Barahona, S.; **Centella, P.**; **Gual-Arnau, X.**; **Ibáñez, M. V.**; **Simó, A.**

Supervised classification of geometrical objects by integrating currents and functional data analysis. (English) Zbl 1460.62209

Test 29, No. 3, 637-660 (2020).

Summary: This paper focuses on the application of supervised classification techniques to a set of geometrical objects (bodies) characterized by *currents*, in particular, discriminant analysis and some nonparametric methods. A *current* is a relevant mathematical object to model geometrical data, like hypersurfaces, through integration of vector fields over them. As a consequence of the choice of a vector-valued reproducing kernel Hilbert space (RKHS) as a test space to integrate over hypersurfaces, it is possible to consider that hypersurfaces are embedded in this Hilbert space. This embedding enables us to consider classification algorithms of geometrical objects. We present a method to apply supervised classification techniques in the obtained vector-valued RKHS. This method is based on the eigenfunction decomposition of the kernel. The novelty of this paper is therefore the reformulation of a size and shape supervised classification problem in functional data analysis terms using the theory of currents and vector-valued RKHSs. This approach is applied to a 3D database obtained from an anthropometric survey of the Spanish child population with a potential application to online sales of children's wear.

MSC:

62R10 Functional data analysis
62H30 Classification and discrimination; cluster analysis (statistical aspects)
46E22 Hilbert spaces with reproducing kernels (= (proper) functional Hilbert spaces, including de Branges-Rovnyak and other structured spaces)
47N30 Applications of operator theory in probability theory and statistics
62P10 Applications of statistics to biology and medical sciences; meta analysis
62P20 Applications of statistics to economics

Keywords:

currents; statistical shape and size analysis; reproducing kernel Hilbert space; functional data analysis; supervised classification methods; discriminant analysis

Software:

fda (R); RKHS; R; Matlab

Full Text: DOI arXiv

References:

[1] Alonso, AM; Casado, D.; Romo, J., Supervised classification for functional data: a weighted distance approach, Comput Stat Data Anal, 56, 7, 2334-2346 (2012) · Zbl 1252.62061

[2] Aneiros, G.; Bongiorno, EG; Cao, R.; Vieu, P., Functional statistics and related fields (2017), Berlin: Springer, Berlin

[3] Araki, Y.; Konishi, S.; Kawano, S.; Matsui, H., Functional logistic discrimination via regularized basis expansions, Commun Stat Theory Methods, 38, 16-17, 2944-2957 (2009) · Zbl 1175.62061

[4] Aronszajn, N., Theory of reproducing kernels, Trans Am Math Soc, 68, 3, 337-404 (1950) · Zbl 0037.20701

[5] Ballester A, Valero M, Nácher B, Piérola A, Piqueras P, Sancho M, Gargallo G, González J, Alemany S (2015) 3D body databases of the spanish population and its application to the apparel industry. In: Proceedings of the 6th international conference on 3D body scanning technologies, Lugano, Switzerland

[6] Ballester, A.; Parrilla, E.; Piérola, A.; Uriel, J.; Pérez, C.; Piqueras, P.; Vivas, J.; Alemany, S., Data-driven three-dimensional reconstruction of human bodies using a mobile phone app, Int J Digit Hum, 1, 4, 361-388 (2016)

[7] Barahona, S.; Gual-Arnau, X.; Ibáñez, M.; Simó, A., Unsupervised classification of children's bodies using currents, Adv Data Anal Classif, 12, 2, 365-397 (2018) · Zbl 1414.62431

[8] Berlinet, A.; Thomas-Agnan, C., Reproducing kernel Hilbert spaces in probability and statistics (2011), Berlin: Springer, Berlin

[9] Berrendero, JR; Cuevas, A.; Torrecilla, JL, On the use of reproducing kernel hilbert spaces in functional classification, J Am

Stat Assoc, 113, 523, 1210-1218 (2018) · Zbl 1402.68152

[10] Bickel, PJ; Li, B.; Tsybakov, AB; van de Geer, SA; Yu, B.; Valdés, T.; Rivero, C.; Fan, J.; van der Vaart, A., Regularization in statistics, Test, 15, 2, 271-344 (2006)

[11] Boj, E.; Caballé, A.; Delicado, P.; Esteve, A.; Fortiana, J., Global and local distance-based generalized linear models, Test, 25, 1, 170-195 (2016) · Zbl 1338.62111

