Statistics and Probability Letters 179 (2021) 109227

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Ordinal classification of 3D brain structures by functional data analysis[☆]

L. Ferrando[a], I. Epifanio[b,*], N. Ventura-Campos[a,c]

[a] Grup Neuropsicologia i Neuroimatge Funcional, Universitat Jaume I, Spain
[b] Dept. Matemàtiques-IF. Universitat Jaume I, Spain
[c] Dept. Educació i Didàctiques Específiques. Universitat Jaume I, Spain

ARTICLE INFO

Article history:
Received 8 November 2020
Received in revised form 28 July 2021
Accepted 14 August 2021
Available online 24 August 2021

Keywords:
Ordinal data
Supervised statistical learning
Functional data
Neuroeducation
Alzheimer's disease

ABSTRACT

We introduce several ordinal classification methods for functional data, specifically multiargument and multivariate functional data. Their performance is analyzed in four real data sets that belong to a neuroeducational problem and a neuropathological problem.

1. Introduction

Supervised learning is one of the most common problems in statistics. Classification is a problem where the objective is to predict a class in a set, which is the so-called response, from a set of inputs. Classes are usually considered unordered, i.e. as levels of a nominal variable, and the majority of classification algorithms are designed for this kind of problem. However, classes are ordered, i.e. labels are levels of an ordinal variable, in many real life problems, such as collaborative filtering, econometric modeling, medicine, psychology, social sciences, and more (Gutiérrez et al., 2016). An example of an ordered variable would be patient condition (good, fair, serious, critical) or the rating of satisfaction (low, indifferent, high). Nevertheless, the literature about ordered classification methods is not very extensive for multivariate data or high-dimensional data (Hornung, 2020; Simó et al., 2020) and even less so for functional data (Wang and Shi, 2014).

In the multivariate case, Gutiérrez et al. (2016) established a taxonomy according to how the order is taken into account in the classification procedure. They proposed three main approaches. The first is the naïve approach, which is the simplest and is very common, not only in the multivariate context but also in the functional context. It consists of using standard classification algorithms as if classes were unordered, i.e. nominal classification. In this case, the ordering information is not taken into account, and that information is lost. The second approach consists of decomposing the ordinal problem into several binary ones. Then, each of them can be solved by standard classification methods and the results are combined to return a label, as described by Frank and Hall (2001). The third approach assumes that an unobserved continuous variable underlies the ordinal response. Some of the methods that belong to this approach are: ordinal logistic regression models, such as the cumulative link models used by Pierola et al. (2016), augmented binary classification problems, such

☆ The data sets and code for reproducing the results are available at http://www3.uji.es/~epifanio/RESEARCH/orderfda.zip.

* Corresponding author.
E-mail address: epifanio@uji.es (I. Epifanio).

https://doi.org/10.1016/j.spl.2021.109227
0167-7152/

as the data replication method by Cardoso and Costa (2007), and ensemble methods, such as the ordinal random forests proposed by Hornung (2020).

According to Gutiérrez et al. (2016), although results from the naïve approach can be very competitive in the multivariate case, taking into account the order improves the performance. The few papers that deal with ordinal classification when inputs are functional data are based mainly on the use of the functional generalized linear model (Aguilera and Escabias, 2008; Barahona et al., 2020) and, in many cases, the order is ignored (Epifanio and Ventura, 2014). To the best of our knowledge, ordinal methods for functional data with the ordinal binary decomposition approach, the second approach, have not been considered until now, and the third approach has not been fully exploited either. On the other hand, classification problems in neuroscience do not usually exploit the ordering information of the different classes in the multivariate case.

The objective of this work is to introduce more ordinal methods for functional data, use them for the first time in two neuroscience problems and compare these methods. As a novelty, we present ordinal methods for the second and third approaches in the functional context. Note also that the functional data in both neuroscience problems are multivariate functional data with multiple arguments, which are not the classical univariate functional data. Furthermore, functional data analysis (FDA) is not usually applied to the analysis of brain structures, with some exceptions (Epifanio and Ventura, 2014; Lila et al., 2016), although it is a natural way to do it. In fact, in the neuroscience literature, many brain structures are represented with a functional basis, such as spherical harmonic (SPHARM) representation, but they do not go any further and functional data procedures are not exploited in this field. To the best of our knowledge, no previous work has considered ordinal classification in neuroscience with functional data as inputs.

The outline of the paper is as follows: Section 2 reviews ordinal classification methods for the multivariate and the functional case. Section 3 contains the proposed functional ordinal classification methods. The results are discussed in Section 4. Section 5 contains the conclusions. The appendix introduces the motivating problems and describes the data in detail.

