

## XIV. Analyse de complexité des données cinématiques tridimensionnelles du genou

Imen Mechmeche<sup>1,2</sup>, Amar Mitiche<sup>1</sup>, Youssef Ouakrim<sup>2,3</sup>, Jacques A. De Guise<sup>2</sup> et Neila Mezghani<sup>2,3</sup>

<sup>1</sup> Institut National de la Recherche Scientifique, INRS-EMT, Montréal, Québec, Canada,

<sup>2</sup> Laboratoire de recherche en Imagerie et Orthopédie, École de technologie supérieure, Centre de recherche du Centre hospitalier de l'Université de Montréal, Montréal, Canada,

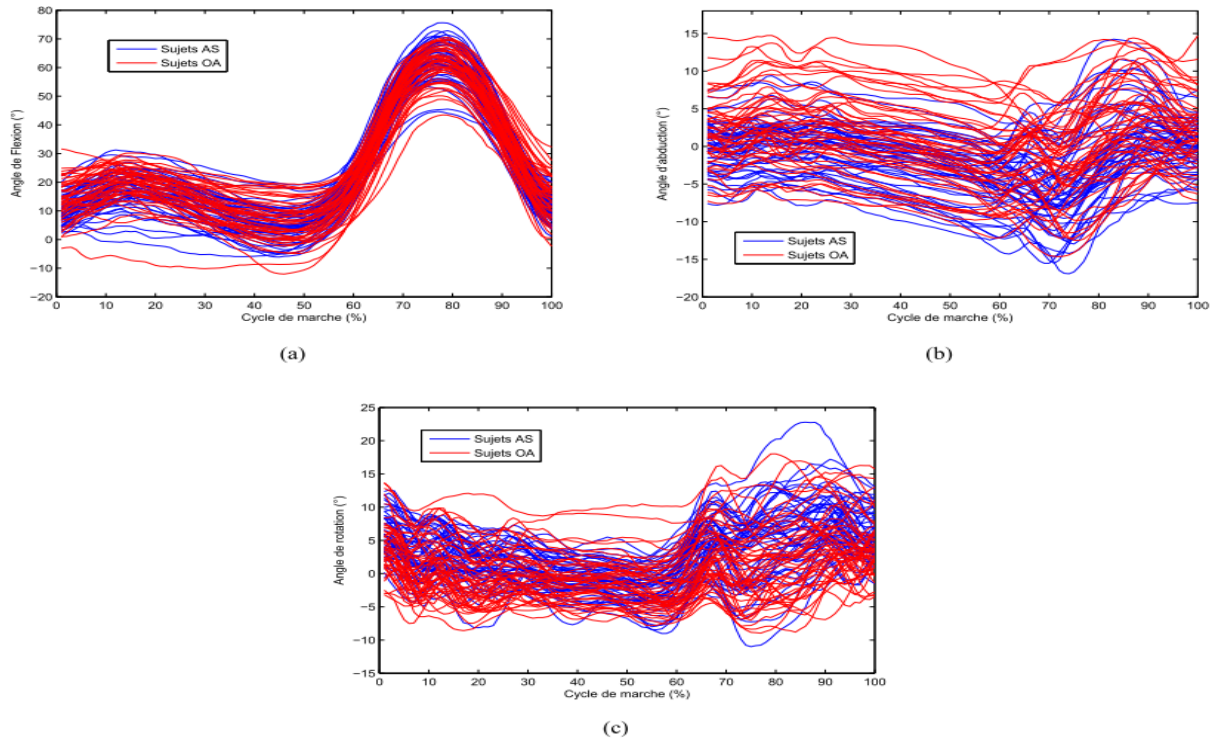
<sup>3</sup> Centre de recherche LICEF, TÉLUQ, Québec, Canada.

### Résumé

L'analyse de la marche à partir des données cinématiques du genou offre des informations importantes sur la fonction de la locomotion humaine. Ces données ont été récemment promues pour le diagnostic des pathologies de genou à travers leur classification. Le but de cette étude est d'analyser la complexité intrinsèque des données cinématiques en vue de leur classification. Nous proposons d'évaluer deux mesures : le chevauchement entre les attributs et la séparabilité des classes. Les résultats obtenus prouvent que l'ordre de complexité des données cinématiques est assez élevé comparé à celui d'autres ensembles de données utilisées dans la littérature comme données de référence.