[12] Bouveyron, C.; Brunet-Saumard, C., Model-based clustering of high-dimensional data: a review, Comput Stat Data Anal, 71, 52-78 (2014) · Zbl 1471.62032

[13] Chiou, JM; Li, PL, Functional clustering and identifying substructures of longitudinal data, J R Stat Soc Ser B (Stat Methodol), 69, 4, 679-699 (2007)

[14] Cuadras, C., Distance analysis in discrimination and classification using both continuous and categorical variables, Statistical data analysis and inference, 459-473 (1989), Amsterdam: Elsevier, Amsterdam · Zbl 0735.62060

[15] Cucker, F.; Smale, S., On the mathematical foundations of learning, Am Math Soc, 39, 1, 1-49 (2001) · Zbl 0983.68162

[16] Cuesta-Albertos, JA; Fraiman, R., Impartial trimmed k-means for functional data, Comput Stat Data Anal, 51, 10, 4864-4877 (2007) · Zbl 1162.62377

[17] Cuevas, A., A partial overview of the theory of statistics with functional data, J Stat Plan Inference, 147, 1-23 (2014) · Zbl 1278.62012

[18] Cuevas, A.; Febrero, M.; Fraiman, R., Robust estimation and classification for functional data via projection-based depth notions, Comput Stat, 22, 3, 481-496 (2007) · Zbl 1195.62032

[19] Dai, X.; Müller, HG, Principal component analysis for functional data on riemannian manifolds and spheres, Ann Stat, 46, 6, 3334-3361 (2018) · Zbl 1454.62553

[20] Delicado, P., Functional k-sample problem when data are density functions, Comput Stat, 22, 3, 391-410 (2007) · Zbl 1197.62041

[21] Delicado, P., Dimensionality reduction when data are density functions, Comput Stat Data Anal, 55, 1, 401-420 (2011) · Zbl 1247.62148

[22] Delicado, P.; Vieu, P., Choosing the most relevant level sets for depicting a sample of densities, Comput Stat, 32, 3, 1083-1113 (2017) · Zbl 1417.62062

[23] Devarajan, P.; Istook, CL, Validation of female figure identification technique (FFIT) for apparel software, J Text Appar Technol Manag, 4, 1, 1-23 (2004)

[24] Di Marzio, M.; Fensore, S.; Panzera, A.; Taylor, CC, Kernel density classification for spherical data, Stat Probab Lett, 144, 23-29 (2019) · Zbl 1407.62184

[25] Dryden, IL; Mardia, KV, Statistical shape analysis: with applications (2016), Hoboken: Wiley, Hoboken

[26] Durrleman S (2010) Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Ph.D. thesis, Université Nice Sophia Antipolis

[27] Durrleman, S.; Pennec, X.; Trouvé, A.; Ayache, N., Statistical models of sets of curves and surfaces based on currents, Med Image Anal, 13, 5, 793-808 (2009)

[28] Eubank, R.; Hsing, T., Canonical correlation for stochastic processes, Stoch Process Their Appl, 118, 9, 1634-1661 (2008) · Zbl 1145.62048

[29] Federer, H.; Fleming, W., Normal and integral currents, Ann Math, 72, 458-520 (1960) · Zbl 0187.31301

[30] Ferraty, F.; Vieu, P., Nonparametric functional data analysis: theory and practice (2006), Berlin: Springer, Berlin · Zbl 1119.62046

[31] Fisher, R., The use of multiple measurements in taxonomic problems, Ann Eugen, 7, 2, 179-188 (1936)

[32] Flores, M.; Gual-Arnau, X.; Ibáñez, M.; Simó, A., Intrinsic sample mean in the space of planar shapes, Pattern Recognit, 60, 164-176 (2016) · Zbl 1414.68093

[33] Fraiman, R.; Gamboa, F.; Moreno, L., Connecting pairwise geodesic spheres by depth: DCOPS, J Multivar Anal, 169, 81-94 (2019) · Zbl 1404.60070

[34] Garcia-Escudero, LA; Gordaliza, A., A proposal for robust curve clustering, J Classif, 22, 2, 185-201 (2005) · Zbl 1336.62179

