la taille de la représentation d'une série temporelle fournie par la méthode SAX, tandis que la méthode SAX-TD la double. De plus, les quatre méthodes précédentes ont deux inconvénients majeurs : elles considèrent des segments de taille fixe, alors que les cycles sont des segments de taille variable, et elles ne prennent pas en compte les propriétés caractéristiques des cycles telles que la durée et la surface sous un cycle (Figure 2). Notre objectif est de proposer une représentation symbolique qui prend en compte plusieurs propriétés pour chaque cycle, mais sans accroître le nombre de symboles utilisés pour la représentation symbolique.

Ces nouvelles représentations symboliques permettront en outre l'utilisation d'un grand nombre d'algorithmes d'analyse de séquences tels que les algorithmes de détection d'anomalie (rechercher une sous-séquence inhabituelle dans une séquence), les algorithmes de découverte de motifs (chercher des sous-séquences qui se répètent)[7], les algorithmes de classification supervisée ou non supervisée, et aussi des algorithmes provenant de l'analyse de texte ou de la communauté de bio-informatique [8] [9].

2 SAX-P : Nouvelle représentation symbolique de séries temporelles cyclique

Définition 1 : Une série temporelle $TS = t_{s0}, t_{s1}, \dots, t_{sm}$ est un ensemble ordonné de variables à valeurs réelles où t_{sm} est la valeur la plus récente.

Définition 2 : Une sous-séquence d'une série temporelle TS de taille m est un segment $T_{si,l}$ de taille $l < m$ constitué de variables consécutives commençant à la position i . Formellement, $T_{si,l} = t_{si}, t_{si+1}, \dots, t_{si+l-1}$

Problème : Étant donnée une série temporelle cyclique T_S et une fonction F qui la divise en k segments consécutifs et disjoints (ici les segments sont des cycles) telle que $F(T_S) = T_{s0,l1}, T_{s1,l2}, \dots, T_{sm-lk,lk}$. Chaque segment $T_{si,li}$ est caractérisé par n propriétés dont l'ensemble est $P(T_{si,li})$. Notre problème est de comparer deux séries temporelles T_S et T_{S0} en se basant sur les propriétés des segments de $F(T_S)$ et de $F(T_{S0})$.

Le principe de SAX-P est le suivant (Tableau 1) :

1. **Découper la série temporelle en segments correspondant aux cycles successifs :** Pour cela, nous définissons un seuil tel que lorsqu'une valeur de la série temporelle est inférieure ou égale au seuil, alors nous revenons en arrière pour trouver un point proche de zéro correspondant à l'instant de début du cycle. Un cycle est constitué de tous les points situés entre deux débuts de cycles consécutifs ou de tous les points situés après le dernier début de cycle.
2. **Calculer un ensemble de propriétés pour chaque segment.** Pour l'analyse des séries temporelles cycliques, nous calculons les propriétés suivantes : la moyenne des valeurs et leur écart type, la valeur minimale, la durée du cycle et la surface sous un cycle. D'autres propriétés pourront toutefois être ajoutées en fonction de l'application visée.
3. **Regrouper dans les mêmes classes les segments ayant des propriétés similaires.** Comme les propriétés ont des unités différentes, le vecteur des propriétés associé à chaque cycle de propulsion a été centré et réduit. A partir de ce vecteur, les cycles de propulsion ont été regroupés en classes grâce à un algorithme de classification non supervisée. Nous avons choisi l'algorithme k-moyennes car il peut traiter efficacement de grands jeux de données [10].
4. **Remplacer chaque cycle de propulsion par la lettre de la classe à laquelle il appartient.** Nous avons ensuite associé une lettre à chaque classe ainsi obtenue et remplacé chaque cycle par la lettre de la classe à laquelle il appartient. Cette dernière étape nous a permis d'obtenir une chaîne de caractères représentant une série temporelle.

La figure 1 illustre le fonctionnement de SAX-P.

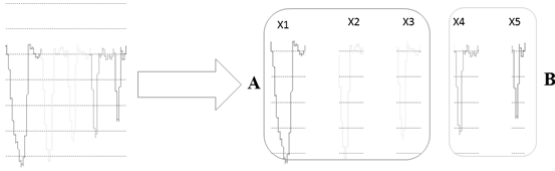


FIGURE 1 – Une série temporelle cyclique est segmentée en 5 cycles de propulsion (X1, X2, X3, X4, X5). Pour chaque cycle, un ensemble de propriétés a été calculé et les cycles ayant des propriétés similaires ont été regroupés dans une même classe (A ou B). La série temporelle initiale est ainsi transformée en une chaîne de caractères (ici : AAABB).