### I. Introduction

L'analyse de la marche à partir de données cinématiques du genou (angles de flexion/extension, d'abduction/adduction et de rotation interne/externe) offre des informations sur la fonction de locomotion humaine [1] [2]. Plusieurs études ont montré la valeur ajoutée et la pertinence d'une analyse cinématique pour l'aide au diagnostic de pathologie de genou à travers leur classification [3] [4]. Cependant, une telle classification est difficile pour deux raisons. Premièrement, les données cinématiques de chaque sujet sont fournies sous la forme d'un vecteur de mesures de grande dimension rendant l'analyse vulnérable à la malédiction de la dimensionnalité [5]. Deuxièmement, ces données présentent une grande variabilité intra-classe et une petite variabilité inter-classe [6]. La figure 1 illustre les données cinématiques tridimensionnelles de deux classes: une classe de données asymptomatiques (AS) et une classe de données arthrosiques (OA). Plusieurs travaux de recherche ont appréhendés la classification de données cinématiques pour discriminer une classe de sujets sains d'une classe de sujets pathologiques [7]. Cependant aucune n'a investigué l'analyse de la complexité intrinsèque de ces données. Le but de cette étude est d'analyser la complexité des données cinématiques tridimensionnelles du genou et ce indépendamment de la méthode de classification à étudier. Il s'agit de la première étude qui aborde la problématique de la complexité de ce type de données avant d'investiguer leur classification.



**Figure 1:** Illustration des données cinématiques tridimensionnelles des deux classes: asymptotiques (AS) et arthrosiques (OA): (a) la flexion/extension, (b) l'abduction/adduction et (c) la rotation interne/externe

## 2. Mesures de complexité

La mesure de la complexité s'est basée sur l'évaluation des frontières des classes de point de vue séparabilité, chevauchement et caractéristiques géométriques. Nous avons évalué deux groupes de mesures: (1) Mesure de chevauchement entre les points d'intérêt et (2) Mesure de la séparabilité des classes [8].

### Mesure de chevauchement entre les attributs

Cette mesure évalue la capacité des attributs à séparer les instances des différentes classes [9]. Dans notre étude, ces attributs sont constitués d'un ensemble de facteurs biomécaniques mesurés sur les courbes de la cinématique. Le chevauchement a été évalué via plusieurs paramètres notamment : Le ratio maximal de Fisher (*Maximum Fisher's discriminant ratio*)  $F_1$ , le chevauchement interclasses (*Overlap of the per-class bounding boxes*)  $F_2$ , et l'efficacité maximale des points d'intérêts (*maximum individuel feature efficiency*)  $F_3$ .

Le paramètre  $F_1$  quantifie la distance entre les centres des deux classes, en se basant sur chacun des attributs individuellement comme décrit par l'équation 1.

$$F_1 = \max_i \frac{(\mu_{i1} - \mu_{i2})^2}{(\sigma_{i1}^2 + \sigma_{i2}^2)} \quad (I)$$

où  $\mu_{i1}$  et  $\sigma_{i1}$  sont la moyenne et la variance des de la  $i^{eme}$  classe par rapport au  $i^{eme}$  attribut. Une valeur élevée de ce paramètre permet de déduire que les classes sont fort probablement séparable au moins par un attribut. Cependant, la géométrie des classes intervient aussi car les classes peuvent chevaucher même si les centres sont éloignés, ainsi ce paramètre seul ne suffit pas pour trancher. Le paramètre  $F_2$  mesure le chevauchement entre les distributions des deux classes. Il est défini par le ratio de l'intervalle de chevauchement contenant des attributs appartenant aux deux classes par l'intervalle entier des attributs :

$$F_2 = \prod_{i=1}^l \frac{\min(\max(f_i, c_1), \max(f_i, c_2)) - \max(\min(f_i, c_1), \min(f_i, c_2))}{\max(\max(f_i, c_1), \max(f_i, c_2)) - \min(\min(f_i, c_1), \min(f_i, c_2))} \quad (2)$$

où  $l$  est le nombre d'attributs,  $f_i$  est le  $i^{eme}$  attribut, et  $c_1, c_2$  les deux classes. Contrairement à  $F_1$  une faible valeur de ce paramètre qui est aussi appelé volume de chevauchement (*Volume of Overlap Region*) permet de conclure que les attributs permettent de discriminer les données de chaque classe. Le 3<sup>eme</sup> paramètre  $F_3$ , calcule l'efficacité de discrimination de chaque attribut et retourne la valeur maximale. Ainsi pour chaque attribut on calcule ratio du nombre d'instances non contenu dans la région de chevauchement par le nombre total des instances. La valeur maximale correspond par la suite au paramètre  $F_3$ .

### Mesure de la séparabilité des classes

La mesure de séparabilité examine les frontières de chaque classe afin d'estimer la complexité de séparer les instances des différentes classes. Elle peut être évaluée via plusieurs paramètres, notamment : la fraction des points sur la frontière de la classe (*The fraction of points on the class boundary*)  $N1$  et le ratio de la moyenne de la distance du voisin le plus proche dans la même classe par la distance du voisin le plus proche de l'autre classe (*Ration of Average Intra/Inter Class Nearest Neighbour Distance*)  $N2$  [10]. Le calcul de  $N1$  est effectué à l'aide de l'arbre couvrant (*minimum spanning tree*) à travers toute la base de données. Ce paramètre correspond au ratio du nombre de nœuds de cet arbre qui sont connectés et appartiennent à différentes classes par le nombre des instances de la base de données. Si un nœud est connecté à plusieurs nœuds de classes différentes, il est compté juste une fois. Une valeur élevée de  $N1$  indique que la majorité des instances sont situées à la frontière de la classe ce qui rend la sa séparation plus difficile. Le paramètre  $N2$  compare l'étalement des instances de la même classe par rapport à la distance des plus proches voisins appartenant aux autres classes. Une faible valeur de  $N2$  permet de conclure que les instances de la même classe sont assez proches dans l'espace des attributs.

$$N2 = \frac{\sum_{i=0}^{N_e} \text{intraDist}(ex_i)}{\sum_{i=0}^{N_e} \text{interDist}(ex_i)} \quad (3)$$

où  $ex_i$  est une instance et  $N_e$  est le nombre d'exemples dans la base de données.