[35] Gerber S, Tasdizen T, Whitaker R (2009) Dimensionality reduction and principal surfaces via kernel map manifolds. In: IEEE 12th international conference on computer vision, IEEE, pp 529-536

[36] Glaunès J, Joshi S (2006) Template estimation form unlabeled point set data and surfaces for computational anatomy. In: 1st MICCAI Workshop on mathematical foundations of computational anatomy: geometrical, statistical and registration methods for modeling biological shape variability

[37] Goia, A.; Vieu, P., An introduction to recent advances in high/infinite dimensional statistics, J Multivar Anal, 146, 1-6 (2016) · Zbl 1384.00073

[38] González J, Muñoz A (2010) Representing functional data in reproducing kernel hilbert spaces with applications to clustering and classification. Tech. rep., Universidad Carlos III de Madrid. Departamento de Estadística

[39] Hall, P.; Poskitt, D.; Presnell, B., A functional data-analytic approach to signal discrimination, Technometrics, 43, 1, 1-9 (2001) · Zbl 1072.62686

[40] Horváth, L.; Kokoszka, P., Inference for functional data with applications (2012), Berlin: Springer, Berlin · Zbl 1279.62017

[41] Hsing, T.; Eubank, R., Theoretical foundations of functional data analysis, with an introduction to linear operators (2015),

Hoboken: Wiley, Hoboken · Zbl 1338.62009

[42] Jacques, J.; Preda, C., Functional data clustering: a survey, Adv Data Anal Classif, 8, 3, 231-255 (2014) · Zbl 1414.62018

[43] James, G.; Hastie, T., Functional linear discriminant analysis for irregularly sampled curves, J R Stat Soc Ser B (Stat Methodol), 63, 3, 533-550 (2001) · Zbl 0989.62036

[44] Kadri, H.; Duflos, E.; Preux, P.; Canu, S.; Rakotomamonjy, A.; Audiffren, J., Operator-valued kernels for learning from functional response data, J Mach Learn Res, 17, 1, 613-666 (2016)

[45] Kendall, D.; Barden, D.; Carne, T.; Le, H., Shape and shape theory (2009), Hoboken: Wiley, Hoboken · Zbl 0940.60006

[46] Kent JT, Mardia KV, Morris RJ, Aykroyd RG (2001) Functional models of growth for landmark data. In: Proceedings in functional and spatial data analysis university of leeds, pp 109-115

[47] Kneip, A.; Utikal, KJ, Inference for density families using functional principal component analysis, J Am Stat Assoc, 96, 454, 519-542 (2001) · Zbl 1019.62060

[48] Kupresanin, A.; Shin, H.; King, D.; Eubank, R., An rkhs framework for functional data analysis, J Stat Plan Inference, 140, 12, 3627-3637 (2010) · Zbl 1327.62379

[49] Lian, H., Nonlinear functional models for functional responses in reproducing kernel hilbert spaces, Can J Stat, 35, 4, 597-606 (2007) · Zbl 1142.62020

[50] Lin, L.; St Thomas, B.; Zhu, H.; Dunson, DB, Extrinsic local regression on manifold-valued data, J Am Stat Assoc, 112, 519, 1261-1273 (2017)

[51] López-Pintado, S.; Romo, J., On the concept of depth for functional data, J Am Stat Assoc, 104, 486, 718-734 (2009) · Zbl 1388.62139

[52] Loubes, JM; Pelletier, B., A kernel-based classifier on a riemannian manifold, Stat Decis Int Math J Stoch Methods Models, 26, 1, 35-51 (2008) · Zbl 1418.62161

[53] Lukić, M.; Beder, J., Stochastic processes with sample paths in reproducing kernel hilbert spaces, Trans Am Math Soc, 353, 10, 3945-3969 (2001) · Zbl 0973.60036

[54] Marron, JS; Alonso, AM, Overview of object oriented data analysis, Biom J, 56, 5, 732-753 (2014) · Zbl 1309.62008

[55] MATLAB (2015) version 8.6.0 (R2015b). The MathWorks Inc., Natick, MA

[56] Meunier P (2000) Use of body shape information in clothing size selection. In: Proceedings of the human factors and ergonomics society annual meeting. SAGE Publications Sage CA: Los Angeles, vol 44, pp 715-718

[57] Müller, HG, Functional modelling and classification of longitudinal data, Scand J Stat, 32, 2, 223-240 (2005) · Zbl 1089.62072