2. Established ordinal classification methodologies

2.1. Multivariate ordinal classification methodologies

Let $\mathbf{X}$ be an $N \times K$ matrix with K inputs in N instances and $\mathbf{y}$ the output vector, which is an ordered factor with Q levels, the ordered classes $C_1, \ldots, C_Q$.

The Frank and Hall (FH) method: Frank and Hall (2001) proposed to decompose the ordinal classification problem into the following binary ones: they discriminated between $C_1, \ldots, C_q$ and $C_{q+1}, \ldots, C_Q$, for $q = 1, \ldots, Q-1$. Each of these binary classification problems yields an estimate p_q of $P(y \in C_{q+1}, \ldots, C_Q|\mathbf{x})$, for a new instance with input $\mathbf{x}$. Then, the predicted probability values of each of the Q classes for the corresponding output y are estimated by: $P(y = C_1|\mathbf{x}) = 1-p_1$; $P(y = C_q|\mathbf{x}) = p_{q-1} - p_q$, $q = 2, \ldots, Q-1$; $P(y = C_Q|\mathbf{x}) = p_{Q-1}$, where $p_q = P(y > C_q \mid \mathbf{x})$ for $q = 1, \ldots, Q-1$. It was proposed for the multivariate case. Note that this method is applicable as long as the binary classifier produces class probability estimates.

Ordered logistic regression: The cumulative link model is explained in detail by Agresti (2002, Ch. 7). The model is $\text{logit}(P(y \le q|\mathbf{x})) = \zeta_q - \eta$, where the logit link function is the inverse of the standard logistic cumulative distribution function, i.e. $\text{logit}(p) = \log(p/(1-p))$, ζ_q parameters provide each cumulative logit, and η is the linear predictor $\beta_1 x_1 + \cdots + \beta_K x_K$. We refer to this method as POLR.

The data replication method: Cardoso and Costa (2007) proposed the data replication method for the multivariate case, where the ordinal classification problem is reduced to binary classification problems by augmenting the features, through their replication, i.e. the original vector is concatenated with each of the extension features. A final classification rule is built based on the results of the binary problems to obtain the prediction of a new instance. We refer to this method as oSVM (ordinal Support Vector Machine), since SVMs are used in the implementation.

Ordinal forest (OF): Hornung (2020) proposed a random forest-based prediction method for ordinal outputs in the multivariate case. The idea of OF is to use optimized score values in place of the category values of the ordinal output and treat the result as a metric output.

2.2. Functional nominal and ordinal classification methodologies

Kernel-Induced Random Forests (KIRF): Fan et al. (2010) proposed KIRF for functional data. Kernel-induced classification trees are built using kernel functions of each two different cases of the training set as candidate splitting rules. These trees are used in KIRF. Fan et al. (2010) proposed KIRF for functional data classification by defining some kernels for functional data. Note that only nominal functional classification was considered with univariate and unidimensional functional data. They considered the kernel function of two curves as the squared Euclidean norm of the principal component (PC) scores of the PC expansion with K components.

FPCA-SVS-LDA: This method is also a nominal classification procedure, like KIRF. It was proposed by Ferrando et al. (2020), and it consists of using functional principal component analysis (FPCA) followed by step-wise variable selection (SVS) for linear discriminant (LDA) classification. It belongs to the first approach as described in Section 1, where the order is not taken into account.

FPCA-POLR: Aguilera and Escabias (2008) used FPCA to solve multicollinearity in functional multinomial logit models for ordinal responses. We use FPCA followed by POLR. It belongs to the third approach.

Table 1
Performance measures for several classifiers for the left putamens data set: accuracy, sensitivity, specificity, precision, and negative predictive value (NPV) for each class and RPS.

Method	Accuracy	Sensitivity			Specificity			Precision			NPV			RPS
		E	ME	NE	E	ME	NE	E	ME	NE	E	ME	NE	
FPCA-SVS-LDA	0.313	**0.267**	0.353	0.278	0.712	0.576	0.653	0.211	0.462	0.227	0.771	0.463	0.711	0.392
FPCA-POLR	0.418	**0.267**	0.529	0.333	0.731	**0.667**	0.714	0.222	**0.621**	0.300	0.776	**0.579**	0.745	0.370
FH-FPCA-SVS-LDA	0.418	**0.267**	0.500	**0.389**	0.731	0.636	0.735	0.222	0.586	0.350	0.776	0.553	0.767	0.356
FPCA-oSVM	0.448	0.200	0.706	0.167	0.923	0.273	0.816	0.429	0.500	0.250	0.800	0.474	0.727	–
FPCA-OF	**0.501**	0.200	**0.912**	0	**1**	0.091	**0.939**	**1**	0.508	0	**0.813**	0.500	0.719	**0.190**
FPCA-KIOF	0.478	0.200	0.647	**0.389**	0.885	0.424	0.796	0.333	0.537	**0.412**	0.793	0.539	**0.780**	0.193

Table 2
Performance measures for several classifiers for the right putamens data set.