### 3. Application aux données cinématiques

Le tableau 1 résume les résultats de l'analyse de complexité. Les mesures de complexité permettent de déduire l'aspect assez complexe du problème de classification des données cinématiques du genou. Nous remarquons que le paramètre  $F1$  est assez faible ce qui reflète un chevauchement important entre les instances des classes dans l'espace des attributs. Une forte valeur de  $N2$  prouve aussi que les instances de la même classe sont assez dispersées. Les trois bases de données Iris, Abalone et Wine, nous ont servi comme références pour avoir une idée sur l'ordre de grandeur de ces paramètres. La base Abalone qui donne de très bons résultats de classification suit parfaitement le comportement que doivent avoir ces paramètres comme décrit précédemment.

**Table 1:** Mesures de complexité des données cinématiques 3D du genou ainsi que pour d'autres bases de données de référence

	$F_1$	$F_2$	$F_3$	$N1$	$N2$
Données cinématiques 3D du genou	0,38	$8,57 \cdot 10^{-14}$	0,065	0,5	0,94
Iris	16,82	0,0048	0,57	0,013	0,095
Abalone	14,03	0	0,99	$4,7 \cdot 10^{-4}$	0,025
Wine	4,29	$3,9 \cdot 10^{-5}$	0,76	0,06	0,49

#### 4. Conclusion

Dans cette étude nous avons analysé la complexité intrinsèque des données cinématiques en utilisant les mesures de chevauchement entre les attributs et la séparabilité des classes. Les résultats obtenus démontrent une grande complexité de ces données comparativement à des bases de données de références. Ces résultats nous orientent vers le besoin d'utiliser des méthodes de prétraitement des données cinématiques avant leur classification et vers l'utilisation de méthodes de classification avancées où les données ne sont pas linéairement séparables.

#### 5. References and Notes

- [1] J. De Guise, N. Mezghani, A. Fuentes, É. Szmuty, G. Grimard, P. Ranger, N. Hagemeister, R. Aissaoui, T. Cresson, J. Clément, et al., "Method and system for human joint treatment plan and personalized surgery planning using 3-d kinematics, fusion imaging and simulation," Jan. 16 2013. US Patent App. 13/743,057.
- [2] V. Medved, Measurement of Human Locomotion. CRC Press, 2000.
- [3] N. Mezghani, S. Husse, K. Boivin, K. Turcot, R. Aissaoui, N. Hagemeister, J. De Guise, et al., "Automatic classification of asymptomatic and osteoarthritis knee gait patterns using kinematic data features and the nearest neighbor classifier," Biomedical Engineering, IEEE Transactions on, vol. 55, no. 3, pp. 1230–1232, 2008.
- [4] N. Mezghani, Y. Ouakrim, A. Fuentes, N. Hagemeister, R. Aissaoui, M. Pelletier, and J. de Guise, "Knee osteoarthritis severity assessment using knee kinematic data classification," Osteoarthritis and Cartilage, vol. 20, p. S97, 2012.
- [5] R. O. Duda, P. E. Hart, et al., Pattern classification and scene analysis, vol. 3. Wiley New York, 1973.
- [6] N. Mezghani, N. Gaudreault, A. Fuentes, A. Mitiche, N. Hagemeister, R. Aissaoui, and J. de Guise, "Identification des patrons de la cinématique du genou par une analyse en composantes principales identification of knee kinematic patterns in normal gait by principal component analysis,"
- [7] N. Mezghani, K. Boivin, K. Turcot, R. Aissaoui, N. Hagemeister, and J. A. De Guise, "Hierarchical analysis and classification of asymptomatic and knee osteoarthritis gait patterns using a wavelet representation of kinetic data and the nearest neighbor classifier," Journal of Mechanics in Medicine and Biology, vol. 8, no. 01, pp. 45–54, 2008.

- [8] T. K. Ho and M. Basu, "Measuring the complexity of classification problems," in Pattern Recognition, 2000. Proceedings. 15th International Conference on, vol. 2, pp. 43–47, IEEE, 2000.
- [9] J. Sotoca, J. Sanchez, and R. Mollineda, "A review of data complexity measures and their applicability to pattern classification problems," Actas del III Taller Nacional de Minería de Datos y Aprendizaje. TAMIDA, pp. 77–83, 2005.
- [10] J.-R. Cano, "Analysis of data complexity measures for classification," Expert Systems with Applications, vol. 40, no. 12, pp. 4820–4831, 2013.