[58] Pelletier, B., Non-parametric regression estimation on closed Riemannian manifolds, J Nonparametric Stat, 18, 1, 57-67 (2006) · Zbl 1088.62053

[59] Peng, J.; Müller, HG, Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions, Ann Appl Stat, 2, 3, 1056-1077 (2008) · Zbl 1149.62053

[60] Pennec, X., Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements, J Math Imaging Vis, 25, 1, 127-154 (2006)

[61] Preda, C., Regression models for functional data by reproducing kernel hilbert spaces methods, J Stat Plan Inference, 137, 3, 829-840 (2007) · Zbl 1104.62043

[62] Preda, C.; Saporta, G.; Lévéder, C., PLS classification of functional data, Comput Stat, 22, 2, 223-235 (2007) · Zbl 1196.62086

[63] Quang, M.; Kang, S.; Le, T., Image and video colorization using vector-valued reproducing kernel Hilbert spaces, J Math Imaging Vis, 37, 1, 49-65 (2010) · Zbl 1392.94042

[64] R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[65] Rham Gd (1960) Variétés différentiables. Formes, courants, formes harmoniques. Hermann et Cie, Butterworths Scientific Publications, Hermann

[66] Ripley, B., Pattern recognition and neural networks (2007), Cambridge: Cambridge University Press, Cambridge · Zbl 0853.62046

[67] Rossi, F.; Villa, N., Recent advances in the use of svm for functional data classification, Functional and operatorial statistics, 273-280 (2008), Berlin: Springer, Berlin

[68] Saitoh, S.; Sawano, Y., Theory of reproducing kernels and applications (2016), Berlin: Springer, Berlin · Zbl 1358.46004

[69] Schölkopf, B.; Smola, AJ; Bach, F., Learning with kernels: support vector machines, regularization, optimization, and beyond (2002), Cambridge: MIT Press, Cambridge

[70] Senkene, E.; Tempel'man, A., Hilbert spaces of operator-valued functions, Math Trans Acad Sci Lith SSR, 13, 4, 665-670 (1973)

[71] Serra, J., Image analysis and mathematical morphology (1982), Cambridge: Academic Press, Cambridge · Zbl 0565.92001

[72] Shin, H., An extension of Fisher's discriminant analysis for stochastic processes, J Multivar Anal, 99, 6, 1191-1216 (2008) · Zbl 1141.62053

[73] Silverman, B.; Ramsay, J., Functional data analysis (2005), Berlin: Springer, Berlin · Zbl 1079.62006

[74] Smale, S.; Zhou, DX, Geometry on probability spaces, Constr Approx, 30, 3, 311-323 (2009) · Zbl 1187.68270

[75] Stoyan, D.; Stoyan, H., Fractals, random shapes and point fields. Methods of geometrical statistics (1994), Hoboken: Wiley,

Hoboken · Zbl 0828.62085

[76] Tenenbaum, JB; De Silva, V.; Langford, JC, A global geometric framework for nonlinear dimensionality reduction, Science, 290, 5500, 2319-2323 (2000)

[77] Turaga, PK; Srivastava, A., Riemannian computing in computer vision (2016), Berlin: Springer, Berlin · Zbl 1335.65003

[78] Vaillant, M.; Glaunès, J., Surface matching via currents, Information processing in medical imaging, 381-392 (2005), Berlin: Springer, Berlin

[79] Viktor HL, Paquet E, Guo H (2006) Measuring to fit: virtual tailoring through cluster analysis and classification. In: European conference on principles of data mining and knowledge discovery. Springer, Berlin, pp 395-406

[80] Vinué, G.; Simó, A.; Alemany, S., The k-means algorithm for 3D shapes with an application to apparel design, Adv Data Anal Classif, 10, 1, 103-132 (2016) · Zbl 1414.62295

[81] Wang, H.; Marron, J., Object oriented data analysis: sets of trees, Ann Stat, 35, 5, 1849-1873 (2007) · Zbl 1126.62002

[82] Wang, JL; Chiou, JM; Müller, HG, Functional data analysis, Ann Rev Stat Its Appl, 3, 257-295 (2016)

[83] Younes, L., Computable elastic distance between shapes, SIAM J Appl Math, 58, 2, 565-586 (1998) · Zbl 0907.68158