Method	Accuracy	Sensitivity			Specificity			Precision			NPV			RPS
		E	ME	NE	E	ME	NE	E	ME	NE	E	ME	NE	
FPCA-SVS-LDA	0.358	0.400	0.412	0.222	0.865	**0.485**	0.612	0.462	0.452	0.174	0.833	0.444	0.682	0.328
FPCA-POLR	0.433	0.333	0.559	0.278	0.827	**0.485**	0.755	0.357	0.528	0.294	0.811	0.516	0.740	0.315
FH-FPCA-SVS-LDA	0.418	**0.533**	0.529	0.111	0.846	**0.485**	0.714	0.500	0.514	0.125	**0.863**	0.500	0.686	0.323
FPCA-oSVM	**0.537**	0.200	0.853	0.222	0.981	0.242	0.898	**0.750**	0.537	**0.444**	0.810	**0.615**	0.759	–
FPCA-OF	0.493	0	**0.971**	0	**1**	0	**0.980**	–	0.500	0	0.776	0	0.727	**0.192**
FPCA-KIOF	0.433	0	0.677	**0.333**	0.846	0.424	0.776	0	**0.548**	0.353	0.746	0.560	**0.760**	0.214

3. Proposed ordinal functional classification methodologies

The proposed methods generalize the previous methods to functional data by exploiting an FPCA decomposition.

FH-FPCA-SVS-LDA: We propose to use the FH method considering FPCA-SVS-LDA as the binary classifier since it returns class probability estimates. In other words, we apply the FH method to the matrices of FPC scores. To the best of our knowledge, this is the first time FH has been used with functional data. This method belongs to the second approach.

FPCA-oSVM: We consider the same idea used by Hall et al. (2001), Epifanio (2008), and Epifanio and Ventura (2011): a feature extraction stage whose resulting features are carried forward to a (nominal) classification stage. However, in our case, the nominal classification stage is changed for an ordinal classification stage. Feature extraction is a powerful preprocessing method for improving the performance of a learning algorithm (see the section on Feature Extraction in Hastie et al. (2009, pp. 126–127)). In this case, FPCA is used for the feature extraction stage. As the FPC scores are multivariate, oSVM is used for the ordinal classification stage. This method belongs to the third approach.

FPCA-OF: We again consider the previous idea. FPCA is also used for the feature extraction stage, but OF is used for the ordinal classification stage. This method belongs to the third approach.

FPCA-KIOF: We consider KIRF for functional data, but in this case, OF is used for ordinal classification with multivariate and multiargument functional data. This method belongs to the third approach.

4. Results and discussion

We apply the methods presented in Sections 2.2 and 3 to the data described in Appendix. In particular, our data are the FPC score matrices of the left and right putamens and the left and right hippocampi, which are described by multivariate and multiargument functional data. The implementation of the methods is explained in Appendix, including the choice of the number of PCs. The final class assignation is implemented by choosing the class with the highest probability. The performance of the methods is assessed by leave-one-out (LOO) cross-validation. In each trial, we leave one subject out and FPCA is applied to the remaining subjects, which form the training set of that trial. The FPC scores of the training set are used with the classification methods to fit each model. We compute the FPC scores of the left-out subject, which is the test set of that trial, and we predict its class and/or save the class probability estimates. We repeat this process for each subject of the data set. In this way, we obtain LOO performance estimates. The performance is evaluated with different measures: the commonly used measures for nominal classification assessment, such as accuracy, but we also use the ranked probability score (RPS), which is specific for ordered classification (see Appendix for details). Tables 1 and 2 show the results for the left and right putamens, respectively, while Tables 3 and 4 show the results for the left and right hippocampi, respectively. The best value in each column appears in bold. RPS can be computed for all the methods, except FPCA-oSVM, since oSVM returns only the predicted class. We can see the performance in different kinds of real problems. On the one hand, the neuroeducation problem is a very difficult classification problem, as it is not easy to distinguish between classes with the putamen shape. On the other hand, the neuropathological problem is an easier classification problem. The classes can be distinguished better, especially for the left hippocampus.