Algorithme 1. SAX-P(D, S, K)

Données :

D : ensemble de séries temporelles cycliques

S : Seuil de détection des cycles

K : Le nombre de lettres ou de classes à former

Résultats : CHAINES ; Représentations symboliques des séries temporelles cycliques de D

1. cycles = F(D, S) ;
2. proprietesCycles = centrerReduire(P(cycles)) ;
3. classes = kMoyennes(K, proprietesCycles) ;
4. CHAINES = representationSymbolique(cycles, classes) ;
5. retourner CHAINES ;

TABLE 1 – Pseudo-code de SAX-P

Après avoir appliqué la méthode SAX-P à des séries temporelles cycliques, nous avons obtenu des représentations symboliques qu'il restait à comparer. La comparaison de deux chaînes de caractères consiste à comparer les caractères qui les constituent, ce qui revient à comparer les classes entre elles. Cette comparaison est effectuée à l'aide de la distance euclidienne entre les vecteurs moyens des propriétés de chaque classe. La comparaison entre les chaînes de caractères est faite à partir d'une distance basée sur la déformation temporelle dynamique, ou Dynamic Time Warping (DTW) [11]. En effet, l'utilisation de DTW associée à l'algorithme des K plus proches voisins est une méthode bien connue et efficace de classification de séries temporelles [12] [13].

3 Application à la locomotion en Fauteuil Roulant Manuel

Le modèle SAX-P a été appliqué aux séries temporelles obtenues par l'enregistrement de moments axiaux exercés sur les roues arrière droite (RD) et gauche (RG) d'un Fauteuil Roulant Manuel (FRM) instrumenté par trois sujets (S1, S2, S3) ayant effectué chacun cinq déplacements rectilignes. Dans un premier temps, les séries temporelles ont été segmentées en cycles de propulsion. Puis, pour chaque cycle de propulsion, neuf propriétés ont été calculées :

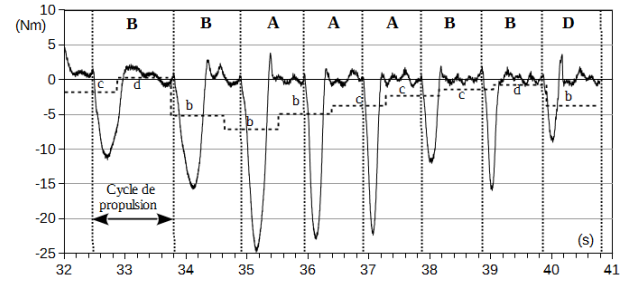


FIGURE 2 – Exemple de série temporelle cyclique : Moment propulsif appliqué par un sujet à la roue arrière d'un Fauteuil Roulant Manuel pendant un déplacement rectiligne. Dans ce cas, le patron est appelé cycle de propulsion. Les barres verticales délimitent les cycles de propulsion (segments) identifiés pendant les déplacements, et les lettres majuscules au-dessus de chaque cycle indiquent les classes auxquelles ils appartiennent. L'application de SAX à la série temporelle la divise en dix intervalles (lettres minuscules) de taille égale, mais sans prendre en compte les cycles de propulsion, l'aire sous le cycle, le temps de cycle et le temps de poussée ne sont pas pris en compte par SAX.

- Des propriétés directement utiles à l'analyse : les valeurs minimale et maximale du moment propulsif, la durée d'un cycle, la durée d'une poussée, la surface sous un cycle ;
- Des propriétés statistiques : la moyenne, la médiane, l'écart type, l'écart interquartile. Puis nous avons regroupé les cycles de propulsion dans 5 classes (A, B, C, D, E), dont les vecteurs moyens des propriétés sont présentés dans le Tableau 2.

En remplaçant chaque cycle de propulsion par la lettre de la classe à laquelle il appartient (Figure 2), on obtient les chaînes de caractères présentées dans le Tableau 3.