According to the results in Table 1, it is clear that taking into account the order improves the performance. FPCA-SVS-LDA is the method with the lowest accuracy and the highest RPS. The methods with the highest accuracy are FPCA-OF,

Table 3
Performance measures for several classifiers for the left hippocampi data set.

Method	Accuracy	Sensitivity			Specificity			Precision			NPV			RPS
		CN	MCI	AD	CN	MCI	AD	CN	MCI	AD	CN	MCI	AD	
FPCA-SVS-LDA	0.750	0.833	0.333	0.900	0.813	0.864	0.944	0.769	0.400	**0.900**	0.867	0.826	0.944	0.088
FPCA-POLR	0.821	**1**	0.333	0.900	0.875	0.955	**0.889**	0.857	0.667	0.818	**1**	0.840	0.941	0.089
FH-FPCA-SVS-LDA	**0.893**	**1**	**0.500**	**1**	**0.938**	**1**	**0.889**	**0.923**	**1**	0.833	**1**	**0.880**	**1**	**0.059**
FPCA-oSVM	0.607	0.833	0.167	0.600	0.750	0.773	**0.889**	0.714	0.167	0.750	0.857	0.773	0.800	–
FPCA-OF	0.750	**1**	0	0.900	0.686	**1**	**0.889**	0.706	–	0.818	**1**	0.786	0.941	0.149
FPCA-KIOF	0.750	**1**	0	0.900	0.813	0.955	0.833	0.800	0	0.750	**1**	0.778	0.938	0.095

Table 4
Performance measures for several classifiers for the right hippocampi data set.

Method	Accuracy	Sensitivity			Specificity			Precision			NPV			RPS
		CN	MCI	AD	CN	MCI	AD	CN	MCI	AD	CN	MCI	AD	
FPCA-SVS-LDA	0.571	0.917	0	0.500	**0.875**	0.818	0.667	0.846	0	0.455	0.933	0.750	0.706	0.204
FPCA-POLR	0.679	0.917	0.167	0.700	**0.875**	0.864	0.778	0.846	0.250	0.636	0.933	0.792	0.824	0.179
FH-FPCA-SVS-LDA	0.679	**1**	**0.333**	0.500	**0.875**	0.818	0.833	**0.857**	**0.333**	0.625	**1**	**0.818**	0.750	**0.137**
FPCA-oSVM	0.607	**1**	**0.333**	0.300	0.625	0.818	**0.944**	0.667	**0.333**	**0.750**	**1**	**0.818**	0.708	–
FPCA-OF	**0.714**	**1**	0	**0.800**	0.813	**1**	0.722	0.800	–	0.615	**1**	0.786	**0.867**	0.160
FPCA-KIOF	0.607	0.917	0	0.600	0.688	0.955	0.722	0.688	0	0.546	0.917	0.778	0.765	0.162

FPCA-KIOF, and FPCA-oSVM, which are also the methods with the lowest RPS: FPCA-OF and FPCA-KIOF. However, many of the subjects are classified in the ME class for FPCA-KIOF, FPCA-OF, and FPCA-oSVM. For example, 61 subjects are assigned to the ME group by FPCA-OF. ME is the most numerous group, but only 34 individuals belong to that intermediate class. There is a high percentage in sensitivity, but a low percentage in specificity for FPCA-OF in the ME class.

The same performance pattern is observed in the results in Table 2. Again, it is beneficial to consider the order because the method of the first (naïve) approach returns the worst performance in terms of accuracy and RPS. In this case, the best accuracy is clearly attained by FPCA-oSVM, i.e. the difference in accuracy compared to the other methods is wider.

As regards the results in Table 3, FH-FPCA-SVS-LDA is the best method in terms of accuracy and RPS. The second best in terms of accuracy is FPCA-POLR. In this case, FPCA-SVS-LDA, which does not take order into account, provides results that are similar to or better than other methods that consider the ordering information. This shows that naïve methods can also be very competitive. The accuracy of FPCA-SVS-LDA is equal to that attained by FPCA-OF and FPCA-KIOF and better than that of FPCA-oSVM. The RPS value for FPCA-SVS-LDA is the second best. Note the difference between the results for putamens and hippocampi. For right putamens the best method was FPCA-oSVM, which is the worst for left hippocampi in terms of accuracy. Therefore, there is no single method that performs best in all possible datasets, as is the case in nominal multivariate classification.