L'étude biomécanique des mouvements humains au cours d'activités telles que la marche, la course, le cyclisme ou encore la locomotion en FRM produit un grand nombre de séries temporelles cycliques qui restent encore généralement inexploitable par les personnes directement concernées en priorité, à savoir les pratiquants et les patients eux-mêmes, mais aussi et surtout les médecins spécialistes en rééducation fonctionnelle et les kinésithérapeutes. Il est donc apparu nécessaire de transformer ces multiples signaux afin de les représenter, sans les dénaturer, sous une forme simplifiée facilitant leur analyse et leur interprétation.

L'analyse des séries temporelles est un domaine de recherche où ont été développées de nombreuses méthodes répondant à cet objectif, telles que : PAA, SAX, ESAX et SAX-TD. Cependant, ces méthodes ne prennent en compte qu'un nombre limité des propriétés des segments qui composent les séries temporelles analysées (moyenne, minimum, maximum, tendance). En outre, ces méthodes sont basées sur des segments de longueur fixe : elles ne sont donc pas adaptées aux séries temporelles cycliques pro-

Classes	A	B	C	D	E
Nb de cycles	18	36	59	18	104
Temps cycle (s)	1.2	1.0	1.0	1.7	0.8
Temps poussée (s)	0.6	0.3	0.4	1.0	0.3
Mz Min (Nm)	-22.3	-17.4	-11.4	-8.7	-6.4
Mz Max (Nm)	0.1	0.1	0.1	0.7	0.1
Moyenne (Nm)	13.6	-8.1	-6.2	-3.0	-3.3
Médiane (Nm)	-16.1	-10.8	-7.4	-4.2	-3.9
IRQ (Nm)	12.4	10.7	6.0	4.3	3.1
Ecart-Type (Nm)	7.1	5.6	3.4	2.6	1.8
Aire (Nm.s)	-7.1	-2.3	-2.2	-1.8	-1.0

TABLE 2 – Vecteurs moyens des propriétés des classes (A, B, C, D, E) utilisées pour la représentation symbolique des moments propulsifs (Mz). SAX-P prend en compte le temps de poussée et le temps de cycle, la surface sous la poussée.

Subject	S1		S2		S3	
Poussée	Drte	Gche	Drte	Gche	Drte	Gche
1	C	A	C	D	E	D
2	B	B	E	E	D	E
3	B	B	C	E	C	E
4	B	B	C	E	E	E
5	B	B	C	D	E	E
6	C	B	E	C		D
7	B	C	E	E		E
8	E		E	E		
9			C	C		
10			E	E		
11			C	E		
12			C	E		
13				E		
DTW	268		354		44	

TABLE 3 – Chaînes de caractères obtenues par la méthode SAX-P appliquée aux séries temporelles des moments propulsifs enregistrés par les capteurs des roues droite (RD) et gauche (RG) d'un fauteuil roulant manuel instrumenté lors de déplacements rectilignes effectués par trois sujets (S1, S2, S3). La distance DTW mesure l'alignement des séquences des roues droite et gauche d'un même sujet pendant un même déplacement : lorsqu'elle est non nulle, elle est une mesure de la dissymétrie de la propulsion du sujet concerné.

duites par l'étude biomécanique de la locomotion humaine et constituées de cycles successifs de durée variable. La méthode SAX-P, présentée ici, permet de segmenter une série temporelle cyclique en cycles successifs et prend en compte simultanément plusieurs propriétés d'un même cycle. Elle se traduit par la représentation symbolique d'une série temporelle cyclique sous la forme d'une chaîne de caractères qui, à la fois, réduit la série temporelle et facilite son interprétation.

Dans le cas de la locomotion en FRM, les séries temporelles cycliques ont été représentées sous la forme de chaînes de caractères qui ont ensuite pu être comparées à l'aide de la méthode DTW. Ces comparaisons ont permis de mettre en évidence que les trois sujets dont les déplacements ont été étudiés ici n'avaient pas un mode de propulsion symétrique (comparaison des cycles de propulsion droits et gauches) et qu'ils avaient des modes de propulsion différents les uns des autres (comparaisons interindividuelles). Les représentations symboliques ainsi obtenues ont un autre avantage ; elles permettent d'utiliser le grand nombre d'algorithmes de traitement de séquences disponibles dans la littérature pour analyser les séries temporelles.

Bien que l'application de SAX-P effectuée ici ne concernent que trois sujets, les résultats de cette analyse sont encourageants car ils démontrent à la fois la pertinence de la méthode proposée et les perspectives qu'elle offre dans le domaine clinique pour évaluer, par exemple, les progrès réalisés par un patient au cours de sa rééducation et tout au long de sa vie, mais aussi pour évaluer les méthodes de rééducation elles-mêmes.