For right hippocampi, the results in Table 4 show that FPCA-OF is the best in terms of accuracy, but FH-FPCA-SVS-LDA is the best in terms of RPS. FH-FPCA-SVS-LDA returns the second-best in accuracy, together with FPCA-POLR. However, the RPS value of FPCA-POLR is the second-worst. The naïve method, FPCA-SVS-LDA, is again the worst in terms of both accuracy and RPS.

5. Conclusion

We have introduced several methodologies for ordinal classification of functional data. This problem has hardly been studied. We have considered methods that have not been considered until now: FH-FPCA-SVS-LDA, a method with the ordinal binary decomposition approach; feature extraction plus augmented binary classification (FPCA-oSVM), feature extraction plus ensembles (FPCA-OF) and kernel-induced ordinal random forests (FPCA-KIOF). They have been analyzed in four real neuroscience data sets, and the results confirm that taking into account ordering information improves performance. We have seen that there is no 'number one' method, but a method can perform better in some data sets and worse in other data sets. This is why having different alternative methodologies to address an ordinal classification problem is a good option.

Although the proposed methodologies have been presented for multiargument and multivariate functional data, they can be used for classical univariate functional data. In fact, as future work, ordinal classification for univariate functional data can be studied for both dense and sparse functional data. Note that FPCA can be performed for sparse functional data (Yao et al., 2005). We can extend more ordinal classification methodologies from the multivariate case (see Gutiérrez et al. (2016) for a survey) to the functional case. Another direction of future work would be to consider more applications, not only in the neuroscience field. Many real-world applications include ordinal classification and ordinal information should not be ignored.

Funding

This work has been supported by the following grants: DPI2017-87333-R and PID2020-118763GA-I00 from the Spanish Ministry of Science and Innovation and UJI-A2017-8 and UJI-B2020-22 from Universitat Jaume I, Spain.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2021.109227.

References

Agresti, A., 2002. Categorical Data Analysis. Wiley.
Aguilera, A., Escabias, M., 2008. Solving multicollinearity in functional multinomial logit models for nominal and ordinal responses. In: Functional and Operatorial Statistics. Springer, pp. 7–13.
Barahona, S., Centella, P., et al., 2020. Generalized linear models for geometrical current predictors, an application to predict garment fit. Stat. Model 20 (6), 562–591.
Cardoso, J.S., Costa, J.F., 2007. Learning to classify ordinal data: The data replication method. J. Mach. Learn. Res. 8 (Jul), 1393–1429.
Epifanio, I., 2008. Shape descriptors for classification of functional data. Technometrics 50 (3), 284–294.
Epifanio, I., Ventura, N., 2011. Functional data analysis in shape analysis. Comput. Statist. Data Anal. 55 (9), 2758–2773.
Epifanio, I., Ventura, N., 2014. Hippocampal shape analysis in Alzheimer's disease using functional data analysis. Stat. Med. 33 (5), 867–880.
Fan, G., Cao, J., Wang, J., 2010. Functional data classification for temporal gene expression data with kernel-induced random forests. In: IEEE Symp. on Comput. Intell. in Bioinformatics and Computational Biology. pp. 1–5.
Ferrando, L., Ventura, N., Epifanio, I., 2020. Detecting and visualizing differences in brain structures with SPHARM and functional data analysis. NeuroImage 222, 117209.
Frank, E., Hall, M., 2001. A simple approach to ordinal classification. In: European Conference on Machine Learning. Springer, pp. 145–156.
Gutiérrez, P., Pérez-Ortiz, M., et al., 2016. Ordinal regression methods: Survey and experimental study. IEEE Know. Data En 28 (1), 127–146.
Hall, P., Poskitt, D.S., Presnell, B., 2001. A functional data–analytic approach to signal discrimination. Technometrics 43 (1), 1–9.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.
Hornung, R., 2020. Ordinal forests. J. Classification 37, 4–17.
Lila, E., Aston, J.A.D., Sangalli, L.M., 2016. Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. Ann. Appl. Stat. 10 (4), 1854–1879.
Pierola, A., Epifanio, I., Alemany, S., 2016. An ensemble of ordered logistic regression and random forest for child garment size matching. Comput. Ind. Eng. 101, 455–465.
Simó, A., Ibáñez, M.V., Epifanio, I., et al., 2020. Generalized partially linear models on Riemannian manifolds. J. R. Stat. Soc. C-Appl. 69 (3), 641–661.
Wang, B., Shi, J.Q., 2014. Generalized Gaussian process regression model for non-Gaussian functional data. JASA 109 (507), 1123–1133.
Yao, F., Müller, H.-G., Wang, J.-L., 2005. Functional data analysis for sparse longitudinal data. JASA 100 (470), 577–590.