4 Conclusion

Dans ce travail, nous avons proposé une méthode de représentation symbolique des séries temporelles cycliques nommée SAX-P. Cette méthode permet de représenter une série temporelle cyclique sous la forme d'une chaîne de caractères, chaque caractère représentant une classe définie par les propriétés des cycles de la série considérée. Les chaînes de caractères ainsi tenues ont ensuite été comparées à l'aide de la distance Dynamic Time Warping utilisée pour l'alignement des séquences. Le modèle SAX-P a été appliqué aux moments propulsifs mesurés pendant les déplacements en ligne droite de trois personnes en FRM. Les résultats préliminaires obtenus ont notamment permis de démontrer que ces sujets avaient des modes de propulsion différents et que les cycles de propulsion d'un même sujet n'étaient pas toujours symétriques. La suite de ce travail va consister à valider le modèle SAX-P. Pour ce faire, nous envisageons de comparer les méthodes SAX, ESAX, SAX-TD et SAX-P sur une tâche de classification supervisée. Dans des travaux ultérieurs, nous nous intéresserons aux différents paramètres de la méthode SAX-P notemment le choix de l'algorithme de classification des cycles de propulsion à utiliser.

Remerciements

Ce travail a été partiellement financé par le Labex IMOBS3.

Références

- [1] R. J. K. Vegter, C. J. Lamoth, S. de Groot, D. H. E. J. Veeger, and L. H. V. van der Woude, “Inter-individual differences in the initial 80 minutes of motor learning of handrim wheelchair propulsion.” *PloS one*, vol. 9, no. 2, p. e89729, Jan. 2014.
- [2] J. Abonyi, B. Feil, S. Nemeth, and P. Arva, “Fuzzy clustering based segmentation of time-series,” *Advances in Intelligent Data Analysis V*, pp. 275–285, 2003.
- [3] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, “Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases,” *Knowledge and Information Systems*, vol. 3, no. January, pp. 263–286, 2001.
- [4] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, “A symbolic representation of time series, with implications for streaming algorithms,” *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery - DMKD '03*, p. 2, 2003.
- [5] B. Lkhagva and K. Kawagoe, “New Time Series Data Representation ESAX for Financial Applications,” in *22nd International Conference on Data Engineering Workshops (ICDEW'06)*. IEEE, Apr. 2006, pp. x115–x115.
- [6] Y. Sun, J. Li, J. Liu, B. Sun, and C. Chow, “An improvement of symbolic aggregate approximation distance measure for time series,” *Neurocomputing*, vol. 138, pp. 189–198, Aug. 2014.
- [7] N. Begum and E. Keogh, “Rare time series motif discovery from unbounded streams,” *Proceedings of the VLDB Endowment*, vol. 8, no. 2, pp. 149–160, Oct. 2014.
- [8] J. Aach and G. M. Church, “Aligning gene expression time series with time warping algorithms,” *Bioinformatics*, vol. 17, no. 6, pp. 495–508, Jun. 2001.
- [9] P. Papapetrou, V. Athitsos, M. Potamias, G. Kollios, and D. Gunopulos, “Embedding-based subsequence matching in time-series databases,” *ACM Transactions on Database Systems*, vol. 36, no. 3, pp. 1–39, Aug. 2011.
- [10] P. Esling and C. Agon, “Time-series data mining,” *ACM Computing Surveys (CSUR)*, vol. 45, no. 1, pp. 1–34, 2012.
- [11] H. Sakoe and S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, Feb. 1978.
- [12] F. Petitjean, A. Ketterlin, and P. Gançarski, “A global averaging method for dynamic time warping, with applications to clustering,” *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, Mar. 2011.
- [13] F. F. Petitjean, G. Forestier, G. I. Webb, A. E. Nicholson, Y. Chen, and E. Keogh, “Dynamic Time Warping Averaging of Time Series Allows Faster and More Accurate Classification,” in *2014 IEEE International Conference on Data Mining*. IEEE, Dec. 2014, pp. 470–479.
- [14] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh, “Querying and mining of time series data,” *Proceedings of the VLDB Endowment*, vol. 1, no. 2, pp. 1542–1552, Aug. 2008.
- [15] T. G. Dietterich, “Machine Learning for Sequential Data : A Review,” *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, pp. 16–30, Aug. 2002.